

2021

International

Conference

On Future Intelligent

Computing

(icfic 2021)

Theme Future Intelligent Computing

Date November 18-19, 2021

Venue Daegu University (Hybrid Online/Offline)

Website <https://icfic.daegu.ac.kr>



WELCOME ADDRESS

2021 International Conference On Future Intelligent Computing

ICFIC

In the face of the 4th Industrial Revolution, an ecosystem of hyper-connectivity, hyper-convergence, and super-intelligence is being formed, and related research is being actively carried out in various fields such as education, culture, and entertainment. Artificial intelligence convergence technology has shown many achievements with the development of computer performance, but has not yet penetrated deeply into our daily life.

Our center will hold the 2021 International Conference on Future Intelligent Computing (ICFIC 2021) at Daegu University. Under the theme of Future Intelligent Computing, this conference will provide a forum for discussions with researchers from universities, industries, and research institutes in related fields such as artificial intelligence, big data, Internet of Things, cloud computing, block chain, and intelligent robots.

We look forward to your submission and participation in this ICFIC 2021, as beneficial topic lectures, presentations, and high-quality academic papers will be presented. Thank you!

Director of ICT Convergence Demonstration Center
Prof. Jeong-Tak Ryu (Daegu Univ.)

Director of Mixed Reality Convergence Research Center
Prof. Kyuman Jeong (Daegu Univ.)



ICFIC2021 COMMITTEE

General Chair

- Jeong-Tak Ryu (Daegu University, Korea)

General Vice Chairs

- Kyuman Jeong (Daegu University, Korea)
- Sang Chul Suh (Texas A&M University, USA)

Conference Secretary and Finance Chair

- Donghwa Lee (Daegu University, Korea)

Organizing Chair

- Kyung-Ki Kim (Daegu University, Korea)

Program Chair

- Yoosoo Oh (Daegu University, Korea)

Publication Chair

- Heungwoo Nam (Daegu University, Korea)

Publicity Chairs

- Junho Yeo (Daegu University, Korea)
- Jung Jaewoo (Daegu University, Korea)

Special Session Chairs

- Hyungwoo Lee (Sookmyung University, Korea)
- Hosub Lee (Daegu University, Korea)
- Hyung Gyu Lee (Dongsung Women's University, Korea)
- Yujung Jung (Daegu University, Korea)
- Do Yongtae (Daegu University, Korea)

International Cooperation Chairs

- Synho Do (Massachusetts General Hospital, Harvard Medical School, USA)
- Hideaki Okada (Kyushu Institute of Technology University, Japan)
- Ka Lok Man (Xi'an Jiaotong-Liverpool University, China)
- Xin-She YANG (Middlesex University, UK)
- Mou Ling Dennis WONG (Swinburne University of Technology, Malaysia)
- Yoon Yong Hwang (Chosun University, Korea)
- Yan Wu (Computer Science, Bowling Green State University, USA)
- Dong hwa Shin (Soongsil University, Korea)
- Yungjun Yoo (University of Maryland, USA)
- Kwangmin Kim (National High Magnet Field Lab, USA)

PROGRAM & CONTENTS**Thursday, November 18**

| Time | Session | Chair | No. | Title | Author(s) | Affiliation |
|----------------------|-------------------------------------|----------------------|------|--|---|-----------------------------------|
| 11/18 15:00-16:30 | Poster Presentation Session 1 | Prof. Junho Yeo | P-15 | A study on the Efficiency Optimization of Wireless Power Transmission by using Mechanical Control | Patrick Danuor, Young-Bae Jung | Hanbat National Univ. |
| | | | P-16 | Investigation of the Transmission Efficiency of Wireless Power Transfer System using Phase Control | Patrick Danuor, Young-Bae Jung | Hanbat National Univ. |
| | | | P-17 | Communication Range Test Using nRF24L01+ Transceiver | Jaeyeong Jung, In Kyung Kim, JunhoYeo, Jeong Tak Ryu | Daegu Univ., DEXCREW |
| | | | P-18 | U-net based Automatic Photo Adjustment | Junggi Lee, Kyeongbo Kong | POSTECH, Pukyong National Univ |
| | | | P-19 | Design of Measuring Equipment for ATS | Mi Rim Seo, Jooyul Park | Robot campus of Korea Polytechnic |
| 11/18 16:30-18:0 | Poster Presentation Session 2 | Prof. Donghwa Lee | P-20 | Angle Insensitive Plasmonic Color Filters Design by Deep Neural Network | Ki Won Jeong, Yun Seon Do | Kyungpook National Univ. |
| | | | P-21 | Implementation of a system for ATS Anomaly Detection | Min Gyeong Kim, Jooyul Park | Robot campus of Korea Polytechnic |
| | | | P-22 | Implementation of a system to verify the method of minimizing the impact of handling products for collaborative robots | Juhyun Park, Jaeho lee, Dongjun Lee, Teakseung Lee, Chang-O Koh, Byeongjun Lee, Sangho Seo, Jaesung Kong | Robot campus of Korea Polytechnic |
| | | | P-23 | Gripper research for thin film parts assembly using collaborative robots | Ju Hyun Jo, Min Kyun Park, Jeong Ri Kong | Robot campus of Korea Polytechnic |
| | | | P-24 | Beauty Sniper : a Web-based Virtual Makeup System | Young-Seok Noh, Seong-Cheol Park, Jun-Won Kim, Hiecheol Kim | Daegu Univ |
| | | | P-25 | A system for detecting safety helmets at construction sites | Jin-Wook Lee, Jun-Sik Kim, Min-hoKwon, Hiecheol Kim | Daegu Univ. |
| | | | P-26 | Audio Quality Restoration Method using Convolutional Neural Network | Hong-Jin Kim, Kyuman Jeong | Daegu Univ. |

PROGRAM & CONTENTS**Friday, November 19**

| Time | Session | Chair | No. | Title | Author(s) | Affiliation |
|------------------|---|---------------------|------|--|---|-------------------------------|
| 09:30-10:00 | Opening Ceremony | | | | | |
| 10:00-10:40 | Keynote Speech | Prof. SangC.Suh | S-01 | Intelligent Computational Edge: From Human/Machine Augmentation to Cognitive Engineering | Prof. Radmila Juric | Univ. of South-Eastern Norway |
| 10-minutes break | | | | | | |
| 10:50-11:10 | Invited Talk | Prof. Kyuman Jeong | S-02 | Linear Regression Analysis of the Impact of COVID-19 Pandemic Health Interventions on Mental Health in the United States | Prof. Sang C. Suh, Amy Hayes | Texas A&M Univ. |
| 11:10-11:30 | | | S-03 | A Machine Learning Approach to Zero-day Detection | Prof. Jinhoo Kim | Texas A&M Univ. |
| 11:30-11:50 | | | S-04 | System Design for Biomedical and Rehabilitation Systems | Prof. Jeonghee Kim | Texas A&M Univ. |
| Lunch | | | | | | |
| 13:00-13:15 | Oral Presentation Session 1 (Future Computing) | Prof. Hyung Gyu Lee | P-01 | Efficient Hyperdimensional Computing using Monte Carlo Tree Search | Toygun Basaklar, Suat Gumussoy, Umit Ogras | Univ. of Wisconsin-Madison |
| 13:15-13:30 | | | P-02 | Enabling Robust Human Activity Recognition through Generative Adversarial Networks | Dina Hussein, Ganapati Bhat | Washington State Univ. |
| 13:30-13:45 | | | P-03 | Fed Group: Efficient Federated Learning via Decomposed Similarity-Based Clustering | Moming Duan, Duo Liu, Xinyuan Ji, Renping Liu, Liang Liang, XianzhangChen, Yujuan Tan | Chongqing Univ. |
| 10-minutes break | | | | | | |
| 13:55-14:10 | Oral Presentation Session 2 (Electric Power Applications for Automotive Industry) | Prof. Jae-Woo Jung | P-04 | Low Cost Shape Sensing Technique for Soft Robotic Actuators | Rafsan Al Shafatul Islam Subad, LiamB. Cross, Kihan Park | Univ. of Massachusetts |
| 14:10-14:25 | | | P-05 | Distance Teaching and Learning for Artificial Intelligence in Engineering Education at University | Jaehyun Lee | MathWorks |
| 14:25-14:40 | | | P-06 | Analytical Model for Circuit Parameter Estimation of Permanent Magnet Synchronous Motor for EV Traction | Min-Ro Park | Korea Institute of Robotics & |
| 14:40-14:55 | | | P-07 | Surrogate Model Based Design Optimization of Ultra-High-Speed Motor Stator Shape to Minimize Loss | Sun-Yong Shin, Ji-Hyeon Lee, Dong-Min Kim, Myung-Seop Lim | Hanyang Univ. |
| 14:55-15:10 | | | P-08 | Comparative Study on the Optimal Design of Linear Oscillating Actuator using Kriging and Deep Neural Network | Peel-Joong Kim, Soo-Hwan Park, Myung-Seop Lim | Hanyang Univ. |
| 10-minutes break | | | | | | |

PROGRAM & CONTENTS

Friday, November 19

| Time | Session | Chair | No. | Title | Author(s) | Affiliation |
|------------------|---|----------------------|------|---|--|--|
| 15:20-15:40 | Oral Presentation Session 3 (Intelligent Systems) | Prof. HoSub Lee | S-05 | [Invited Talk] A Ring-shaped Soft Robotic Actuator for Mimicking Gastric Motility | Prof. Peter Xu | Univ. of Auckland |
| 15:40-15:55 | | | P-09 | Real-Time Vision-based Segmentation Algorithm for Autonomous Vessels | Hanguen Kim, Donghoon Kim | Seadronix Corp. |
| 15:55-16:10 | | | P-10 | Multi-sensor Fusion-based Vessel Monitoring System using Artificial Intelligence | Hanguen Kim, Donghoon Kim | Seadronix Corp. |
| 16:10-16:25 | | | P-11 | Development of Military Tactical Training Simulator Using LVC-based Dron | Sungwook Jung, Jonghong Park, Sung-chan Choi, Wonseok Jung, and Il-yeop Ahn | Korea Electronics Technology Institute(KETI) |
| 10-minutes break | | | | | | |
| 16:35-16:55 | Oral Presentation Session 4 (IoT/RFID) | Prof. Youchung Chung | S-06 | [Invited Talk] Ear tag type sustainable sensing platform for livestock | Prof. Haruichi Kanaya | Kyushu Univ. |
| 16:55-17:10 | | | P-12 | UHF RFID Sensor Tag Comparison and Temperature Error Analysis for Covid-19 Cold Chain Logistics | Minkyong Kim, Youchung Chung | Daegu Univ. |
| 17:10-17:25 | | | P-13 | Broadband Planar Array Antenna with High-Efficiency for Synthetic Aperture Radar Applications | Patrick Danuor, Young-Bae Jung | Hanbat National Univ. |
| 17:25-17:40 | | | P-14 | A Study on Correlation between the Millimeter-wave Array Antenna and Digital Wireless Communication in the 5G Mobile System | YeJune Seo, Munsu Jeon, JunghyunCho, Jiyeon Jang, Yejin Lee, and Sungtek Kahng | Incheon National Univ. |
| Closing Ceremony | | | | | | |

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | <i>X</i> |
| Poster Preferred | |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|-----------------------------------|
| Title | Professor |
| Name | Jung Young-Bae |
| Affiliation | Hanbat National University |
| Address | Daejeon 34158, South Korea |
| E-mail | ybjung@hanbat.ac.kr |

A study on the Efficiency Optimization of Wireless Power Transmission by using Mechanical Control

Patrick Danuor and Young-Bae Jung*

Electronic Engineering Department, Hanbat National University
Daejeon 34158, South Korea

*Corresponding Author: ybjung@hanbat.ac.kr

Abstract – In this paper, a study of the effect of coil mechanical rotation on the transmission efficiency of a magnetic resonance power transfer system is presented. In the study, a model consisting of an arrangement structure composed of multiple coils is used. The study shows that the transmission efficiency can be optimized through the mechanical control in addition to the phase control of the individual coils in the existing coil arrangement structure.

Keywords: Coil arrangement, magnetic field, mechanical rotation, transmission efficiency, wireless power transfer.

1 Introduction

There has been a growing interest in wireless power transmission (WPT) technology as the use of electronic devices such as smart phones; tablets, etc. have witnessed a massive proliferation. The concept of wireless power transfer has attracted a lot of attention and has witnessed a significant amount of research in both industry and academia [1-3].

Inductive coupling, which is the near-field method of wireless power transfer technology, is already in fierce competition with many companies securing similar technologies, as high efficiency can be achieved as far as the coils are close together [4,5]. Thus to achieve high efficiency at larger distances, magnetic resonance coupling is used. Mid- and long-range wireless power transmission, required by society and industry, has received insufficient amount of research [6].

The transmission efficiency of wireless power transfer technology however, is very low at long distances thus wireless power transfer based on magnetic resonance has become an active research subject worldwide.

This paper presents a study to optimize the power transmission efficiency through a magnetic field coil arrangement structure and a coil rotation method for mid-range wireless power transmission at the level of tens of centimeters (cm).

2 Proposed Method

In order to investigate the transmission efficiency of wireless power transfer through the mechanical rotation of

the coils, a test bed was constructed. In the experimental set-up, the transmitter coils, i.e. $T_1, T_2 \dots T_n$ are placed at a predetermined distance away from each other. The receiver coils are placed 40 cm above the transmitter coil arrangement. Figure 1 illustrates the conceptual diagram of the wireless power system.

To verify the effect of the mechanical rotation of the coils on the transmission efficiency, the transmission coils are rotated as demonstrated in Figure 2. The intensity of the induced voltage was measured at the receiver coil as the transmitter coils were being rotated and a significant voltage difference of 1.94V was recorded.

This significant change in voltage confirms the effect that the mechanical rotation of the coils has on the transmission efficiency of wireless power transfer systems.

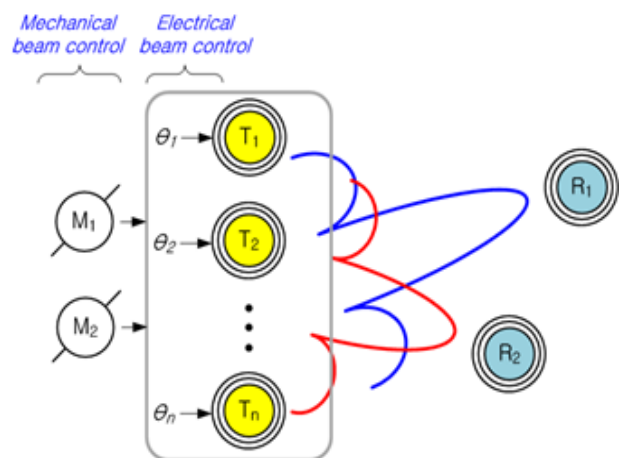


Figure 1. Conceptual diagram of the WPT system using Array of Coils

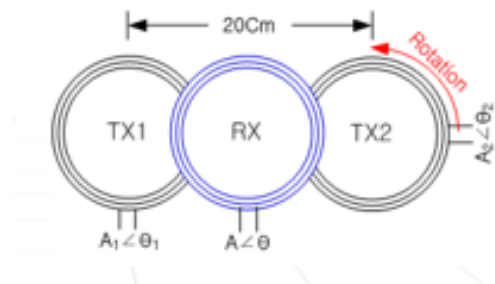


Figure 2. Schematic diagram of WPT Mechanical Control Test

3 Conclusions

In this paper, we have presented a study on the effect of mechanical control on the transmission efficiency of wireless power systems via magnetic resonance coupling. Our study confirms that the transmission efficiency is greatly affected when the coils are mechanically rotated.

Acknowledgements

This work was conducted with the support of the Information and Communication Technology Promotion Center with the Jeong (Ministry of Science and ICT) in 2017 (No. 2017-0-00653).

References

- [1] J. Jadidian and D. Katabi, "Magnetic MIMO: How to charge your phone in your pocket," Proc. of the Annual International Conference on Mobile Computing and Networking, Sep. 2014.
- [2] Kazmierkowski, M. P., and A. J. Moradewicz. "Contactless energy transfer (CET) systems—a review." *2012 15th international power electronics and motion control conference (EPE/PEMC)*. IEEE, 2012.
- [3] Lee, C. K., W. X. Zhong, and SY Ron Hui. "Recent progress in mid-range wireless power transfer." In *2012 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 3819-3824. IEEE, 2012.
- [4] Kurs, Andre, Aristeidis Karalis, Robert Moffatt, John D. Joannopoulos, Peter Fisher, and Marin Soljačić. "Wireless power transfer via strongly coupled magnetic resonances." *science* 317, no. 5834 (2007): 83-86.
- [5] Agbinya, Johnson Ihyeh. "Investigation of near field inductive communication system models, channels and experiments." *Progress In Electromagnetics Research B* 49 (2013): 129-153.
- [6] Van Thuan, Nguyen, Seok Hyon Kang, Jung Han Choi, and Chang Won Jung. "Magnetic resonance wireless power transfer using three-coil system with single planar receiver for laptop applications." *IEEE Transactions on Consumer Electronics* 61, no. 2 (2015): 160-166.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|-----------------------------------|
| Title | Professor |
| Name | Jung Young-Bae |
| Affiliation | Hanbat National University |
| Address | Daejeon 34158, South Korea |
| E-mail | ybjung@hanbat.ac.kr |

Investigation of the Transmission Efficiency of Wireless Power Transfer System using Phase Control

Patrick Danuor and Young-Bae Jung*

Electronic Engineering Department, Hanbat National University
Daejeon 34158, South Korea

*Corresponding Author: ybjung@hanbat.ac.kr

Abstract – In this paper, the transmission efficiency of wireless power transfer via magnetic resonance coupling is studied using phase control. The proposed model set up consist of an arrangement structure composed of multiple transmitter coils with a single receiver coil. From the study, it was verified that the wireless power transmission efficiency is affected as the phases of the transmitter coils are varied. Thus, phase variation can be implemented in wireless power transfer systems to optimize the transmission efficiency for the enhancement of wireless power transfer.

Keywords: High efficiency, resonant coupling, phase, wireless power transfer.

1 Introduction

Wireless power transfer (WPT) technology via magnetic resonance coupling is one the promising technologies for energy transfer, as it can eliminate the use of the wires and batteries, thus increasing the mobility, convenience, and safety of an electronic device for all users [1, 2].

The most widely used WPT technique is the inductive coupling method which finds applications in charging of handheld devices such as smart phones and RFID tags etc. [3]. Here, induction coils are used to transfer electrical energy from a power source to an electrical load across an air gap [4]. The transmitter coil produces an electromagnetic field and transfers the energy to a receiver coil. The receiver coil further converts it into electrical power.

A number of research works have been done concerning the near-field method, with its applications being utilized by many companies [5, 6]. However, the mid-to-long range wireless power transmission has witnessed a relatively insufficient amount of research.

In this paper, the effect of phase control of a multi-coil system on the transmission efficiency of wireless power transfer system is studied. The results show that the coil phases can be optimized to enhance the wireless power transmission efficiency.

2 Proposed Method

In order to study the effect of phase variation on the wireless power transmission efficiency, a test bed was constructed. In the test bed, the two transmitter coils are employed and separated from each other by a

predetermined distance. The receiver coil is 40 cm above the transmitter coil arrangement. Figure 1 illustrates the schematic diagram for the phase control test.

In the test, the phases of each transmitter coils are varied from 0° to 324° at intervals of 36° . As the phases were changing, the voltage at the rectifier circuit of the receiving coil was checked and recorded.

From the test, it was observed that when similar phases were used at the transmitter coils (i.e. $TX1_\theta = TX2_\theta$), the maximum voltage recorded was 2.06V. However when the phase difference between the transmitter coils was 180° , (i.e. $TX1_\theta=0^\circ$ and $TX2_\theta=180^\circ$ / $TX1_\theta=180^\circ$ and $TX2_\theta=0^\circ$) a voltage close to 0V was recorded. The experimental set-up is shown in Figure 2.

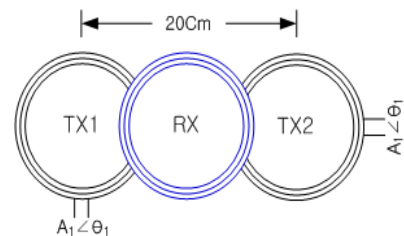


Figure 1. Schematic diagram of WPT Phase Control Test

3 Conclusions

In this paper, we have presented a study on the effect of phase control method on the transmission efficiency of wireless power transfer technology via magnetic resonance coupling. Our study shows that the transmission efficiency

is greatly affected by varying the phases of the transmission coils.

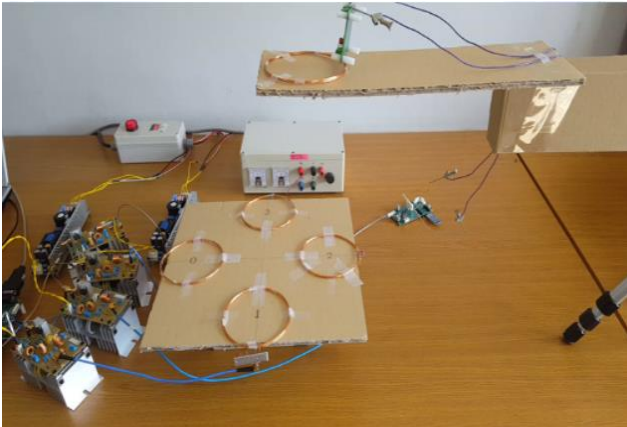


Figure 2. Experimental Set-up for WPT Phase Control Test

Acknowledgements

This work was conducted with the support of the Information and Communication Technology Promotion Center with the Jeong (Ministry of Science and ICT) in 2017 (No. 2017-0-00653).

References

- [1] Ibrahim, F.N.; Jamail, N.A.M.; Othman, N.A. (2016). "Development of wireless electricity transmission through resonance coupling". 4th IET Clean Energy and Technology Conference (CEAT 2016). Pp. 33 (5).
- [2] N. Shinohara, *Wireless Power Transfer via Radiowaves*. John Wiley & Sons, 2014.
- [3] R. Roy and K. Suresh, "A review of contactless energy transfer system", *International Research Journal of Engineering and Technology*, vol. 03, no. 04, p. 766, Apr-2016.
- [4] A. Tomar and S. Gupta, "Wireless power Transmission: Applications and Components", *International Journal of Engineering Research & Technology*, vol. 1, no. 5, 2012.
- [5] "Analysis, experimental results, and range adaptation of magnetically coupled resonators for wireless power transfer," *IEEE Trans. Ind. Electron.*, vol. 58, no. 2, pp. 544-554, Feb. 2011.
- [6] C. Zhang, D. Lin and S. Y. R. Hu, "Efficiency optimization method of inductive coupling wireless power transfer system with multiple transmitters and single receiver," in *Energy Conversion Congress and Exposition (ECCE)*, Milwaukee, WI, USA, 2016.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | O |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Communication Range Test Using nRF24L01 Transceiver |
| Name | Junho Yeo |
| Affiliation | Daegu University |
| Address | Department of information and communication engineering, Daegu University, 201 Daegudae-ro, Gyeongsan 38453, Gyeongbuk, Korea |
| E-mail | jyeo@daegu.ac.kr |

Communication Range Test Using nRF24L01+ Transceiver

Jaeyeong Jung¹⁾, In Kyung Kim²⁾, Junho Yeo^{3)*}, Jeong Tak Ryu¹⁾

¹⁾ Department of electronics engineering, Daegu university, Gyeongsan, Korea

²⁾ DEXCREW, Daegu, Korea

³⁾ Department of information and communication engineering, Daegu university, Gyeongsan, Korea

*Corresponding Author: jyeo@daegu.ac.kr

Abstract – Experiment results for communication range tests using an nRF24L01+ transceiver are presented in this paper. An Arduino-based radio communication system with the nRF24L01+ transceiver is designed for communication range tests. Communication range tests are conducted using two Arduino-based radio systems with the nRF24L01+ transceiver in an open area using four different transmitter (Tx) power settings. Each communication test is designed to transmit different data for a total of 100 times at 500ms intervals. Success rate for each communication test is calculated by counting for the case when the sending system successfully received the returned data from the receiving system in 500 msec. It is found that maximum communication range for RF24_PA_MAX is about 75 m, which is the longest among the four settings. Note that the output power of the power amplifier in the RF transmitter for RF24_PA_MAX is the highest among the four.

Keywords: Communication range, Internet of things (IoT), nRF24L01+, Arduino, Transmitter (Tx) power setting.

1 Introduction

The Internet of Things (IoT), which combines communication functions and various sensors with objects to connect to the Internet, is a core technology of the fourth industrial revolution, and is widely applied to smart homes, smart factories, and smart cities. Various wireless communication technologies are required to realize the IoT. Candidate wireless technologies would be cellular, Wi-Fi, Bluetooth, and IEEE 802.15.4-based wireless technologies. IEEE 802.15.4 is a standard for low-data rate wireless personal area network (WPAN). ZigBee and Z-wave are the examples of IEEE 802.15.4-based WPAN wireless technologies [1].

ZigBee has been most widely used in smart home and building automation applications in a form of a wireless ad-hoc network using star, tree, and mesh topologies. However, ZigBee modules are expensive compared to other modules using different protocols [2]. As an alternative to ZigBee modules, nRF24L01+ modules from Nordic semiconductor, Norway are available. The nRF24L01+ is a cheap single chip radio transceiver operating in the global, license-free ISM (industrial, scientific, and medical) band at 2.400 – 2.4835 GHz [3]. It has been widely used for a wide range of applications, such as wireless mouse and keyboard, wireless remote control, asset tracking, monitoring, medical sensor, somatosensory equipment, home automation, active RFID, NFC, and gaming [4].

In this paper, the results for communication range tests using the nRF24L01+ transceiver in an open area are presented. In order to design a radio communication system with the nRF24L01+, a microcontroller unit (MCU) and a

few external passive components are required. It can be operated and configured by using the MCU through a serial peripheral interface (SPI). In this work, an Arduino Uno board, which is the most popular open-source hardware platform, has been chosen for controlling and sending/receiving messages using the nRF24L01+. Four different Tx power settings with 0 dBm, -6 dBm, -12 dBm, and -18 dBm are used for the communication range test.

2 Design of Arduino-based Radio System with nRF24L01+ transceiver

An Arduino Uno-based radio communication system with the nRF24L01+ transceiver is designed for communication range tests in an open area.

2.1 nRF24L01+ Transceiver Module

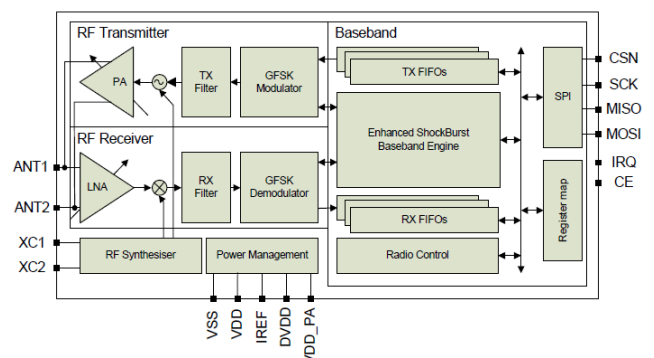


Figure 1. Block diagram of nRF24L01+ transceiver [3].

The nRF24L01+ consists of RF transmitter/receiver, RF synthesizer, Tx/Rx filter, GFSK (Gaussian frequency shift keying) modulator/demodulator, and baseband logic circuits, as shown in Figure 1. Available data rates are 250 kbps, 1Mbps, and 2 Mbps. RF channel bandwidth is 1 MHz for 250 kbps and 1 Mbps, whereas it is 2 MHz for 2 Mbps.

In order to adjust the output power of the power amplifier (PA) in the RF transmitter, the PA control is used. In Tx mode, PA control has four programmable steps, RF24_PA_MIN, RF24_PA_LOW, RF24_PA_HIGH, and RF24_PA_MAX, as shown in Table 1.

Table 1. RF output power setting [3].

| Arduino Code | SPI RF-SETUP (RF_PWR) | RF output power (dBm) | DC current Consumption (mA) |
|--------------|-----------------------|-----------------------|-----------------------------|
| RF24_PA_MAX | 11 | 0 | 11.3 |
| RF24_PA_HIGH | 10 | -6 | 9.0 |
| RF24_PA_LOW | 01 | -12 | 7.5 |
| RF24_PA_MIN | 00 | -18 | 7.0 |

Receiver sensitivities for 250 kbps, 1Mbps, and 2 Mbps data rates are -94 dBm, -85 dBm, and -82 dBm, respectively. Receiver current consumption for 250 kbps, 1Mbps, and 2 Mbps data rates are 12.6 mA, 13.1 mA, and 13.5 mA, respectively.

Figure 2 shows the schematic of the nRF24L01+ module with single ended matching network crystal, bias resistor, and decoupling capacitors for RF layouts with single ended 50 ohm RF output. The PCB layout and photograph of the nRF24L01+ module with an internal meander-line antenna is shown in Figure 3.

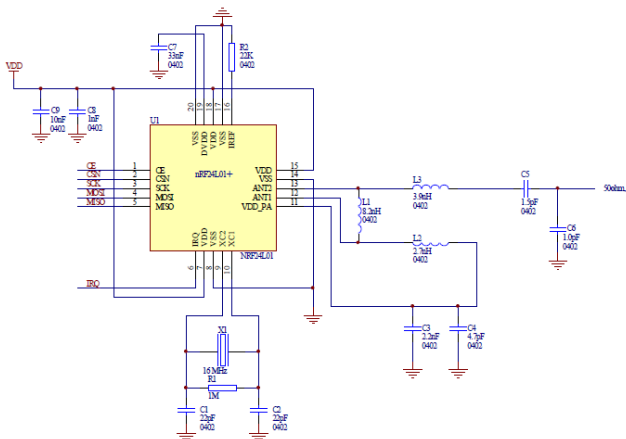


Figure 2. Schematic of the nRF24L01+ module [3].

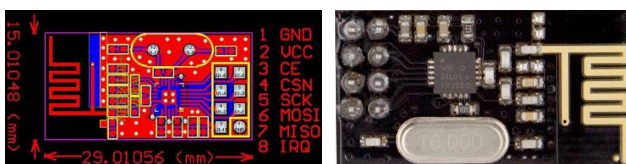


Figure 3. PCB layout and photograph of the nRF24L01+ module [5].

2.2 Arduino-based Radio System with nRF24L01+ transceiver

An Arduino Uno board is used to design a radio communication system with the nRF24L01+ transceiver. A schematic and photograph of the Arduino-based radio system with the nRF24L01+ transceiver is shown in Figures 4 and 5

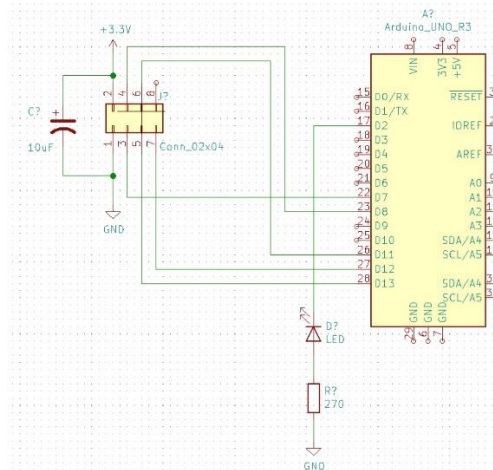


Figure 4. Schematic of the Arduino-based radio system with the nRF24L01+ transceiver.

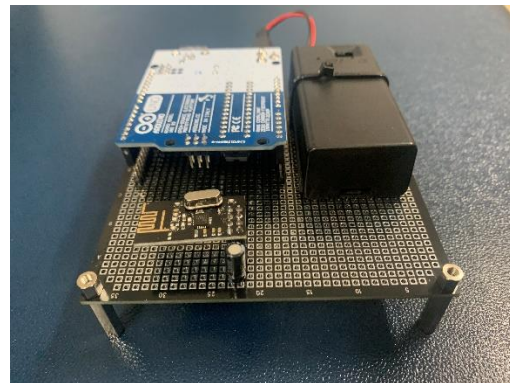


Figure 5. Photograph of the Arduino-based radio system with the nRF24L01+ transceiver.

3 Experiment Results

Communication range tests are conducted using two Arduino-based radio systems with the nRF24L01+ transceiver in an open area, as shown in Figure 6. Note that Two Arduino-based radio systems are installed at a position of 1 m above the ground using a tripod support.

Communication test is carried out by dividing the sending and the receiving systems. The sending system transmits data to the receiving system first, and after receiving the data, the receiving system must be transferred back to the sending system before it is judged that communication is complete. The sending system transmits once and sends data again after a predetermined delay time

of 500 msec. The sending data consisting of a predetermined character data and an increased communication count is sent to the receiving system. After 100 transmissions, each communication test ends. When data is received from the sending system, the receiving system returns the predetermined character data and a count value by adding 100 to the number of the received transmission count. Success rate for each communication test is calculated by counting for the case when the sending system successfully received the returned data from the receiving system in 500 msec.



Figure 6. Communication range test setup.

Figure 7 shows communication range test results for four different Tx power settings, RF24_PA_MIN, RF24_PA_LOW, RF24_PA_HIGH, and RF24_PA_MAX. Distance between the sending and the receiving systems is varied starting from 5 m to maximum communication range for each Tx power setting. Note that a total of three communication tests are conducted at each distance.

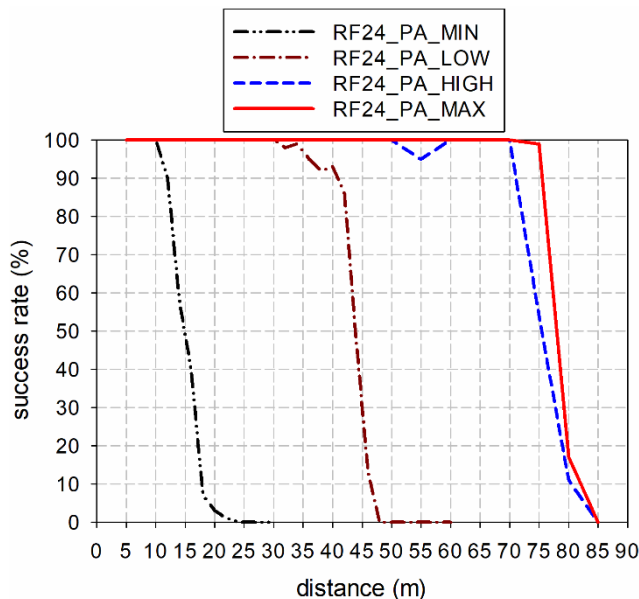


Figure 7. Communication range test results.

It is observed from Figure 7 that maximum communication range for RF24_PA_MAX is the longest

among the four settings. Communication success rate for RF24_PA_MAX remains around 100% until 75 m. It is slightly reduced to 70 m for RF24_PA_HIGH. For RF24_PA_LOW, it is further reduced to 30 m. The shortest communication range among the four settings is 10 m for RF24_PA_MIN.

4 Conclusions

We have presented experiment results for communication range tests using nRF24L01+ transceiver. An Arduino Uno-based radio communication system with the nRF24L01+ transceiver is designed for communication range tests. Experiment results show that maximum communication range for RF24_PA_MAX is about 75 m, which is the longest among the four settings.

The proposed Arduino-based radio communication system with the nRF24L01+ transceiver will be used for various IoT and ITS (intelligent traffic system) applications.

References

- [1] R. Mubashar, M. A. B. Siddique, A. U. Rehman, A. Asad, and A. Rasool, "Comparative performance analysis of short-range wireless protocols for wireless personal area network," *Iran Journal of Computer Science*, Vol 4, pp. 201–210, Apr. 2021.
- [2] H. Saha, S. Mandal, S. Mitra, S. Banerjee, and U. Saha, "Comparative performance analysis between nRF24L01+ and XBEE ZB module based wireless Ad-hoc networks," *International Journal of Computer Network and Information Security*, Vol. 9, pp. 36-44, Jul. 2017.
- [3] Nordic Semiconductor, "nRF24L01 single chip 2.4GHz transceiver product specification", Sep. 2008.
- [4] R. Sidqi, B. R. Rynaldo, S. H. Suroso, and R. Firmansyah, "Arduino based weather monitoring telemetry system using NRF24L01+," *The 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Surabaya, Indonesia, Vol. 9, Nov. 2017.
- [5] <https://devzone.nordicsemi.com/f/nordic-qa/57978/nrf24-modules-pcb>.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | U-net based Automatic Photo Adjustment |
| Name | Kyeongbo Kong |
| Affiliation | Pukyong National University |
| Address | 45, Yongso-ro, Nam-gu, Busan, Republic of Korea |
| E-mail | kbkong@pknu.ac.kr |

U-net based Automatic Photo Adjustment

Junggi Lee¹, Kyeongbo Kong^{2*}

¹POSTECH, ²Pukyong National University

¹77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Republic of Korea

²45, Yongsu-ro, Nam-gu, Busan, Republic of Korea

*Corresponding Author: kbkong@pknu.ac.kr

Abstract - Nowadays, photo editing methods have been popular to photographers. Since the camera cannot take the nature image, they want to adjust global tone or colors to improve the image quality. However, it is hard for photographers who are not familiar with photo editing to retouch photos. Also, it is hard for human to retouch photos because it needed consistent style to edit photos and took many times. In this paper, we proposed U-net based automatic photo adjustment. Since the encoder part of the network could find features of given image and decoder part of the network could make the output image with deconvolutional layer, the output image could predict retouched style of the expert.

Keywords: DNN, Photo editing, Scene parsing, U-net, Photo enhancement.

1 Introduction

After the increasing number of digital cameras including DSLR or smartphones, photo editing methods have been popular to photographers. Because the captured photos are just normal and cameras cannot take the nature or object as it is, they want to adjust global tone or colors to improve the image quality. However, it is hard for photographers who are not familiar with photo editing to retouch photos. Also, it is hard for human to retouch photos because it needed consistent style to edit photos and took many times.

For these reason, there are some methods to study automatic photo adjustment. Automatic photo adjustment is to match the style of professional photo retouchers and automatically adjust to the trained style with reference style. With automatic photo adjustment, photographers can improve the image quality through stylistic color and global or local tone adjustment.

2 Related Works

2.1 Automatic photo enhancement using deep neural network

There is a method to automatically adjust photo using Deep Neural Network (DNN) [1]. It used hand-crafted features such as intensity distributions, scene brightness and equalization curve etc [2] to train DNN. Also, it used pixel-wise features which are colors in Lab color space and contextual features in each pixel. Thus each pixel with above features was given to the network, and the output of

the DNN was predicted pixel of corresponding ground truth. The whole architecture of DNN in [1] was shown in Figure 1.

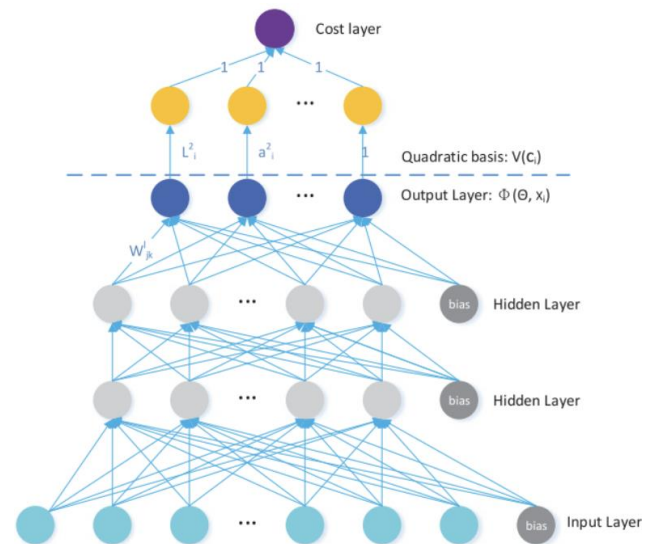


Figure 1. The architecture of DNN in [1].

2.2 MIT-Adobe fiveK dataset

For dataset to retouch photographs, there is a dataset with 5000 photographs taken with SLR camera [2], MIT-Adobe fiveK dataset. For the dataset, there were five retouchers who were in an art school to study adjusting the tone of photographs. They retouched each photographs with their own style. Each retouching style is quite different and we will use the retouched result of expert C that [1] also used.

[2] used the Euclidean distance in CIE Lab color space considering human perception to evaluate the result of retouching. In the dataset, the difference between the input image and retouched image was 5.5 on average in each pixel and the maximum was 23.7. We will also use the distance in Lab color space to compare the output images with other methods.

3 Proposed Method

In [1], it used hand-crafted global, color and contextual feature of given image to train its network. However, after Convolutional neural networks (CNNs) has achieved high performance in classification and object recognition, features from CNNs can represent image more than hand-crafted features. Thus, we will use CNN based network to get global and local feature of given image. Also, we can make the predicted output image with deconvolutional network like [3, 4]. We will introduce a network like Autoencoder to get the segmented result.

3.1 U-net

Like Autoencoder, U-net consists of a contracting path as encoder and an expansive path as decoder [5] as shown in Figure 2 . They used the output of U-net for segmentation in medical imaging community. For structure, it consisted of several skip layers to get more precise output result like [3]. Its feature map in each layers was transferred to decoder and concatenated with corresponding upsampled decoder feature map.

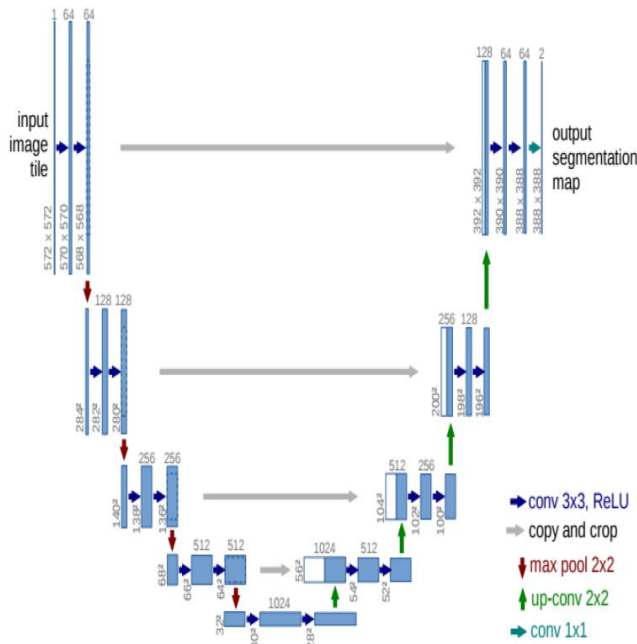


Figure 2. The architecture of U-net..

We will use U-net to get features from given image, and make the output image through expansive path in U-net.

Although original U-net was for segmentation, our U-net can predict enhanced image using ground truth image set.

3.2 Scene Parsing

Since the dataset might be retouched in semantic units such as object or background, semantic and contextual information of given image is quite implemental to match the ground truth enhanced image. Also, [1] used contextual features of image to train their network. Thus, we will train U-net with a scene parsing dataset in [6] which consists of 150 semantic categories. Then, the decoder part of the trained network is retrained by MIT-Adobe fiveK dataset to get the final predicted output image. Figure 3 shows our proposed method. First, the upper part of the network is pre-trained with the scene parsing dataset. Next, the encoder part of the network is fine-tuned for enhancement task with lower decoder part of the network. Finally, we can get the output image predicted to style of expert C.

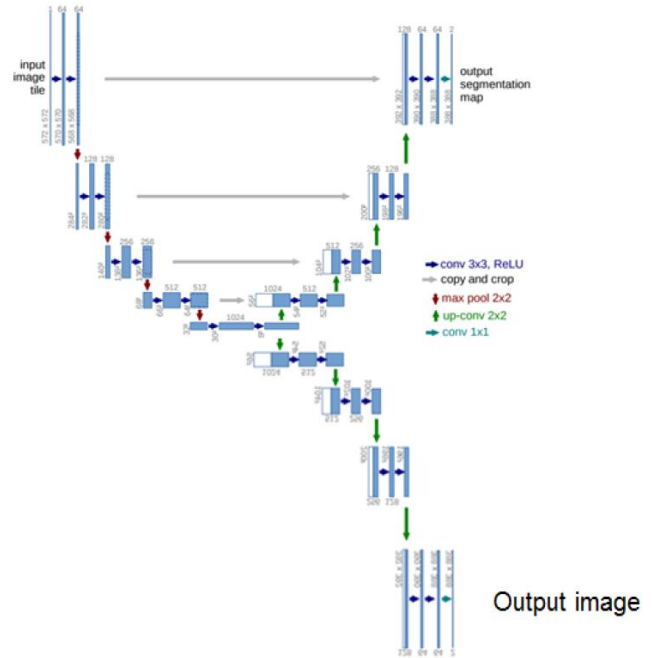


Figure 3. The architecture of proposed method.

4 Experiments

4.1 System Setup

We simulated two kinds of U-net, without and with scene parsing. For U-net without scene parsing, We just made the network same as original U-net. The input size of the network was [288, 512] because U-net used 2x2 pooling layer five times. In encoder part, the number of channel started with 8 and doubled to upper layers. For U-net with scene parsing, we trained the U-net with the scene parsing dataset. Then, we retrained the decoder part with MIT-Adobe fiveK dataset using same encoder part lowering

learning rate. We used 4750 images from the dataset to train, and remained 250 images were used for test.

4.2 Quantitative Result

Table 1 shows the quantitative result of the proposed method. The values in the table are L2 distances in the Lab color space. The performance of proposed method is worse than DNN [1]. The reason why proposed method is worse than previous method is difference in the number of dataset. Because the input of DNN [1] is pixel, the number of dataset is much larger than proposed method that uses image for input.

For scene parsing, the performance of proposed method without scene parsing is better than that of method with scene parsing. First, we cannot make the convergence of pre-training network with the scene parsing dataset for lack of time. If there is a pre-trained model with U-net, the result of proposed method with scene parsing gets better than method without scene parsing. Secondly, the semantic and contextual information is less significant than global or color feature in the dataset. In MIT-Adobe fiveK dataset, the expert retouched the image using global transformation [7]. Thus, the performance of proposed method with scene parsing is worse than method without scene parsing.

Table 1. Quantitative result.

| | DNN [1] | Proposed without scene parsing | Proposed with scene parsing |
|------------------------------------|---------|--------------------------------|-----------------------------|
| L_2 distances in Lab color space | 9.85 | 13.66 | 16.04 |

4.3 Qualitative Result

Figure 4 shows the Qualitative results from the test set. The result of proposed method can adjust the ground truth overall, but is worse than DNN [1].

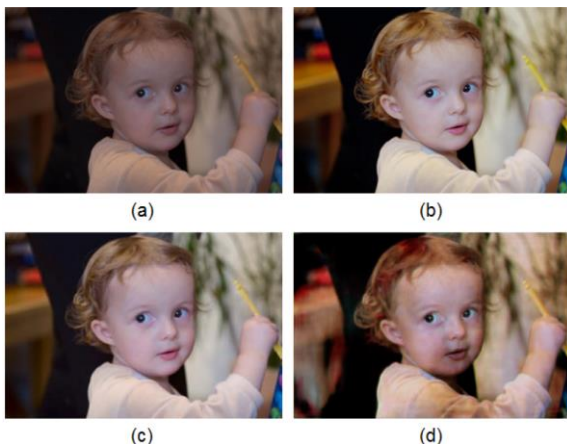


Figure 4. Qualitative result of proposed method, (a) Input; (b) Ground truth; (c) DNN [1]; (d) Proposed method without scene parsing.

5 Conclusions

In this paper, we proposed U-net based automatic photo adjustment. Because the encoder part of the network could find features of given image and decoder part of the network could make the output image with deconvolutional layer, the output image could predict retouched style of the expert. However, the result was worse than DNN [1] since the number of dataset (images) was much smaller than DNN [1] (pixels). We think that adjusting whole image is much harder than adjusting the pixel of image because the complexity of image and pixel is quite different. Therefore, the network size should be larger than before and lots of training dataset are needed.

References

- [1] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu, "Automatic photo adjustment using deep neural networks". *ACM Transactions on Graphics (TOG)*, 35(2):11, 2016.
- [2] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs", *Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.
- [4] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation.", *In Proceedings of the IEEE International Conference on Computer Vision*, pages 1520–1528, 2015.
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation.", *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [6] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Bar-riuso, and Antonio Torralba, "Scene parsing through ade20k dataset", *In Proc. CVPR*, 2017.
- [7] Sung Ju Hwang, Ashish Kapoor, and Sing Bing Kang, "Context-based automatic local image enhancement" *In European Conference on Computer Vision*, pages 569–582. Springer, 2012.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Prof. |
| Name | Joo Yul Park |
| Affiliation | Robot campus of Korea Polytechnic |
| Address | 1, Robot Campus-ro, Yeongcheon-si, Gyoungsangbuk-do, Korea |
| E-mail | jooyul.park@kopo.ac.kr |

Design of Measuring Equipment for ATS

Mi Rim Seo, Jooyul Park *

Department of Robot Electronics, Robot campus of Korea Polytechnic
1, Robot Campus-ro, Yeongcheon-si, Gyeongsangbuk-do, Korea

*Corresponding Author: jooyul.park@kopo.ac.kr

Abstract – The ground signal equipment of the railway has a long equipment, and there are various signals such as ATS (Automatic Train Stop), ATC (Automatic Train Control), and crossing signals. In this paper, we propose a method to detect ATS signals on the ground in real time using an on-board detection device. We used the DSP(Digital Signal Processor) of TI chip to collect the magnitude and frequency characteristics of the ground ATS signal in a high-speed train. The collected signal sends to the main control device through internal communication and analyzed. Using the results of this study, it will be possible to develop a system that can analyze various ground signals such as ATC and crossing signals.

Keywords: ATS, Onboard Signal, Railway

1 Introduction

This paper is about the development of on-board signal system that detects the signal of the ATS installation section among the ground signal facilities of general railroads. ATS is a device that places a signal at a certain distance, detects the train installed on the track, and automatically applies the emergency brake if the signal is not observed.[1] Depending on the section of the ATS installation, 4 traffic lights (4 lights) and 5 lights (5 lights) are used. Currently, they are used on general trains (Saemaoul and Mugunghwa).[2] ATS has a ground element composed of a capacitor and a coil installed on the ground, and when the on-board device passes over the ground surface, a specific frequency is generated by resonance.[2] In this paper, we propose a method to measure the normal operation of an ATS ground company on the ground through a signal detector on the vehicle.

2 System Implementation

The configuration of ATS signal detector receives the response signal using the antenna and matching transformer as shown in Figure 1, uses a low-pass

hardware filter, and uses the digital signal processing processor to check the magnitude and frequency of the response signal. These processes are verified in real time. The magnitude, frequency, and input time of the received signal are measured. Figure 2 is a hardware board for measuring ATS input signal. Real-time measurement is performed using TI's DSP (TMS28335).

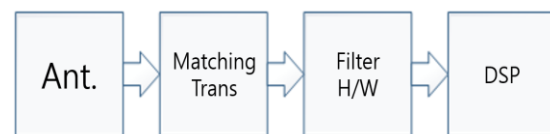


Fig. 1 Block Diagram of ATS board

Figure 3 is a screen of TI's Code Composer development tool to check whether the frequency and magnitude of a signal are normally received when a signal is input using a frequency generator. Using this development tool, we checked whether the function works normally. Figure 4 shows the actual ground signal on-board detection device, and the ATS analysis board is installed to deliver the real-time analysis result to the main control device using CAN communication.



Fig. 2 ATIS board of On-board Measuring Equipment

| Expression | Type | Value | Address |
|----------------------|---------------|-------|-----------------|
| freq | unsigned long | 78 | 0x0000C00E@Data |
| max_mag | unsigned int | 2045 | 0x0000C007@Data |
| j | unsigned int | 86 | 0x0000C005@Data |
| + Add new expression | | | |

Fig. 3 Real Time Frequency analysis of TI's Code-Composer

3 Conclusions and Future works

As a result of this study, normal operation was confirmed by conducting an experiment using a signal generator in the present laboratory. In the future, we plan to continue research on detecting various signals on the ground and collecting data through the development of on-board detection equipment and detection information analysis system for general railroad ground signal facilities.

References

- [1] J.H. Park, " A Study on the Definition of Terms for Domestic Train Control System", Proceedings of the Korean Society For Railway, 2015
- [2] Design Guidelines Auto Train Stop Device(ATS) Railway, KR S-07020, Korea Railroad Authority, 2016



Fig. 4 On-board Measuring Equipment for Ground Signaling System

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Angle Insensitive Plasmonic Color Filters Design by Deep Neural Network |
| Name | Yun Seon Do |
| Affiliation | School of Electronic and Electrical Engineering, Kyungpook National University |
| Address | 80, Daehak-ro, Buk-gu, Daegu, South Korea |
| E-mail | yuns.do@knu.ac.kr |

Angle Insensitive Plasmonic Color Filters Design by Deep Neural Network

Ki Won Jeong and Yun Seon Do *

School of Electronic and Electrical Engineering, Kyungpook National University,
Daehak-ro, Buk-gu, Daegu, South Korea

80,

*yuns.do@knu.ac.kr

Abstract - Plasmonic color filter (PCF) is more durable than conventional color filters and is suitable for high-resolution displays. PCFs determine filtering characteristics by adjusting design parameters as a type of structural color. However, if the structure is complicated, trial and error increase because many cases need to be verified. In addition, the resonance caused by phase matching between structures changes according to the angle of incident light. We reduce researchers' trial and error through deep neural network model (DNN) that enable reverse design. In addition, we propose a PCFs that shows characteristics independent to the angle of incident light through a coaxial hole and disk arrangement structure.

Keywords: Nanotechnology, Neural networks, Photonics, Optics

1 Introduction

Light incident on the metal vibrates free electrons on the metal surface to generate surface plasmon (SP). If there is a periodic arrangement of holes or grooves in the metal surface, surface plasmon resonance (SPR) occurs at a specific wavelength according to the period.[1] Localized surface plasmon resonance (LSPR) generated in the hole, propagated surface plasmon resonance (PSPR) propagating the metal surface, and incident light interact to generate extraordinary optical transmission (EOT).[2] Plasmonic color filters (PCFs) using EOT phenomena are more durable to heat, light, and chemicals than conventional color filters because periodicity of metal structures, not chemical properties, determines filtering properties. In addition, PCFs can be manufactured within small and thin dimensions in submicrons, so they can be used in small pixel displays and sensor technologies.[3] However, as the angle of incident light increases, the filtering color changes as the transmission peak is divided. Since these characteristics affect the viewing angle of the display and sensor, problem solving is necessary for actual use.

In this paper, we propose a new structure that coaxially hole and disk array plasmonic color filters (CHDPCFs). Compared to a PCF that uses only a hole array, the proposed color filter has a reduced peak shift and a reduced transmittance depending on the angle of incident light. The new structure was designed using reverse design deep neural network (DNN) model to reduce trial and error and inefficiency. Unlike conventional simulations, this design method can immediately recognize the design parameters of PCFs

with specific optical characteristics. It will be helpful for early researchers with limited background knowledge. This study is expected to contribute to increasing research efficiency and commercialization of plasmonic color filters.

2 Results

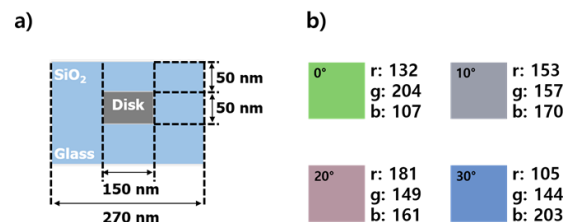


Figure 1. a) Structure of HPCF, b) RGB color of HPCF

Figure 1. a) is a hole array PCF structure. The HPCF has a 50 nm thick aluminum (Al) film with a hole diameter of 150 nm and a period of 270 nm. The holes in the film are filled with SiO₂ and stacked to 50 nm over the film. We investigated the optical response of PCFs using the 3D finite-difference time-domain method (FDTD solutions, Lumerical Inc., Canada).

When light is normally incidence on HPCF, green is transmitted. The RGB color values are 132, 204, and 107. As the angle of incidence increases to 30°, the transmitted RGB color coordinates change to 105, 144, and 203. The distance between the two color coordinates is 116.38.

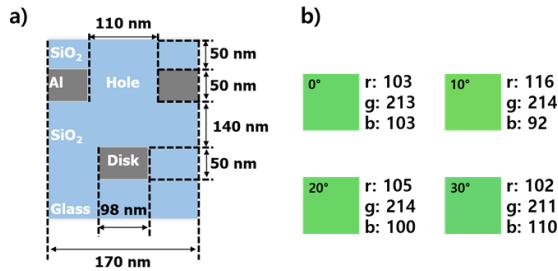


Figure 2. a) Structure of CHDPCF, b) RGB color of CHDPCF

In the suggested CHDPCF, a hole array Al film having a diameter of 110 nm is coaxially deposited on an Al disk array having a diameter of 98 nm. Both the disk and the hole array film are 50 nm thick. The distance between the hole and the disk is 140 nm, and the hole and disk array have a period of 170 nm. SiO₂ of 50 nm is additionally deposited on the hole array film.

When light is normally incidence on CHDPCF, green is transmitted. The color RGB values are 103, 213, and 103. CHDPCF still transmits green even when the angle of incidence increases to 30°. The coordinates of the transmitted color are 102, 211, and 110. The distance between the two color coordinates is 7.34.

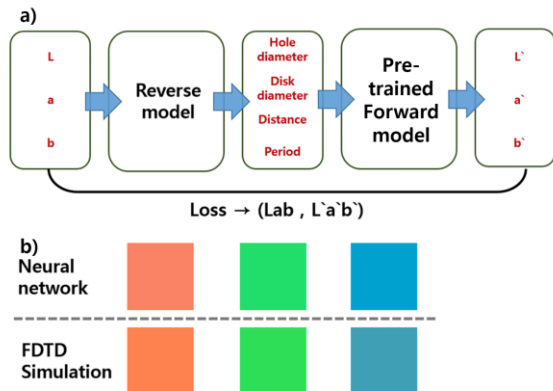


Figure 3. a) Deep neural network model for design PCF, b) Predict color from neural network and FDTD simulation result with same output parameter of reverse model

Design parameters to be considered in CHDPCF are the diameter of the hole, the diameter of the disk, the distance between the hole and the disk, and the period of the array. Identifying optical properties by changing parameters in each conventional way requires a lot of background knowledge. Moreover, it is very inefficient to repeat many cases and trials and errors. In this study, a reverse design DNN model was created using simulation data of CHDPCF. This model outputs a CHDPCF design parameter with the color space (CIELAB) desired by the researcher.

The reverse design DNN model is made based on the pre-trained forward DNN model. The forward DNN model is a model that predicts color space through design parameters. The forward DNN model may replace the FDTD simulation because the accuracy exceeds 99%. The reverse design DNN model predicts design parameters through color space, which are input values, and predicts colors again through the pre-trained forward DNN model. Both the input and output of this model are CIELAB, and training proceeds in a direction where the difference between the two coordinates decreases. By connecting the trained reverse model and forward model in series, researchers can use one neural network model that can find and verify design parameters of the desired color.

Figure. 3. b) shows the color when a specific color is input to the reverse design DNN model, and the output parameters are put into the forward DNN model and FDTD simulation, respectively. As a result, the similarity between the two colors means that the reverse model can find design parameters through input colors. Depending on the size of the structure, a simulation process that takes minutes from FDTD can be performed in seconds using reverse design models and forward models.

3 Conclusions

This study suggest a new structure with characteristics independent of the incident light angle. The proposed CHDPCF has less color change according to the incident light angle compared to the conventional hole array PCF. The CHDPCF will affect the industrial development of PCFs. In addition, a reverse design DNN model was used for the efficiency of the research. This not only reduces trial and error of early researchers in the field, but also shows the possibility that more complex structures can be studied more efficiently.

References

[1] William L. Barnes et al., "Surface plasmon subwavelength optics", *Nature*, Vol. 424, pp. 824-830, Aug. 2003.
 [2] T. W. Ebbesen et al., "Extraordinary optical transmission through sub-wavelength hole arrays", *Nature*, Vol. 391, pp. 667-669, Feb. 1998.
 [3] S. Daqiqeh Rezaei et al., "Nanophotonic Structural Colors", *ACS Photonics*, Vol. 8, pp. 18-33, Jul. 2020.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Prof. |
| Name | Joo Yul Park |
| Affiliation | Robot campus of Korea Polytechnic |
| Address | 1, Robot Campus-ro, Yeongcheon-si, Gyoungsangbuk-do, Korea |
| E-mail | jooyul.park@kopo.ac.kr |

Implementation of a system for ATS Anomaly Detection

Min Gyeong Kim, Jooyul Park *

Department of Robot Electronics, Robot campus of Korea Polytechnic
1, Robot Campus-ro, Yeongcheon-si, Gyongsangbuk-do, Korea

*Corresponding Author: jooyul.park@kopo.ac.kr

Abstract - Failure of the on-board signaling system causes serious problems in the safe operation of trains. Recently, as the driverless train system is installed, the reliability and availability of the on-board system become very important factors. The goal of this paper is to predict and repair on-board equipment failures in advance through an AI-based failure prediction system. It was implemented to predict the failure time of the ATP (Automatic Train Stop) system through multivariate anomaly identification by collecting various information from the PDT Line entry stage for the pre-departure test until entering the main Line and sending it to the artificial intelligence platform.

Keywords: ATS, Onboard Signal, Failure Prediction, AI

1 Introduction

When an abnormality occurs in an unmanned vehicle, safety risks increase and passenger inconvenience occurs due to delay in operation. So, Periodic preventive maintenance is carried out to check whether there are any abnormalities in the parts that cause problems. [1,2] For the safe and reliable operation of railways, it is necessary to develop a component status diagnosis technology through automated analysis of power train components and signal-related devices responsible for safety. With the development of deep learning technology for real-time data processing and anomaly detection, it is necessary to apply AI-railroad fields. In this paper, a study was conducted to detect abnormal signs of the ATS receiver in the on-board system in advance using an artificial intelligence platform and to secure the reliability of train operation through this. [3]

2 System Implementation

As shown in Fig. 1, the failure prediction system is structured collects the current and voltage of the ATS receiving board, the average amount of the ADC level of the received signal inside the DSP, and the number of CAN communication errors. It sends the collected data to the artificial intelligence platform in real time. It analyzes this and predicts the failure.

In case, to keep the value of the antenna input signal constant. Assuming that the train stops at the exact stop position on the PDT line, the input magnitude of the

signal is analyzed in real time through the ADC. When the DSP chip of the ATP receiving board performs ADC for decoding, the average voltage value of this value is transmitted through CAN communication and analyzed together, and distortion is generated using a function generator to create a different value than usual. In addition, the voltage and current supplied to the DSP chip were measured, and an abnormality was intentionally generated by using a hot air blower in the DSP to promote deterioration. These data were collected as shown in Fig. 2 and delivered to the AI platform.

3 Conclusions and Future works

Although there are various causes of failure and it is difficult to obtain sufficient data, it was confirmed that the input value and current amount of the ADC were different from normal when the DSP was operating abnormally, and it was outside the normal range. In the future, we will secure more data and improve the learning model to conduct research to increase accuracy so that it can be applied to actual train operation.

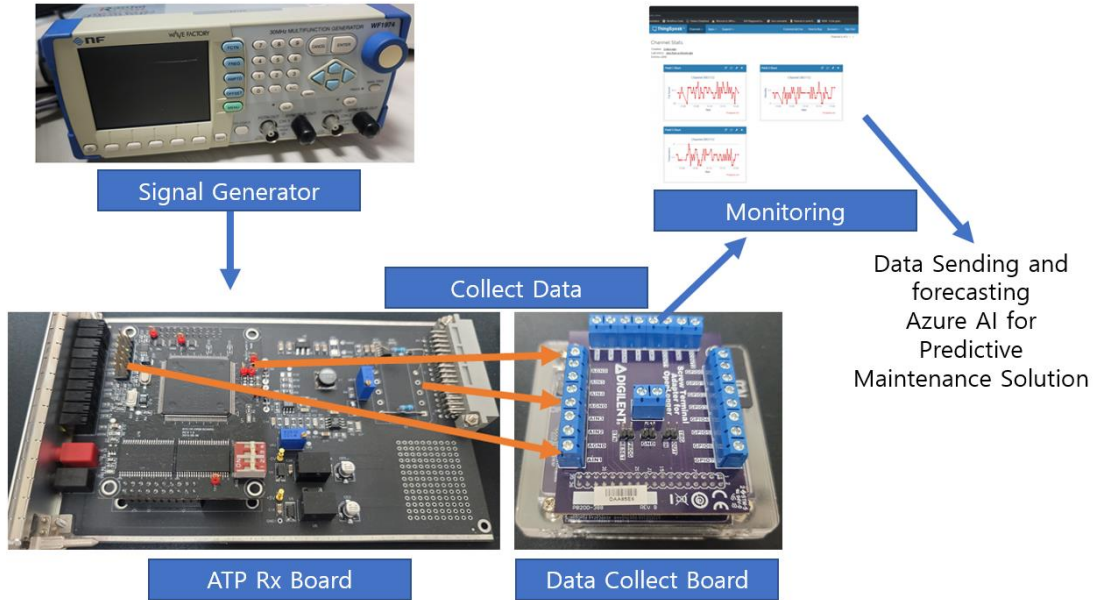


Fig. 1. Proposed System Diagram

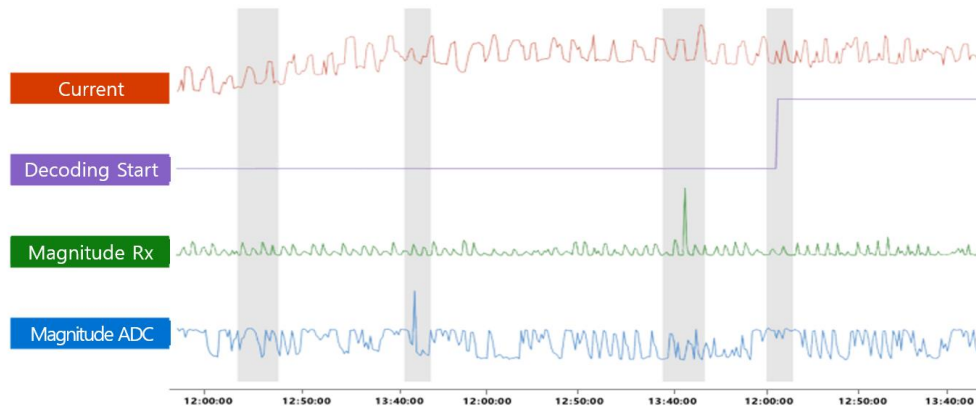


Fig. 2. Implementation system

References

- [1] Seok-Youn Han, Ki-Jun, Park, "Directions for the Application of Predictive Maintenance based on the Data of Train Control and Monitoring System of EMU", Proceedings of the Korean Society For Railway, pp. 509-511, October, 2018 .
- [2] Kwan-Sup Lee, Jong-Woon Kim, "A Study on Condition Based Maintenance for Rolling Stock", Proceedings of the Korean Society For Railway, pp. 1218-1221, October, 2016.
- [3] Yechan Jeon, Yonghyun Lee, Dong-Ju Kim, "Development of Rotating Equipment Anomaly Detection Algorithm based-on Artificial Intelligence", Proceedings of the Korean Society of Computer Information Conference, pp. 57-60, July, 2021.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Prof. |
| Name | Jaesung Kong |
| Affiliation | Robot campus of Korea Polytechnic |
| Address | 1, Robot Campus-ro, Yeongcheon-si, Gyeongsangbuk-do, Korea |
| E-mail | kongjs@kopo.ac.kr |

Implementation of a system to verify the method of minimizing the impact of handling products for collaborative robots

Juhyun Park, Jaeho lee, Dongjun Lee, Teakseung Lee, Chang-O Koh, Byeongjun Lee, Sangho Seo, Jaesung Kong *

Department of Robot Electronics, Robot campus of Korea Polytechnic
1, Robot Campus-ro, Yeongcheon-si, Gyeongsangbuk-do, Korea

*Corresponding Author: kongjs@kopo.ac.kr

Abstract - A verification system was implemented to optimize the robot control parameter setting to minimize the impact on the product in transport using a collaborative robot. A LabVIEW-based PC program was implemented to observe the real-time status of the robot. An embedded device including joystick components was implemented for various experimental conditions. UR3e, the product of the Universal Robots company, has been used. It was confirmed experimentally that product impact can be reduced by changing parameters such as speed, acceleration, and radius mixing used for movement control.

Keywords: Multi-Joint Robot, Cobot, impact, LabVIEW.

1 Introduction

The industrial use of multi-joint cooperative robots has recently increased significantly due to their unique low accident risk, human-friendly control method, and reasonable price. On the other hand, there are areas that need improvement in terms of industrial use, such as low motion accuracy, motion shake, and low payload compared to existing articulated industrial robots. In this paper, we are going to talk about the device setup for the study of how to minimize the impact of the handling product in the utilization of the multi-joint cooperative robot.

2 System Implementation

Figure 1 shows the signal flow diagram of the implemented system. A joystick-type control device has been configured to represent the external control device of the robot. A PC was used to transmit signals from the control device to the robot and to monitor the status of the robot in real time. The joystick control device is manufactured based on Arduino and has a function to receive signals from analog joystick components and transmit them to the PC through serial communication. The PC converts the signal obtained from Arduino into an appropriate signal for robot control through pre-processing and transmits control information to the robot through MODBUS-based communication. The PC program was written based on LabVIEW for real-time visual display of various transmitted data. The robot handles the target product based on the information transmitted to the PC. As the robot, URe3 model of

Universal Robots company, one of the latest and most popular products in the world, has been used. Fig. 2 shows the implemented system. We intend to utilize the force/torque sensor mounted on this robot as information representing the impact information received by the product. Through partial experiments that controlled not only position, but also speed and acceleration, it was partially confirmed that unnatural robot motion that occurred in simple control could be improved.

3 Conclusions and Future works

In the future, we intend to conduct research to quantify and optimize this area. Finally, through this study, it is expected that collaborative robots can broaden the field of application by enabling high-precision/low-impact handling.

References

- [1] Christoph Byner, Björn Matthias, Hao Ding , "Dynamic speed and separation monitoring for collaborative robot applications – Concepts and performance", *Robotics and Computer-Integrated Manufacturing*, Vol. 58, pp. 239-252, August 2019
- [2] Wang, L., "Collaborative robot monitoring and control for enhanced sustainability", *Int J Adv Manuf Technol*, Vol. 81, pp. 1433-1445, 2015.
- [3] Ana M. Djuric, R.J. Urbanic and J.L. Rickli, "A Framework for Collaborative Robot (CoBot) Integration in Advanced Manufacturing Systems", *International Journal of Materials and Manufacturing*, Vol. 9, No. 2. pp. 457-464, 2016.

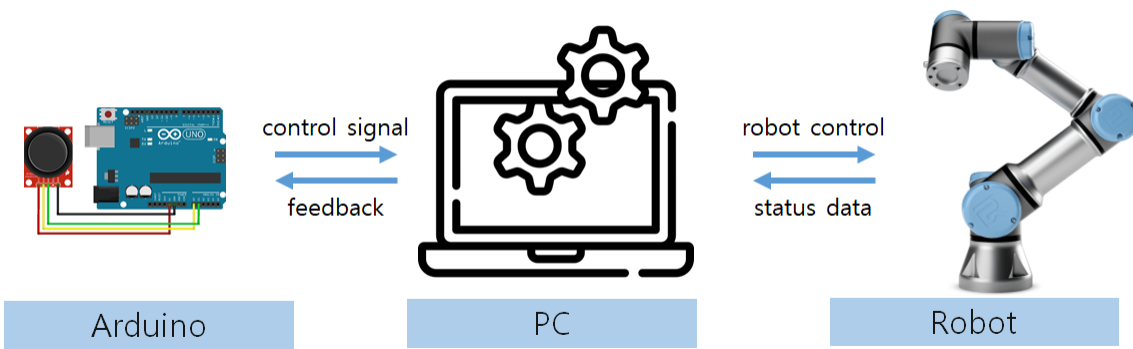


Fig. 1. Signal flow diagram among equipment



Fig. 2. Implementation system

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Prof. |
| Name | Jeong Ri Kong |
| Affiliation | Robot campus of Korea Polytechnic |
| Address | 1, Robot Campus-ro, Yeongcheon-si, Gyoungsangbuk-do, Korea |
| E-mail | balllee@kopo.ac.kr |

Gripper Research for Thin film parts Assembly using Collaborative Robots

Ju Hyun Jo, Min Kyun Park, Jeong Ri Kong *

Department of Robot Machinerics, Robot campus of Korea Polytechnic
1, Robot Campus-ro, Yeongcheon-si, Gyoungsangbuk-do, Korea

*Corresponding Author: balllee@kopo.ac.kr

Abstract – The finger part of the gripper for thin film products is designed to be able to hold thin parts, and a buffer spring is installed to alleviate the impact that may occur in contact with the bottom surface when gripping parts. The ball bush and stopper were designed as elements for linear driving of the shaft and mounted on the stopper plate. In order to test the motion detection of the gripper, proximity sense and fiber sense were installed, and the fiber sensor was installed as a top-view type screw type with adjustable detection distance, so that it can be applied to grippers of various shapes.

Keywords: Washer, Gripper, Supply, Feeding

1 Introduction

As we enter the era of the 4th industry, we are trying to solve the difficulties of manufacturing sites and the manpower shortage of small and medium-sized enterprises by introducing robots.[1] With the introduction of the smart factory, automation continues to grow in response to the demands of the times to increase corporate profits, improve productivity, reduce manufacturing costs, and solve manpower problems. However, when thin-walled parts are put into the assembly process, full automation is difficult.[2] This thesis research relates to the development of a gripper unit to be linked with a robot in the thin film product supply process, which is currently being done manually.

2 Gripper's Finger Encounter

In order to improve the current situation in which the assembly of thin film products that require sophisticated work is supplied by hand, the assembly process of a sequential supply method with the fingers of a fully automated gripper after sorting through a sorting process for smooth assembly with the other part in the input process method was studied. The finger part of the gripper for thin-film products is designed to hold thin parts, and a cushioning spring is installed to

mitigate the impact that may occur by contacting the bottom surface when gripping parts.[3] The ball bush and stopper are designed as elements for linear drive of the shaft and are mounted on the stopper plate. Proximity detection and optical fiber detection were installed to test the motion detection of the gripper, and the optical fiber sensor was installed in a top-view screw type with adjustable sensing distance so that it can be applied to various types of grippers. The superiority of the gripper unit has been confirmed by changing the jaw angle of both fingers while the gripping and pulling phenomenon is repeated according to the shape of the gripper's fingers.

Table 1 The specification of gripper unit

| Machine unit | Material/Standard |
|------------------|-------------------|
| Finger Jaw-A | S45C |
| Finger Jaw-B | S45C |
| Moving Plate | SS41 |
| Finger Cylinder | MHZ2-25D |
| Proximity Sensor | E3E-X3D1-M1GJ |
| Fiber Sensor | E32-T11R |
| Hinge Pin-1 | SHCMG5-18 |
| Hinge Pin-2 | MSB8-10 |
| Shock Absorber | GASB20-16 |
| Ball Bush | LMK30LUU |

3 Conclusions and Future works

In this study, an experiment was carried out for the purpose of linking the thin film product supply line and the robot as a study for the smooth supply and demand of parts and transportation. It was confirmed that the linkage operation was performed smoothly due to the change of the gripper's finger angle. In order to improve production efficiency and work environment in the future, we plan to improve the gripper so that it can fix and transfer flexible materials such as fibers and paper among thin-film products, and conduct research to develop a gripper moving unit.

References

- [1] 임정우, 조동혁, 이승엽 외 2명 (2017) 스마트팩토리 기반 제조공정 혁신에 관한 연구:전자제품 조립 제조라인에 적용 사례, 대한경영학회논문집, pp. 1609-1630
- [2] 공정리, 정명희, 김해지, (2020) 와셔공급용 그리퍼 무빙 유닛에 대한 연구, 한국기계가공학회 학술발표대회논문집, pp. 246-246
- [3] 공정리, (2019) 복잡한 형상의 소재를 이송하기 위한 갠트리 로더 시스템과 지그 개발에 대한 연구, 한국기계가공학회논문집, pp. 67-72

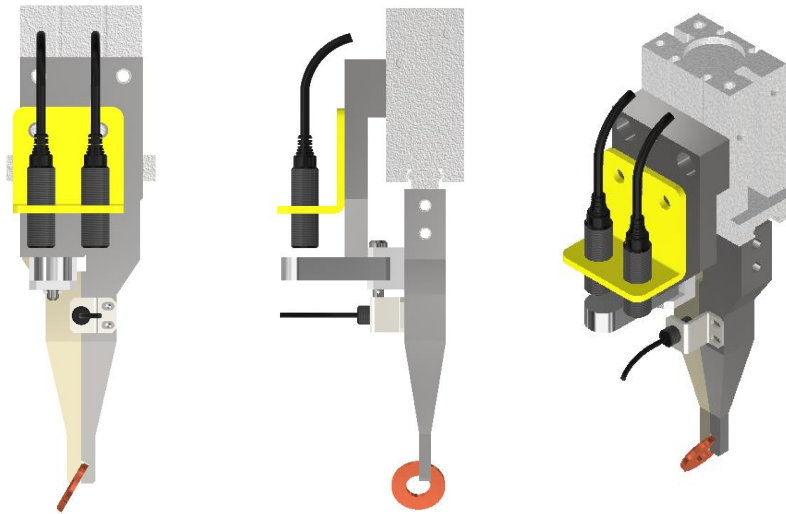


Fig. 1 3D Modeling of gripper

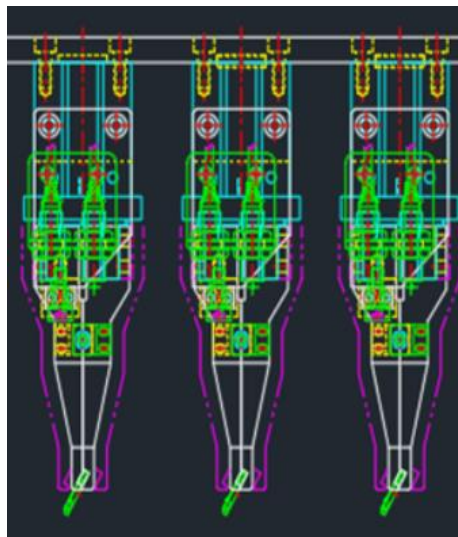


Fig. 2. 2D Design considering scalability

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | O |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Beauty Sniper : a Web-based Virtual Makeup System |
| Name | Hiecheol Kim |
| Affiliation | Daegu University |
| Address | 201 Daegu-daero, Jinryang-eup, Gyeongsan-si, Gyeongsangbuk-do, Korea |
| E-mail | hckim@daegu.ac.kr |

Beauty Sniper : a Web-based Virtual Makeup System

Young-Seok Noh, Seong-Cheol Park, Jun-Won Kim, Hiecheol Kim

AI Center in Daegu University
201 Daegu-daero, Jinryang-eup, Gyeongsan-si, Gyeongsangbuk-do

*Corresponding Author: hckim@daegu.ac.kr

Abstract - As the beauty industry flourishes, makeup is emerging as an important factor regardless of age or gender, and people are very interested in decorating their appearance. However, when you try to do makeup for the first time, you often don't know what color and style suits your face, and it is difficult to do makeup without a fixed style. This paper presents a web-based system designed to allow users to synthesize a makeup on a user's bare face image without having to make-up themselves. Given the user's bare face image, our system produces, using a GAN deep-learning model, a set of make-up images which can be referred before conducting real makeup. Moreover, our system integrated recent advanced web technologies including Flask web server, AddThis API, and Disqus API to enable its functionalities available over the web.

Keywords: AI, Systems, Computers, Web, Beauty, photo.

1 Introduction

As peoples tend to pursue external satisfaction day by day, the importance of makeup is increasing day by day in the life of men and women of all ages. The increase in domestic cosmetics production announced by the Ministry of Food and Drug Safety in Figure 1 reflects this trend well.



Figure 1. The trend of domestic cosmetics production by year

However, due to the nature of the environment, there are many people who do not have a chance to access makeup, and especially in Korea, men are not yet aware of makeup, so even if they try to pay attention to it, it is not easy to access.

This paper proposes a web-based virtual makeup system that makes it easier for such people to access makeup. This system can lower entry barriers to access to the makeup field by allowing you to check in advance what kind of a makeup style goes well with one's face. In addition, when purchasing cosmetics that match the makeup style that goes well with one's face, it can be helpful in determining the lake of cosmetics and the color of your lips. This system allows users to keep pace with the latest makeup trends anytime, anywhere, regardless of location or time.

Recently, deep learning technology is developing

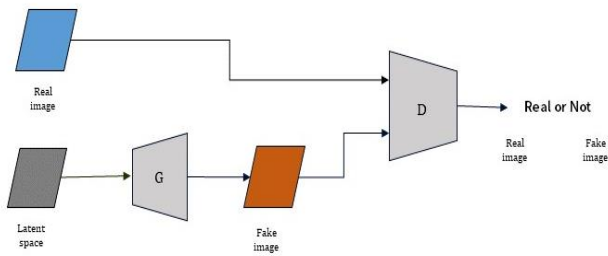
very rapidly [1-4]. Among these techniques, the GAN technology capable of learning and generating images was applied to the virtual makeup system [5]. In the implementation of our system, BeautyGAN[6] was used for the production of virtual makeup and all services were built on the web so that users could create and share images with makeup on their own photos.

This paper is organized as follows. Section 2 presents a brief overview of BeautyGAN. Section 3 explains the architecture of the proposed system and its implementation. Section 4 presents our virtual Markup System deployed over the web. Finally, the concluding remarks are provided in Section 5.

2 The BeautyGAN

The BeautyGAN is a GAN model [5]. GAN is an abbreviation for the Generative Adversarial Network and is the most commonly used model in the field of image generation among several deep neural network models [6]. While CNN, a basic deep learning model, is widely used in classification problems, GANs allow to generate images similar to training images. As shown in Figure 3, GAN consists of Generator and Discriminator which play a hostile role against each other in model training. Generator G learns the actual data distribution, and discriminator D distinguishes whether it is original data or fake data generated from G. The training process of constructor G can be seen as a process of minimax game between Generator and Discriminator, which increases the probability of deceiving discriminator D by generating highly elaborated fake images and increasing discriminator

D's capability to distinguish between the real and the fake



data.

Figure 3. GAN architecture

Unlike general GANs, BeautyGANs used in this system have one generator and two discriminators as shown in Figure 4. Two images are entered into the generator of the first stage: I_{src} is an image of a bare face without makeup, and I_{ref} is an image with makeup of the style you want. As a result of the generator, an image of applying I_{ref} 's makeup to I_{src} and an image of erasing the makeup from I_{ref} are obtained. Then, in the middle stage, a total of four images are created: the I_{ref} image, the I_{srcB} image with makeup applied to the src image, the I_{refA} image with makeup removed, and the I_{src} image. Based on the presence or absence of makeup, two images are paired and determined as Discriminators. Two makeup-applied images (I_{ref} Real Image + I_{srcB} Fake Image with Makeup) DA will be applied with two makeup-free images (I_{refA} Fake Image + I_{src} Real Image with Makeup removed). In the second stage of the Generator, the I_{srcB} image to which the makeup is applied and the I_{refA} image to which the makeup is removed are used as inputs. As a result, makeup is applied to the I_{refA} image without makeup, and the I_{srcB} image with makeup is obtained as a result. Through this process, the original image and the resulting image are compared with each other. Three loss values are used to determine performance, as shown in Figure 4. The first is cycle consistency loss comparing I_{src} images as a result of applying and erasing the makeup to the I_{src} image without makeup and the src image. The second is the perception loss that compares the I_{srcB} image with the I_{ref} image's makeup applied to the I_{src} image and the I_{refA} image without the makeup of the I_{ref} image. The last third is to compare the results of makeup segmentation such as face, eye, shadow, and lip in the I_{ref} image with the two images to which makeup was applied.

Table 1. shows the size of the actual learned dataset used for training BeautyGAN. Considering the relatively small size of training datasets, the performance of the model is very good. For more natural synthesis than the present, one can proceed with training with an additional makeup dataset. Due to lack of available dataset, in our system, the implementation of image synthesis is based on the original author's pre-trained model.

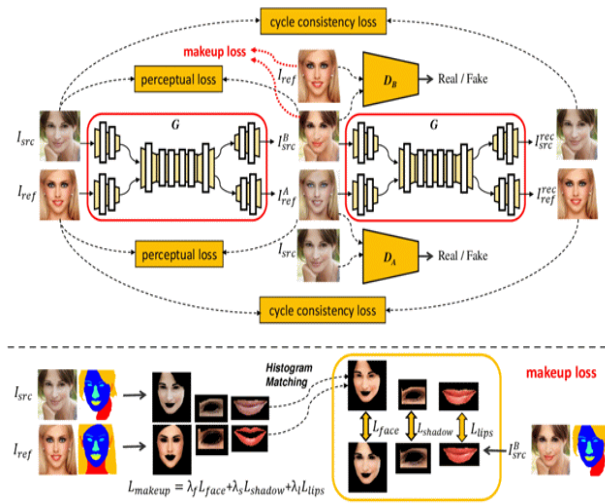


Figure 4. BeautyGAN architecture [6]

| | Dataset | Subjects | Images per subject | Total number of images |
|---|----------|----------|--------------------|------------------------|
| 1 | YMU | 151 | 4 | 604 |
| 2 | VMU | 51 | 4 | 204 |
| 3 | MIW | 125 | 1-2 | 154 |
| 4 | MIFS | 214 | 4 or 2 | 642 |
| 5 | Ours(MT) | 3000+ | 1-2 | 3834 |

Table 1. Dataset[6]

3 Implementation of the Proposed System

In this system, after creating an anaconda virtual environment and installing necessary packages, a web server was built using Flask which is one of the Python web frameworks available for a PC environment. Port forwarding was applied to access this web server from an external network. Therefore, if a Flask web server is running on the main PC, it is possible to access the server through an external network. For the functionality of makeup synthesis, a Tensorflow-based pre-trained BeautyGAN model [7] was used. Figure 2 shows a web server configuration diagram that performs port forwarding.

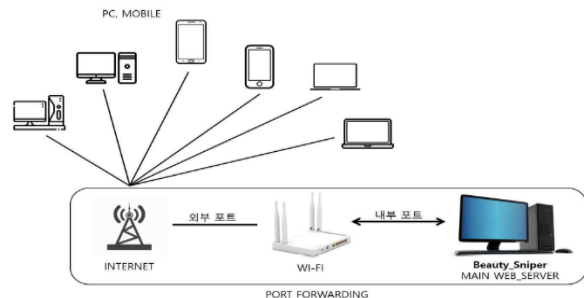


Figure 2. Web server diagram

Once we finish implementing our system, we have performed the makeup image synthesis for our test dataset. As we used the pre-trained BeautyGAN model, the synthesis works very reasonably. As discussed earlier, input to the system is two kinds : one is a bare face image without makeup and the other is a reference image with makeup. The resulting synthesis image is the makeup image of the bare face image augmented with the makeup part of the reference image. Figure 5 show the 3 examples. Each row in the figure represents an examples of which left two images corresponds respectively base face and makeup image, the right-most one is the synthesis image produced from the beautyGAN.



Figure 5. Model execution results

This system does not focus solely on synthesizing makeup and showing results, but implements a sharing function that allows users to share makeup-synthesized result images to other web or SNS spaces, and a comment window that allows users to enter comments on the web. The sharing function was implemented using AddThis API[8]. AddThis is a social bookmark service that can be integrated into a website using a web widget. The comment window was implemented using the Disqus API[9]. Disqus is an American blog comment hosting service for websites and online communities using network platforms.

4 The Production System

Figure 6 shows the Beauty Sniper web site built upon the implementation of the proposed system. This section presents the functionality of this virtual Makeup system.

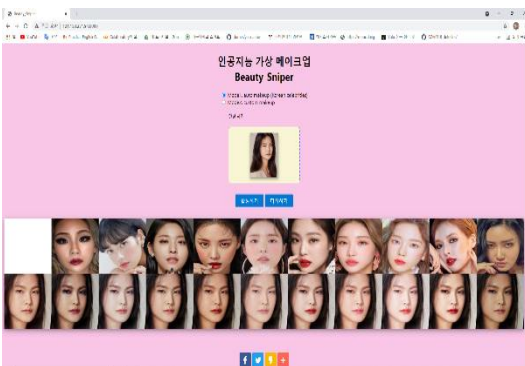


Figure 6. The main web screen

As a principle service, the Beauty Sniper system provides a makeup image synthesis over the web. Figure 7 shows the result screen in which the makeup synthesized image is output. Once uploading the bare face image and the makeup image, the synthesis proceeds by selecting the 'Execution' button. After finishing the synthesis, the synthesis makeup image are displayed along with the input bare face image for comparison

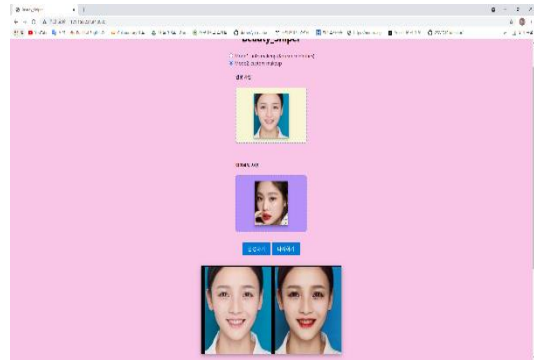


Figure 7. Synthetic result

In our system, users can share the synthesis image with others via SNS. Figure 8 shows an example use case. The SNS icons such as facebook and twitter are provided on the web screen. Additionally, by selecting the SNS icon at the bottm of the web, the web URL may be copied and shared.

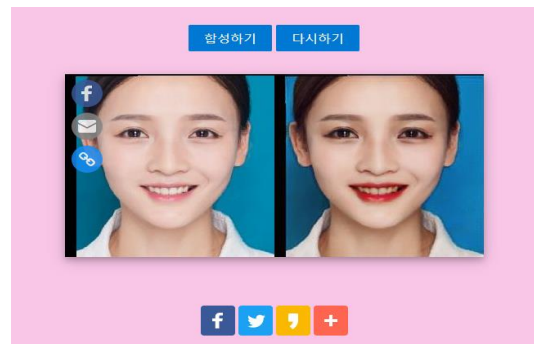


Figure 8. Sharing function

Finally, our system provides a space for placing some comments on some resulting image. Users can freely enter the their comments on a specific systhesis markup images to provide their reviews about the markup.

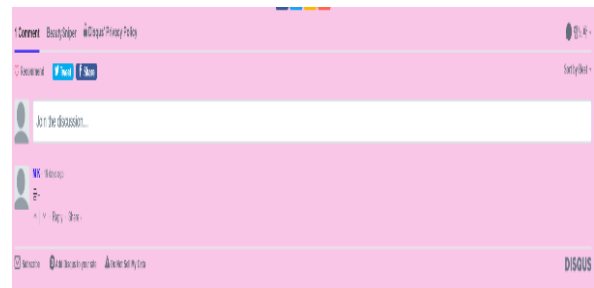


Figure 9. The space for comments

5 Conclusion

In this paper, we present a web-based virtual makeup system using the BeautyGAN object detection model. Given input bare face images, our implementation produces a makeup image synthesized using BeautyGAN. Based on web technologies such as Flask Web Server, AddThis API, and Disqus API, our system also provides some community functions which enable users to share and review the generated markup images over the web. In this research, the main focus was on implementing the initial prototype available. As a future research, we plan to build a virtual makeup system with improved quality by conducting through performance analysis and then improving the BeautyGAN model.

References

- [1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *Nature* 521.7553, pp. 436-444, 2015
- [2] A. Vaswani, et al., "Attention Is All You Need", Conference on Neural Information Processing Systems (NIPS), pp. 5998-6008, 2017
- [3] He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385*, 2015
- [4] S. Ren, et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *arXiv:1506.0149*, 2016
- [5] T. Li, et al., "BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network", Proceedings of the 26th ACM international conference on Multimedia, pp. 645-653, 2018
- [6] I. Goodfellow, et al., Generative Adversarial Nets, *Advances in Neural Information Processing Systems(NIPS)*, 27, 265-283, 2014
- [7] <https://github.com/Honlan/BeautyGAN>
- [8] <http://www.addthis.com>
- [9] <http://disqus.com>

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | O |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Professor |
| Name | Hiecheol Kim |
| Affiliation | Daegu University |
| Address | Jinryung Gyung-san Gyunbuk-do, Korea |
| E-mail | hckim@daegu.ackr |

A system for detecting safety helmets at construction sites

Jin-Wook Lee, Jun-Sik Kim, Min-ho Kwon, Hiecheol Kim

Affiliation

AI Center, Daegu University
201 Daegu-daero, Jinryang-eup, Gyeongsan-si, Gyeongsangbuk-do

<https://daegu.ac.kr/>

*Corresponding Author: hckim@daegu.ac.kr

Abstract – According to a nationwide inspection report of the risk of construction sites, the number of sites warned with danger was twice as many as that of sites without warning. Among them, workers' poor wearing of personal protective gear was the second most common in cadastral sites by the report. This paper implements a system that focuses on safety in construction sites by enforcing workers to wear helmet through automatic detection and alarming. The system automatically detects and judges in real time through cameras whether workers at construction sites wear safety helmets. Our implementation is based on a state-of-the-art object detection model, YOLOv5. To make our system more robust in ensuring safety, warning messages are repeatedly transmitted through a speaker when a worker does not wear the safety helmet. Providing the capability of detecting multiple safety helmets, the detection accuracy of our system reaches about 85% for our test data set.

Keywords: Safty, Helmet, Construction site, Object detection.

1 Introduction

The number of accidental deaths per 100,000 workers in the Korean construction industry was the highest among OECD member countries at 25.45, more than three times the OECD average of 8.29. Though the situation became more enhanced, the number of accidental deaths per 100,000 Korean industrial workers in 2020 was 3.61, still well above the average of 2.43 in 35 OECD member countries at the time, followed by Canada (5.84), Turkey (5.17·16), Chile (4.04) and Luxembourg (3.69) [1]. As such, it has long been pointed out that the safety of construction industry workers is not guaranteed facing with potential accidents in their everyday working. This paper proposes a construction site safety helmet detection system that aims at contributing to lowering the rate of accidents and deaths during labor by improving the insensitivity of workers in the construction industry.

In our approach, deep learning technology, which is rapidly developing in recent years, is applied [2-5]. In the field of safety helmet detection in construction site, several researches have been performed, S. Song and J. Kwak (2021) reported a safety helmet detection system using YOLOv5, and S. Park et al., (2019) presented R-Nfer, an object detection algorithm based on image data [6,7]. These researches only aim to detect the wearing of safety helmets, not providing any means to efficiently prevent accidents by making workers voluntarily wear safety helmets.

This paper implements a system that determines whether construction workers wear safety helmets using the YOLOv5[8-11], which is most commonly used among deep learning object detection models, and repeatedly sends separate alarm messages through speakers until they wear them.

In this paper we provide a brief overview of the YOLOv5 as the basis of our object detection backbone in Section 2. Then, we present the implementation details of the proposed system in Section 3, and finally provide concluding remarks in Section 4

2 Overview of YOLOv5

In our implementation, we used YOLOv5 for the functionality of object detection. YOLOv5 is one of the representative single-step object detection algorithm. It is a new version of YOLOv3 improved with respect to FPS(Frame Per seconds) and mAP (mean Average Precision). Unlike previous YOLO versions, yolov5 is divided into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x by size. As the size of the model increases, the accuracy increases, but the speed slows. For example, as shown in Figure 1, model 's' is somewhat less accurate than the other models, while benefiting highest speed among them. [10].

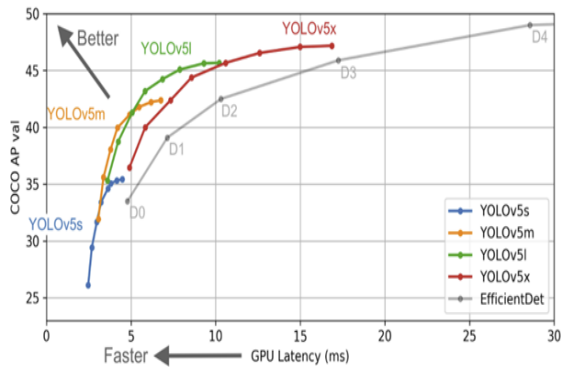


Figure 1. YOLOv5 Performance difference by model [10].

In the architecture, YOLOv5 is composed of the parts, backbone, PANet, and output, as shown in Figure 2.

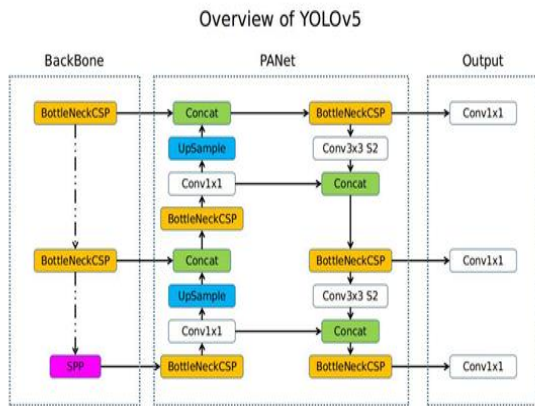


Figure 2. The overall structure of YOLOv5. [11]

Backbone is the part that extracts the feature map from the image. There are four types of Backbone, including the smallest YOLOv5_s to m, l, and x. In this paper, YOLOv5s was used. As shown in Figure 3, the backbone consists of four convolution layers and three pooling layers. Each convolution layer is coupled with a batch normalization layer, therefore undergoing a batch normalization process after performing a convolution operation. The Leaky ReLU is used as the activation function. The bottleneck module is a block with short-cut connection used in ResNet. The Bottleneck CSP module, which can be seen as the core, consists of four convolution layers. The SSP module is the Special Pyramid Pooling Layer used in YOLOv3-SPP, and features of 5*5, 9*9, and 13*13 were used as spatial bins. Finally, a one-dimensional arrangement fixed to a size of $5 + 9 + 13 = 27$ was created, allowing it to enter the input from the fully connected layer.

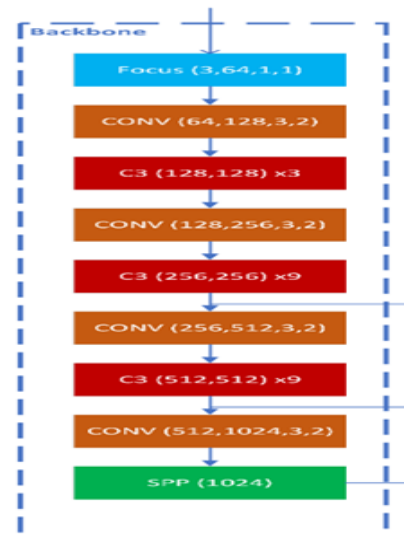


Figure 3. The structure of the backbone.[11]

The PANet network is to serve as a training model. As shown in Figure 4, it consists of four convolution layers and two upsampling layers, and the original data is created by decrypting the file reduced to Pooling through upsampling. Head is one of the configurations of PANet and consists of a series of Convolution, upsampling, concat, and pooling layers. Head is to find the location of an object based on the extracted feature map. After initially setting up the anchor box, there are a total of nine anchor boxes because they create final bounding boxes on a scale of a total of three, i.e., small objects with pixel information, intermediate objects with 16 pixel information, and large objects with 32 pixel information), and use three anchor boxes on each scale [11].

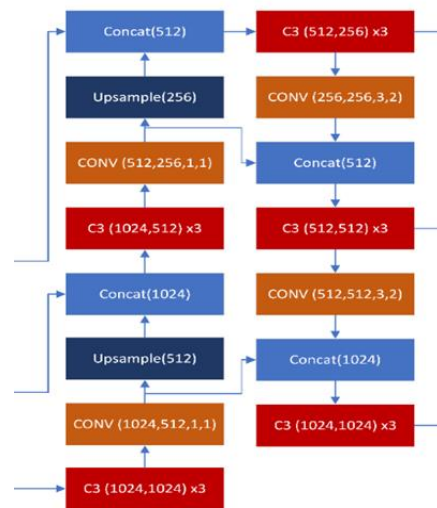


Figure 4. Structure of PANet Network[11]

3 Implementation of our System

3.1 Overview of the Implementation

The proposed system focused on implementing a system can be readily available at construction sites, and implemented a system using a desktop computer connected with a webcam and speakers that can be heard throughout the construction site. Training the deep learning model was carried out by Google Colab. Table 1 shows the HW components used in our implementation.



| HW | | Function |
|---|---------|----------------------|
|  | Webcam | Object Detection |
|  | Speaker | Voice Comment Output |

Table 1. HW in our implementation

Table 2 shows our SW development environment. All necessary packages are stored in "requirements.txt" to be automatically installed in the pip command.

| SWEnvironment | |
|--------------------------------------|---|
| Development environment | GoogleColab |
| Object detection model | YOLOv5s |
| Installation SW/Package Requirements | <ul style="list-style-type: none"> ▪ Python Package: requirements.txt ▪ Pytorch: 1.8.0 version or higher ▪ Python version 3.7 ▪ CUDA version 10.2 |

Table 2. SW composition.

3.2 Training the Model

In this paper, the dataset for training and testing used in our system consists of 10,000 images. As the training accuracy becomes higher as the size of the dataset becomes larger, we tried every effort to gather a reasonable size of datasets. We gathered the image from web and manually made bounding-box for each image. An example training image is shown in Figure 5. Two classes are used for labeling each training image. One of the labels indicates 'pass' which means wearing a safety helmet, and the other is 'fail' which means not wearing a safety helmet. The repository of our dataset containing images and labels is shown in Figure 6.



Figure 5. Example training image with bounding-box

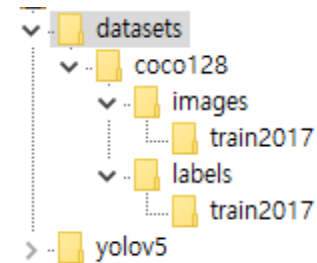


Figure 6. Designating a learning path.

Training the Yolov5 with our dataset was conducted in Google Colab for 100 epochs. As for the batch-size, if the value is too high, the memory capacity required for training becomes larger than the capacity provided by the system, which leads to a situation where the training task would fail, whereas if it is too small, the training job becomes very slow. Through explorative experiments, we choose the batch-size as 64.

For training our dataset, we execute train.py, a python program that trains the model for our custom training dataset over the pre-trained model in the yolov5 folder. When the execution is completed, a sub-folder named 'exp' is created in the run/train folder as shown in Figure 7. As a result of the training, a pt file containing the trained model is created in the exp folder.

Our implementation includes an alarming function. Over the detection of the safety helmet. It aims to repeatedly transmit these comments. The Python play sound library for Windows was used to reproduce the desired sound. The voice file used gtts (Google Text-to-Speech). gtts is a module made by Google, and there is also a method of using the Clova API provided by Naver. In this implementation, we generated an mp3 file saying, "Please wear a safety helmet" using gtts, and implemented the code to execute it when the recognition result is 'fail' [12].

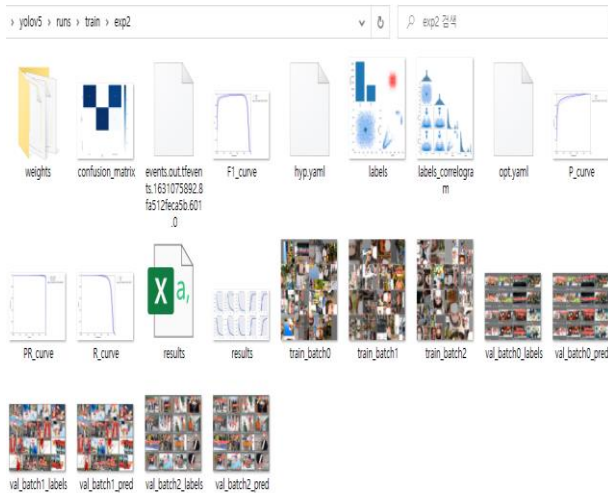


Figure 7. Data folder with trained image

As mentioned earlier, the trained dataset consists of 10,000 images with labeled with ‘pass’ and ‘fail’. Over the trained model, we performed inference for each image in the training dataset. As the final result, we get an average accuracy of 85%. Figure 8. shows an example of an inference result for an input image.



Figure 8. Case of execution result.

4 Conclusions

This paper implemented a system to determine whether workers in construction sites wear safety helmets using the YOLOv5 model, one of the deep learning models. We gathered and preprocessed 10,000 images from the web which is a reasonable size of datasets. To make our system suitable for robust safety management at private construction sites, alarming capability is also implemented. Overall, our system shows the detection accuracy of 85% which is a reasonable in the current datasets and YOLOv5 object detection algorithm. As future works, we plan to conduct a thorough performance evaluation for enhancing the object detection algorithm on

one hand. On the other hand, we plan to extend the coverage of objects subject to detection such as safety hats, vests, and shoes.

References

- [1] Newspaper web site, <https://www.yna.co.kr/view/AKR20210512080000502>.
- [2] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." Nature 521.7553, pp 436-444, 2015
- [3] A. Vaswani, et al., "Attention Is All You Need", Conference on Neural Information Processing Systems (NIPS), 2017
- [4] He, Kaiming, et al. "Deep residual learning for image recognition." arXiv preprint arXiv:1512.03385, 2015
- [5] S. Ren, et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", arXiv:1506.0149, 2016
- [6] S. Song and N. Kwak (2021), "Detection Of Safety Helmets Using YOLOv5" Proceedings of the conference by the Korean Contents Association, pp 483-484, Aug. 2021
- [7] S. Park, et al., "Image-Based Automatic Detection of Construction Helmets Using R-FCN and Transfer Learning", Journal of the Korean Society of Civil engineers, 39(3), 399-407, June 2019
- [8] J. Redmon, et al., "You Only Look Once: Unified, Real-Time Object Detection", arXiv:1506.02640v5, 2016
- [9] yolov5 github web site : <https://github.com/ultralytics/yolov5>
- [10] YOLOv5 – Introduction web site, <https://bigdata-analyst.tistory.com/194>
- [11] Overview of YOLOv5 model structure web site, <https://github.com/ultralytics/yolov5/issues/280>
- [12] Python gtts Module web site, <https://info-lab.tistory.com/234>

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---------------------------------|
| Title | Full Professor |
| Name | Kyuman Jeong |
| Affiliation | Daegu University / Korea |
| Address | Gyeongsan, Korea |
| E-mail | kyuman.jeong@gmail.com |

Audio Quality Restoration Method using Convolutional Neural Network

Hong-Jin Kim ¹⁾, Kyuman Jeong ²⁾

College of ICT, Daegu University, Gyeongsan, Korea

*Corresponding Author: kyuman.jeong@gmail.com

Abstract - Artificial intelligence technology, in which computers perform human-like actions or behaviors, is becoming popular. Particularly, efforts are being made to implement technologies that classify objects or respond to user behavior. It is also attracting attention in fields that require much time and effort, such as restoring paintings drawn in the past. It is expected that it can be used in various fields as well as an image restoration technique using three-dimensional data. In particular, audio data has changed from the way of using physical storage devices in the past to the way of being provided on a network basis. In this paper, we propose an algorithm to recover high - quality audio data from the internal storage device that can be self - produced by receiving compressed audio data. We propose a method of restoring audio data that is arranged and reproduced by changing time-dependent one-dimensional data using lossless audio data and lost audio data after compression through a deep learning technology, CNN (Convolutional Neural Network).

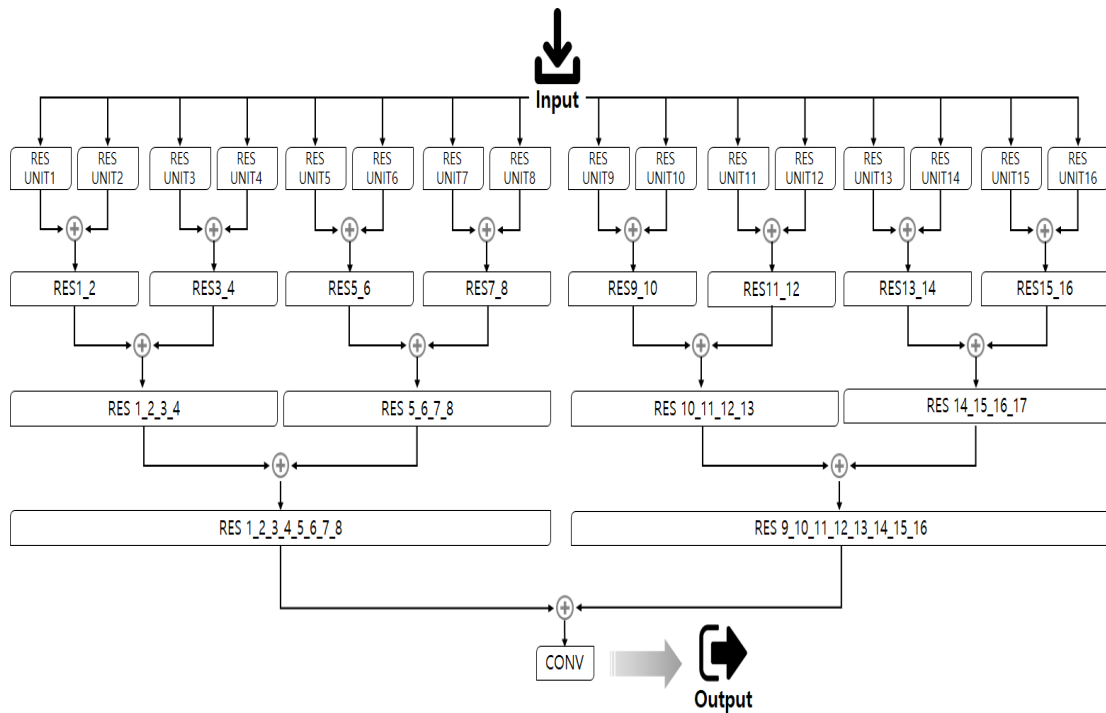


Figure 1. Overall Process

1 Introduction

Recording information contributes to the greatest contribution to the advancement of human technology. In the past, it was common to pass on to future generations by recording texts or pictures, but in modern times, sound and video can be recorded, and various information can be easily accessed. In particular, the act of recording sound rather than video has been developed so that it can be easily transmitted and miniaturized from phonographs to LPs, cassette tapes, CDs, and MP3s. The biggest reason that we were able to move from the material realm to the intangible is the recording method in a file format that can be easily transmitted. With the rapid rise of the Internet and smartphone penetration, the way data was used in the past using physical storage devices has changed. In particular, more data than a storage medium that a machine can hold, such as a network-based cloud service, can be provided through the Internet, and streaming services that receive video or sound data wirelessly are in the spotlight. In this paper, to solve the size and time constraints in Internet-based data transmission, we propose a technology that reduces the size of transmitted data and enables efficient audio streaming through self-restoration in the mechanical device responsible for physical storage and execution.

2 Proposed Method

In this paper, we study how to convert low-quality audio data into high-quality data using CNN used for media restoration. In the process of restoring sound quality using deep learning, WAV format data extracted from CD is converted into low-quality data and high-quality data, and used as input values and GT (GroundTruth) to be used for learning, respectively. Repeated learning We want to design an algorithm that restores audio data through The basis of the model used for learning is to compare and evaluate various models using CNN to select the model with the highest degree of completeness, and to introduce the process of designing a new model by modulating the model if necessary.

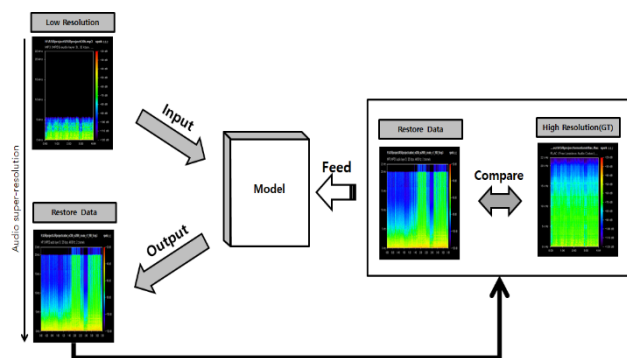


Figure 2 Audio Quality Restoration Algorithm

2.1 Data set generation

In the learning data generation process, 101 sound source CDs are used, and since each audio file has a different playback time, it is necessary to establish a unified standard. The data generation process is a total of two steps, extracts the audio data stored in the first CD, converts it into a lossless compression format Flac, which is the standard of the restoration point, and saves it. Next, in order to generate low-bitrate audio data to be used as input data, the Flac file extracted from the CD is down-sampled and generated. The number of extracted sound sources is a total of 1,137, and the compressed data must be decoded using the Pydub library that supports FFmpeg (Febrice Bellard et). Convert to format.

2.2 Learning model

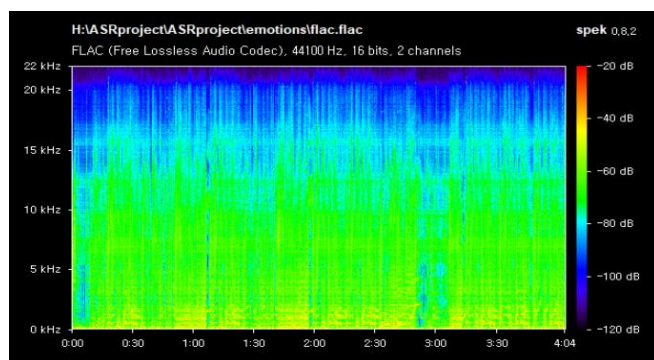
In this paper, we implement and implement a method to restore audio quality through extension by controlling the audio bandwidth using deep learning. The methods used for implementation were DRRN (Deep Recursive Residual Network) and SRResNet (Super-Resolution Residual Network) models used in Image Super-Resolution. Unlike image data, audio data stored in a one-dimensional array is used. Therefore, a new model is designed and proposed to have higher long-term accuracy, and it undergoes a process of comparative analysis with various previously presented CNN models.

Compared with ResNet and DRRN, our proposed model is shown in Figure 1, and RES_UNIT uses the Residual Unit structure of the DRRN algorithm.

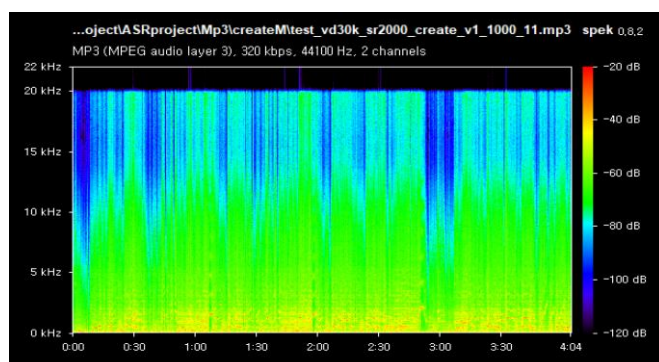
For the initial input, not a single residual unit, but multiple residual units are used to extract features or patterns of different data, and convolution is performed to the next layer by adding two equation pairs to the passed results, and the last one is passed through the residual unit. One result is passed through a single convolution map to derive the output.

2.3 Experimental results

After dividing the total 1,137 raw data by channel, it was studied using 4,512,753 datasets of High Resolution data and Low Resolution data, respectively, prepared by dividing into 2,000 sizes. The spectrum of the sound source to be used as the test set after learning is shown in Figure 3 below.



(a) Original sound



(b) Learning result

Figure 3. Spectrum for the original sound source (a) and the result obtained through training (b)

3 Conclusions

As models to be used in the training process, ResNet and DRRN, which are CNN models that are often used to restore high-resolution images from low-resolution, were used. When restoring the sound quality, it was confirmed that the model used in the existing Image Super-Resolution could not obtain good performance. As shown in Figure 34, although the high-pitched data is alive, there is a problem that it does not match the original spectrum. which is still different from In addition, noise generated during the restoration process can be identified. The time until the sound source derives the result through the model also takes about 3 to 5 minutes when using the current 3 minute sound source. Since audio has a lower data dimension than images, the proposed model uses a number of Residual Units at the initial input rather than the size or number of filters or the depth of the layer, and learns in pairs. It has been confirmed, and it is expected that better results can be obtained if the number of training times, the size and number of filters, and the size of the data to be learned are adjusted.

References

- [1] Ying Tai, Jian Yang and Xiaoming Liu. (2017). Image super-resolution via deep recursive residual network. In CVPR
- [2] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. (2016). Accurate image super-resolution using very deep convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition
- [4] V. H. Quintana and Edward J. Davison. (1974). Clipping-off gradient algorithms to compute optimal controls with constrained magnitude. International Journal of Control, vol. 20, no. 2, pp. 243-255
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems.
- [6] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. (2016). Deeply-recursive convolutional network for image super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- [7] Joan Bruna, Pablo Sprechmann and Yann LeCun. (2016). Super-resolution with deep convolutional sufficient statistics. In International Conference on Learning Representations (ICLR)
- [8] Volodymyr Kuleshov, S. Zayd Enam and Stefano Ermon. (2017). Audio super-resolution using neural nets. Presented at the 5th International Conference on Learning Representations (ICLR)

Intelligent Computational Edge: From Human/Machine Augmentation to Cognitive Engineering

Radmila Juric

University of South-Eastern Norway

Since the vision of The Computer of the 21st Century was debated in IEEE in 2002, rapid developments of ubiquitous and pervasive technologies have created a computational world where barriers between hardware and software are blurred and disappearing. Wireless and mobile technologies changed hand-held devices into powerful windows to a new era of client-server and cloud computing, and at the same time they provided undisputed computational power. Consequently, humans are being seamlessly integrated into the modern computational world by triggering computations at any time and any place. Since our infatuation with cloud computing started fading away, it was natural to start looking at computational edge and investigate how it can address various problems dictated by the nature of clouds and their efficiency. Edge, fog and dew computing followed, together with computational dust, and a proliferation of instances of Internet of Things and Internet of Everything have become our reality.

Developing intelligence at any computational *edge* is inevitable, considering our tendency of using predictive, learning and semantic technologies whenever possible. We claim that we are coming closer to the human level of reasoning because we run algorithms prevalent in the current world of artificial intelligence. However, how far are we from intelligent computational *edge*? There are three important aspects of computational *edge* which illustrate its promising future.

Firstly, human interaction with and contribution to the augmentation of semantics of and computations at the *edge* are important. We need to address numerous issues: from inevitable heterogeneities of everything and needs for interoperable hardware/software counterparts, to interpretation of semantics of data and the context in which we run computations. Human reasoning/intervention might be crucial in interpreting the meaning of data and algorithms at the *edge* and thus create intelligence. Second, we need to re-think the role of cloud computing. Should we complement it with human-centered computing at the *edge* because the *edge* may secure computing continuum? Clouds, edges, fog and dust may negotiate on the most suitable format of computational choices, but according to contexts in which we need them. However, it remains to be seen which role the *edge* will play in running predictive analytics and learning algorithms. Consequently, which types of devices should we develop for creating a real intelligent computational *edge*? Are we going to witness a convergence of human and artificial intelligence, interwoven with physical things?

This may bring us closer to augmented machines and humans, and augmenting cognition, all through *edge* computing. Cognitive engineering is then just around the corner. We hope to utilize *human cognition*, create modern human-machine systems, and harness *computational engineering*, where humans and machines seamlessly co-habit. Intelligent computational *edge* does have a role to play in all of them.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|------------------|---|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|-------------|-----------------------------------|
| Title | Regents Professor |
| Name | Dr. Sang C. Suh |
| Affiliation | Texas A&M University-Commerce |
| Address | P.O. Box 3011, Commerce, TX 75429 |
| E-mail | sang.suh@tamuc.edu |

Author

| | |
|-------------|-----------------------------------|
| Title | |
| Name | Amy W. Hays, MS |
| Affiliation | Texas A&M University-Commerce |
| Address | P.O. Box 3011, Commerce, TX 75429 |
| E-mail | ahays2@leomail.tamuc.edu |

Linear Regression Analysis of the Impact of COVID-19 Pandemic Health Interventions on Mental Health in the United States.

Amy W. Hays and Sang C. Suh

Texas A&M University-Commerce
P.O. Box 3011, Commerce TX 75429

ahays2@leomail.tamuc.edu and sang.suh@tamuc.edu

Abstract - Many government policies were put in place during the SARS-CoV-2 coronavirus (COVID-19) pandemic to help save lives. We examine the effects of lockdown, mask, entertainment, and education policies on the changes in the mental health of the United States population before, during, and after the policies were implemented during the pandemic. Linear regression models were used to analyze open-source datasets to determine if the policies or policy length affected the changes in the mental health of the U.S. population. The number of cases was also compared to mental health. None of the policies affected the percentage of adults who had symptoms of depression or anxiety over time based on the high p-values from the linear regressions. Although the length of the lockdowns did have a negative effect. We also found a linear relationship between the number of cases and mental health. Government policies mandates did little to ease the symptoms of depression or anxiety in the overall population. More demographic studies need to be conducted and increasing mental health resources should still be a government priority during a pandemic.

Keywords: policy, policies, mandates, COVID-19, SARS-CoV-2, coronavirus, mental health, depression, anxiety, lockdown, masks, linear regression.

1 Introduction

In 2020, the SARS-CoV-2 coronavirus (COVID-19) started a worldwide pandemic. Governments had to make public health choices on how they will help their citizens and economies survive. In the United States, there were many public health policies implemented by state and county officials that may have affected the physical, mental and economic health of their communities. Some of these policies included the use of masks, restrictions on gatherings, and what businesses could be open.

There are many articles regarding how the COVID-19 pandemic has had a negative impact on mental health around the world. One study did a meta-analysis that included 36 different studies with over 117,044 people 32% experienced stress, anxiety, or depression since the pandemic started [1]. The U.S. National Center for Health Statistics and the U.S. Census Bureau created the Household Pulse Survey to help track the changes in mental health in the population of the U.S. [2]. The 2019 National Interview Survey showed that 10.8% of adults over the age of 18 suffered from anxiety or depression [3]. During the pandemic, the Household Pulse Survey showed those numbers to reach as high as 42.6% [2].

In March of 2020, there was a study done on the impact of social distancing during the start of the pandemic that found the occurrence of mental health increased as social distancing was implemented and over time. They couldn't determine causality just a correlation

and suggested further research needed to be done [4]. In April of 2020, it was shown the one in four U.S. adults had serious mental distress [5].

In this study, we examine the impact of the different government policies on mental health by looking at the percentage of adults that have symptoms of depression or anxiety, before, during, and after the policies were mandated. The null hypothesis would be that there was no impact from government policies on the mental health of the population. We also examine how the length of time the policy was in place and see if there is any effect from the duration of the policies. The null hypothesis is that there was no impact from the length of the government policies on the mental health of the population. Then we examine the effect of the number of cases on mental health, with the null hypothesis also being no effect from the number of cases on mental health.

2 Proposed Method

2.1 Datasets

The U.S. National Center for Health Statistics and the U.S. Census Bureau created the Household Pulse Survey [2] that was used for the mental health data. The data covers the mental health of each state in the U.S. from April 23, 2020, through June 9, 2021. Each sample is taken and the responses were covered over the previous seven days. This time was referred to as the time period

statistic. The mental health values were percentages of the state adult population that were suffering from depression or anxiety. The questions used in the Household Pulse Survey were of a modified version of the two-item Patient Health Questionnaire (PHQ-2) and the two-item Generalized Anxiety Disorder (GAD-2) scale [2].

The mental health data was combined with the state policies data in the COVID-19 State and County Policy Orders Database from HealthData.gov [6] an official website of the United States Government. This data is a collection of data from 3 different other sources [7–9]. The data ranges from March 23, 2020, through June 1, 2021.

2.2 Policy Data Set-Up

The data had 63 different policies that we combined into 9 group policies. We examined 2 (lockdowns and masks) of these 9 (lockdown, masks, personal and religious, government, entertainment, health, education, business, and travel) group policies. The lockdown group policy consisted of the 'Shelter in Place', 'Nursing Homes', 'Quarantine', and 'Nursing Home Visitations' policies. The mask group policy only consisted of the 'Mask Requirement' policy.

Once this data was combined, we calculated how many days each state policy was enacted which is the policy duration statistic. We then calculated the policy action statistic by if the date of the mental health sample was before, during, or after that policy. Then we took the mean for each time period and then for each policy action statistic (before, during, and after the policy).

Another dataset was created to look at the pattern of cases vs mental health. This was done by taking the mean of all the mental health samples grouped by sample start date. Then, merged that data with the COVID-19 cases data from the CDC's COVID Data Tracker [10].

2.3 Analysis

Linear least-squares regression for two sets of measurements was chosen to see any basic trends. The p-value from the data during the policies was used. Based on the decision rule and in a 95% confidence level, a low p-value of < 0.05 would suggest we would reject our null hypothesis that the government policies did not have an impact on mental health. Thus indicating the policies did affect mental health.

The mental health percentages were compared for each of the policies and linear regression models were run for each policy action statistic (before, during, and after) and compared over the time period. The policy duration and mental health percentages were compared for each of the policies and linear regression models were run for when only the policy was active to see if we could determine if the length of the time the policy was active affected mental health. The final comparison was the mental health averaged by date compared to the daily new

COVID cases and plotted on a dual y-axis over time to show the overall pattern of mental health and if it reflects caseloads.

3 Results

3.1 Lockdowns

Over the time periods, the lockdowns had p-values of 0.189, 0.521, and 0.590 for before, during, and after the policies respectively (see Figure 1). Thus we cannot reject the null hypothesis. This suggests that over time the lockdown policy had very little effect on the mental health of the population.

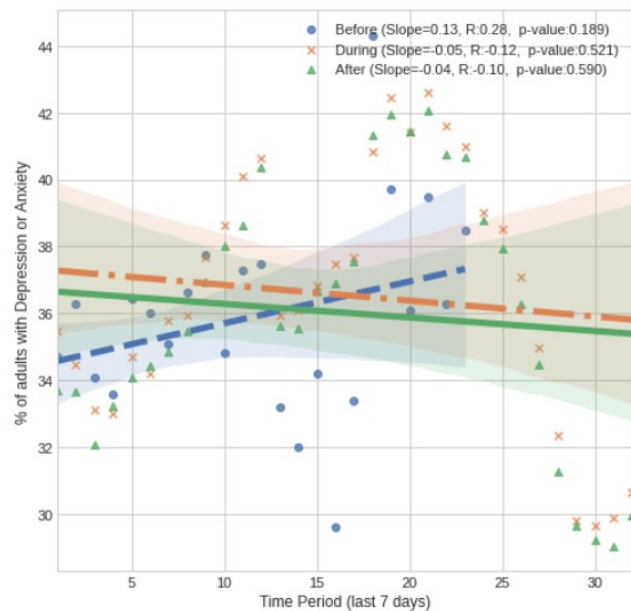


Figure 1. Percent of adults who have depression or anxiety over time before during and after lockdown policies in the U.S.

For the comparison of mental health over the length of lockdown policy was in place, the slope was 0.02 with an R-value of 0.56 and a p-value of 0.005 (see Figure 2). With this low p-value, we can reject the null hypothesis and show that there was a correlation of how long the lockdown policy was in place with increased symptoms of depression or anxiety. And the longer the policies were in place, the more percentage of adults suffering from depression or anxiety go up.

3.2 Masks

Over the time periods, the mask requirement mandates p-values of 0.670, 0.852, and 0.229 for before, during, and after the policies respectively (see Figure 3). With these high p-values, we cannot reject the null hypothesis. Suggesting over time the mask policies had very little effect on the mental health of the population.

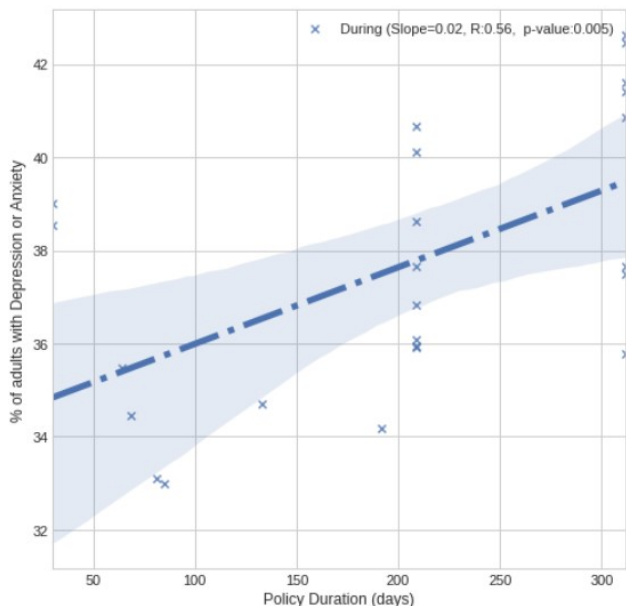


Figure 2. Percent of adults who have depression or anxiety over the length the lockdown policy was active

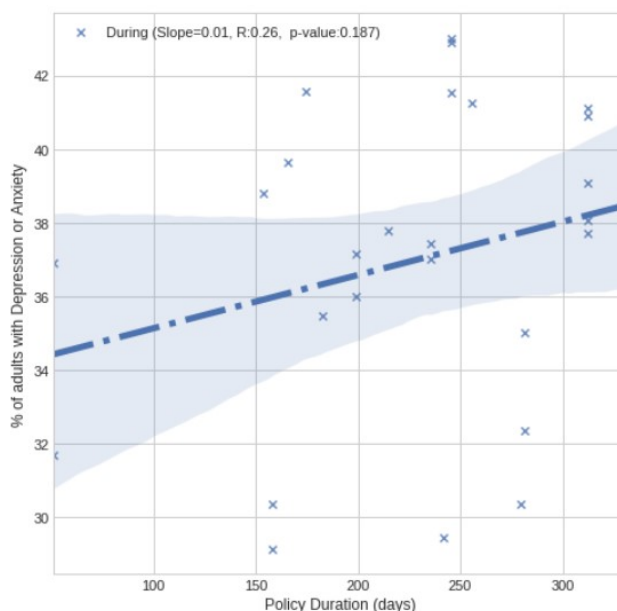


Figure 4. Percent of adults who have depression or anxiety over the length of the mask policy

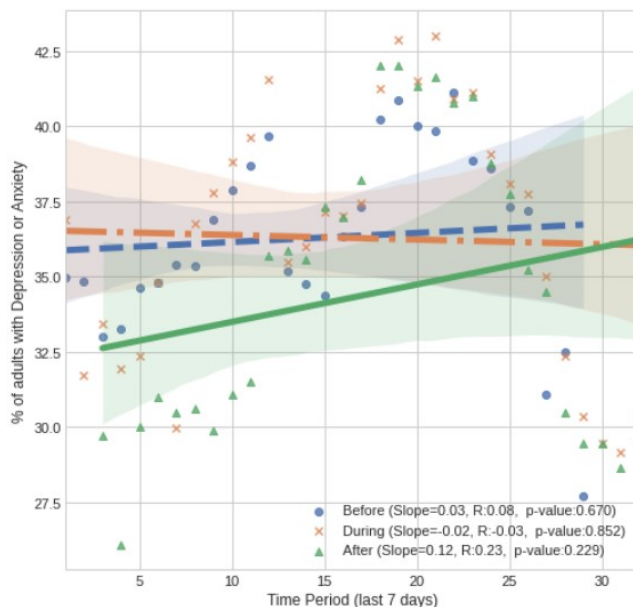


Figure 3. Percent of adults that have depression or anxiety overtime before, during, and after the mask policies

When comparing the percentage of adults with depression or anxiety over the length of the active mask requirement policies, the p-value was 0.187 (see Figure 4). The null hypothesis, in this case, cannot be rejected and suggests the length of the mask mandates had little effect on the mental health of the population

3.3 Mental Health Cases over Time

Because each of the graphs for all the policies showing the same pattern over time, we decided to plot the average of the percentage of adults having symptoms of depression or anxiety on the left y-axis and the number of daily new cases plotted on the right y-axis over time (see Figure 5). The patterns match very closely. We included the date of December 14, 2020, when the vaccines were made available and you can see that that single date had a very large impact on the mental health even as the caseloads rose for a short period after the vaccines came out and then dropped the first half of 2021.

So we took the mental health data and did a linear regression test over the new cases (see Figure 6) and the p-value was 0.000 suggesting that we could reject the null hypothesis that the number of cases did not have an effect and the graph shows a linear relationship. As the number of cases goes up so does the percentage of adults having symptoms of depression and anxiety.

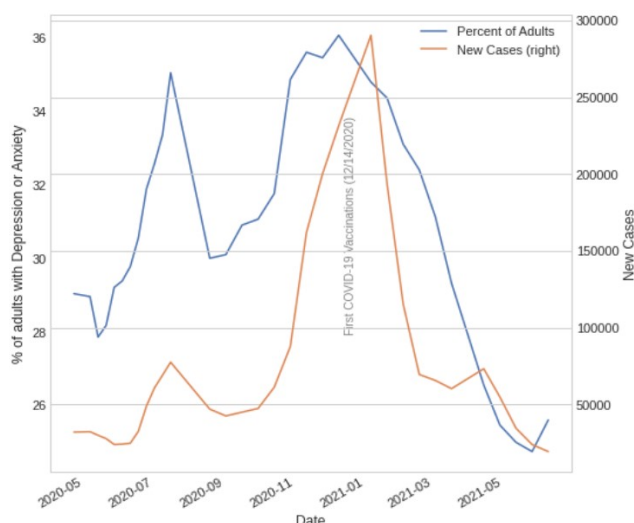


Figure 5. Percentage of adults with symptoms of depression or anxiety [2] (left y-axis) and new COVID-19 cases [10] (right y-axis) over time. The first COVID-9 vaccinations were started on December 14, 2020 [11], and annotated in the chart.

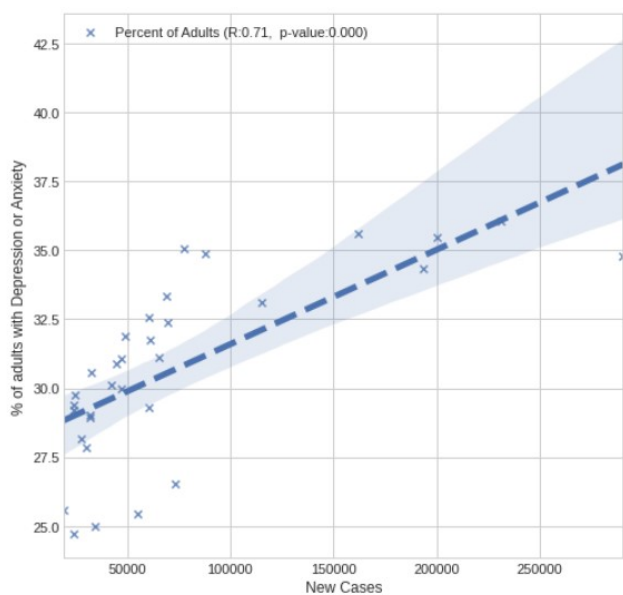


Figure 6. Percentage of adults with symptoms of depression or anxiety over the number of cases.

4 Discussion

4.1 Summary

Over the time period, while the policies were active, none of the policies had a low enough p-value to reject the null hypothesis that the policies made no difference on the percentage of adults who had symptoms of depression or anxiety. Thus proving the null hypothesis that there was no impact on the mental health of the populations by the government-mandated policies in response to COVID-19. For the length of the mask policies were active the null hypothesis, of no effect, was proven. Although in the case of the lockdowns, we could reject the null hypothesis and

we observed percentages of adults with symptoms of depression and anxiety increase the longer the policy was in place.

The largest impact on mental health was the number of cases over time that we saw in Figure 5. And then with the vaccine, mental health immediately started to improve along with the lowering caseloads in the first part of 2021. In Figure 6, we showed that the p-value of the linear regression between the percentage of adults having symptoms of depression or anxiety and the new cases overall was significant. And we could reject the null hypothesis and therefore proving that the number of cases made the percentages of adults having symptoms of depression or anxiety go higher.

4.2 Assumptions and Data Integrity

The assumption was made that the data in the open-source datasets were accurate. When we checked, we did find some errors with dates that caused us to remove some data. We also noticed some of the Arkansas state data was not updated when spot-checking with the Arkansas Department of Health COVID-19 Health Guidelines [12]. It is unknown if other state policies were missing. This study only looked at state data and not at the county policies. The reason for that choice is the mental health data was at the state level.

4.3 Recommendations and Future Research

An important fact to keep in mind when looking at these health statistics or any health statistic is that everyone is an individual and should be treated as such and not as a statistic. In that, consideration should be given by those who put these policies in place, and when making these policies their effects on the individual as well as the whole should be considered as best as possible. They should consider the mental side effects of the policies and ways to help mitigate those possible side effects. The pandemic itself over time had its effect on the percentage of adults that had signs of depression or anxiety that can be observed in all the data graphs as shown in Figures 5 and 6. The baseline from the year before indicated 10.8% [3] of adults showed signs of depression or anxiety, and that value increased to 42.6% [2] during the pandemic.

Many factors can lead to depression or anxiety. Some are environmental and some are biological. It was found that even COVID-19 can cause suicidal tendencies that are biologically induced [13]. We recommend that mental health resources be increased in response to the pandemic in general.

After mid-July of 2020, mental health seemed to improve for a short time even though the pandemic continued. We wonder if there was a small amount of time that some of these policies may have helped the mental health improve, but then the second wave started to pick up in mid-October and make the mental health numbers rise again overshadowing the possible positive impacts of

the policies. More detailed research would need to be done to determine these possible effects.

This study shows that all the policies had little effect on mental health overall. Except, over longer durations, the lockdown policies harmed mental health. Long periods of lockdown policies are not suggested. It has been shown that a specialized COVID intervention group helped students improve their mental health during the COVID-19 pandemic isolation in Spain [14]. With this knowledge, we recommended for special mental health programs be put into place when lockdown policies are enacted.

Even with these results, the amount and extent of these policies should be monitored closely for side effects in many different areas of concern, such as with mental health, physical health, and economic health of the society and of individual demographics. Some of these policies may have more side effects on some demographics over others. For example, the education policies will affect the families more than individuals or elderly adults. The children being at home instead of in school makes it harder for parents to work outside the home. The American Psychological Association found 3 in 5 parents experienced increased stress since the pandemic started [15]. The CDC has reported that the pandemic is affecting the mental health of young adults, Hispanic people, black people, unpaid caregivers for adults, essential workers, and people with preexisting mental health problems more than the other populations [16]. Further research would have to be done to assess the impacts of the policies on specific demographics.

Another concern with these policies, like the masks requirements, is that a false sense of security can be established causing more people to take risks they likely would not have taken. If the masks are not worn properly or cleaned properly they can infect someone or cause other respiratory issues. The most effective prevention from catching a contagious illness is not to be exposed at all. Even masks have only about an 80% protection rate [17].

5 Conclusions

We have shown that the lockdown, mask, entertainment, and education policies had little effect on the percent of adults having symptoms of depression or anxiety over the time of the pandemic. But the length of the lockdowns showed to have a negative effect the longer the policy was in place. The number of new cases had a direct significant impact on the percentage of adults having symptoms of depression or anxiety. More research should be done concerning the effectiveness of these policies in controlling the spread of COVID-19, along with the economic impacts of these policies. We also suggest looking into how government policies regarding COVID-19 affect different demographics as it would be helpful to evaluate the effectiveness of these policies.

Acknowledgments

Special thanks go to Cece Gassner, J.D. Vice President for Research, Texas A&M University-Commerce for her constant support for research.

References

- [1] N. Salari et al., "Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis," *Globalization and Health*, vol. 16, pp. 1–11, 2020, doi:<https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-020-00589-w>.
- [2] CDC - NCHS - National Center for Health Statistics, "Mental Health - Household Pulse Survey - COVID-19," Anxiety and Depression, Household Pulse Survey, Feb. 22, 2021. <https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm> (accessed Jun. 30, 2021).
- [3] Terlizzi EP and Schiller JS, "Estimates of Mental Health Symptomatology, by Month of Interview: United States, 2019," National Center for Health Statistics, Mar. 2021.
- [4] B. Marroquín, V. Vine, and R. Morgan, "Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources," *Psychiatry Research*, vol. 293, p. 113419, Nov. 2020, doi: 10.1016/j.psychres.2020.113419.
- [5] J. M. Twenge and T. E. Joiner, "Mental distress among U.S. adults during the COVID-19 pandemic," *Journal of Clinical Psychology*, vol. 76, no. 12, pp. 2170–2182, 2020, doi: 10.1002/jclp.23064.
- [6] HealthData.gov Team, "COVID-19 State and County Policy Orders | HealthData.gov," COVID-19 State and County Policy Orders | Healthdata.gov, Jul. 13, 2021. <https://healthdata.gov/dataset/COVID-19-State-and-County-Policy-Orders/gyqz-9u7n> (accessed Jul. 13, 2021).
- [7] J. Raifman, D. Jones, J. Bor, S. Lipson, J. Jay, and P. Chan, "COVID-19 US state policy database (CUSP)," Google Docs, 2020. <https://github.com/USCOVIDpolicy/COVID-19-US-State-Policy-Database> (accessed Jul. 18, 2021).
- [8] "Stay-at-home order issued in the United States in response to the 2020 coronavirus pandemic." <https://www.wikidata.org/wiki/Q88509703> (accessed Jul. 18, 2021).
- [9] "Virtual Student Federal Service," 2020. <https://vsfs.state.gov/> (accessed Jul. 01, 2020).
- [10] CDC, "COVID Data Tracker," Centers for Disease Control and Prevention, Mar. 28, 2020. <https://covid.cdc.gov/covid-data-tracker> (accessed Jul. 23, 2021).
- [11] A. S. for P. Affairs (ASPA), "COVID-19 Vaccine Distribution: The Process," HHS.gov, Dec. 16, 2020. <https://www.hhs.gov/coronavirus/covid-19-vaccines/distribution/index.html> (accessed Jul. 23, 2021).
- [12] "COVID-19 Health Guidances Arkansas Department of Health." <https://www.healthy.arkansas.gov/programs->

services/topics/covid-19-health-guidances (accessed Jul. 19, 2021).

[13] D. Banerjee, J. R. Kosagisharaf, and T. S. Sathyanarayana Rao, "'The dual pandemic' of suicide and COVID-19: A biopsychosocial narrative of risks and prevention," *Psychiatry Res*, vol. 295, p. 113577, Jan. 2021, doi: 10.1016/j.psychres.2020.113577.

[14] S. Gorbeña et al., "The effects of an intervention to improve mental health during the COVID-19 quarantine: comparison with a COVID control group, and a pre-COVID intervention group," *Psychology & Health*, vol. 0, no. 0, pp. 1–16, Jun. 2021, doi: 10.1080/08870446.2021.1936520.

[15] "Stress in America: One Year Later, A New Wave of Pandemic Health Concerns: (502832021-001)." American Psychological Association, 2021. doi: 10.1037/e502832021-001.

[16] M. É. Czeisler et al., "Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic — United States, June 24–30, 2020," *MMWR Morb Mortal Wkly Rep*, vol. 69, 2020, doi: 10.15585/mmwr.mm6932a1.

[17] Y. Wang, Z. Deng, and D. Shi, "How effective is a mask in preventing COVID-19 infection?," *MEDICAL DEVICES & SENSORS*, vol. 4, no. 1, p. e10163, 2021, doi: 10.1002/mds3.10163.

A Machine Learning Approach to Zero-day Detection

Jinoh Kim

Computer Science Department
Texas A&M University
Commerce, TX 75428, USA

Jinoh.Kim@tamuc.edu

Abstract – Zero-day vulnerabilities are unknown, and thus, have no readily available patches and remedies to apply to defend the resources to be protected. Therefore, existing approaches based on known patterns may not be effective enough to detect newly emerging threats targeting zero-day vulnerabilities. For example, signature-based detection relies on common byte sequences gathered from known malicious activities, thus showing fairly high performance to detect, which has no idea about new classes of attacks, however. With a recent remarkable advance of machine learning (ML) technologies, numerous ML-based detection techniques have been investigated, but a vast majority of such techniques still assume the availability of existing patterns to construct their learning model (i.e., supervised learning), which is thus also limited to known attacks only. To identify zero-day exploits, one possible approach is semi-supervised learning, which relies on the profiling of benign activities without requiring existing attack samples. This talk introduces two candidate approaches for realizing semi-supervised learning: (1) one-class classification and (2) autoencoding architecture. This talk will also discuss the potential and challenges of the two approaches with an initial evaluation study in the context of malware detection as a use-case study.

Keywords: Zero-day, Security, Profiling, Semi-supervised learning, Machine learning

System Design for Biomedical and Rehabilitation Systems

Prof. Jeonghee Kim

Texas A&M University
Department of Engineering Technology and Industrial Distribution
Fermier Hall 008A | 106 Ross Street | College Station, TX 77843-3367

jeonghee.kim@tamu.edu

Abstract - Many of the assistive technologies and treatment systems currently available for movement disorders are not easy to use due to side effects, lack of efficacy, invasiveness, and bulkiness of the system. In addition, diseases are often misdiagnosed in their early stages or inefficiently treated due to the lack of efficient quantification tools. The goal of my research is to improve access to medical service for people with neurological disorders and physical disabilities, such as Parkinson's disease, dementia, stroke, and spinal cord injuries and to provide efficient information to medical personnel. I aim to develop a wearable health monitoring systems and non-invasive neuromodulation by conducting Ability-based design (1) to quantitatively *monitor/characterize* their performance and (2) to *augment/modulate* their limited performance using a smart optimization algorithm. In this seminar, I will introduce my current research projects including the tongue-based assistive technology and its applications, the tremor monitor/diagnosis system for longitudinal home-based system, and the real-time neuromodulation for tremulous behaviors. And then, I will discuss how to approach the pathological behavior using wearable, mobile and non-invasive technologies to improve the quality of life.

Keywords: Ability-based Design, Biomedical System, Neuroengineering, Quantitative Analysis, Rehabilitation System.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | X |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | |
| Name | Umit Ogras |
| Affiliation | University of Wisconsin – Madison |
| Address | 1415 Engineering Dr, Madison, WI 53706, USA |
| E-mail | uogras@wisc.edu |

Efficient Hyperdimensional Computing using Monte Carlo Tree Search

Toygun Basaklar¹, Suat Gumussoy² and Umit Ogras¹

¹University of Wisconsin – Madison, ²USA, Siemens Research

Abstract - Hyperdimensional computing (HDC) has emerged as a new light-weight learning algorithm with smaller computation and energy requirements compared to conventional techniques. In HDC, data points are represented by high dimensional vectors (hypervectors), which are mapped to high dimensional space (hyperspace). Typically, a large hypervector dimension (≥ 1000) is required to achieve accuracies comparable to conventional alternatives. However, unnecessarily large hypervectors increase hardware and energy costs, which can undermine their benefits. This paper presents a novel search algorithm to reduce the hypervector dimension while maintaining the accuracy and improving the robustness of the classifier. To this end, we employ Monte Carlo Tree Search (MCTS). The proposed approach decreases the hypervector dimension by $128 \times$ while maintaining or increasing the accuracy achieved by conventional HDC.

Keywords: Wearable IoT Devices, Hyperdimensional Computing, Monte Carlo Tree Search.

1 Introduction

The emergence of IoT devices [1, 2] and the reduction of their size and cost [3] leads to a rapidly growing subfield of IoT, namely wearable IoT [4]. Wearable IoT devices fuse data from multiple sensors, such as inertial measurement units (IMU) and biopotential amplifiers, to achieve accurate real-time tracking. This enables a variety of health applications such as remote health monitoring, gesture recognition, human activity recognition etc. [5, 6, 7]. However, size and weight constraints of wearable IoT devices limit the battery capacity [7, 8]. This necessitates the device to operate under a tight energy budget ($\sim \mu W$). However, the intensity of multi-modal sensor data is very high to operate under such resource-scarce device. Cloud computing is not a suitable option for health applications since offloading the data to the cloud increases the communication energy [9] and raises privacy concerns [10]. Hence the need to locally process the sensor data, incorporating machine learning (ML) algorithms on edge devices to alleviate these concerns [11]. However, robust ML approaches such as deep neural networks and other sophisticated algorithms are not viable *due to the limited computational and energy capacity of wearable edge devices*. Therefore, there is a strong need for *computationally light* learning algorithms that can provide high accuracy, real-time inference, and robustness to noisy sensor data.

Brain-inspired hyperdimensional computing (HDC) [12] is an emerging computing paradigm that supports fast, robust, and low-power learning and provides competitive accuracy to state-of-the-art machine learning algorithms for various health applications, such as human activity recognition [13], bio-signal processing [14], and gesture

recognition [15] along with other domains of applications such as language processing, speech recognition, and DNA sequencing. HDC is based on a theoretical model of memory and cognition in the human brain [12]. The human brain operates on the information from the large size of neuronal interactions which includes the interaction of billions of neurons. Motivated by this, HDC models the *association* of a sensory input with the human memory [12, 13] using high dimensional representations of the data points. Instead of using bit-wise representations of the actual input data points' values, HDC maps data points in the original input space to a *random* hyperdimensional vectors, called *sample hypervectors*. This high-dimensional space generally consists of thousands of dimensions. To train the HDC classifiers, the sample hypervectors that have the same labels are combined linearly to obtain ensemble class hypervectors, called *class encoders*. In HDC, the class encoders act as an associative memory of the human brain, constructed from the neuronal representations of the sensory inputs to the brain which are represented as sample hypervectors. The new sensory input to the brain is represented as a *query hypervector* (Q) which is generated in the same way as the sample hypervectors. A cognitive task (inference) associates the new sensory input with the associative memory (class encoders) by finding the most similar class encoder to (Q), generally using cosine similarity or the Hamming distance.

HDC provides hardware friendly binary operations which can be processed efficiently on resource constrained wearable IoT devices. The conventional hardware platforms can leverage the nature of the simple vector arithmetic on binary/non-binary hypervectors used in HDC. A recent study [16] shows that custom hardware implementations of HDC can further improve the energy

efficiency and inference speed while reducing the design complexity. However, the core structure of HDC is based on the existence of billions of nearly orthogonal hypervectors that can be combined to represent different features in the original input space [17]. In order to have huge number of nearly orthogonal hypervectors, with high probability, a sufficiently large dimension is needed [16, 18]. Due to the *random mapping* process, the decrease in dimension leads to a decrease in probability of random hypervectors being orthogonal which in turn decreases the overall inference accuracy [19]. This adds a *redundancy* to the algorithm as the number of bit-wise operation increases with the increase in dimensionality which implies unnecessary use of computational resources. Therefore, there is a critical need to eliminate the redundancy arising from the random mapping process of HDC such that *problem specific mapping* can be achieved to increase the accuracy, robustness and efficiency of HDC with *smaller dimensions*.

This paper presents a novel search algorithm utilizing the reinforcement learning concepts to optimize the mapping process of the input data points in the high-dimensional space. As a result of the mapping process, sample hypervectors of different classes should be spread as far as possible in the high-dimensional space whereas sample hypervectors of the same class should be clustered and placed close to each other [20]. Using this geometrical interpretation of hypervectors and a novel non-uniform quantization approach [20], we employ Monte Carlo tree search (MCTS) [21] to achieve increased accuracy in smaller dimensions. Since each node in the tree may have huge number of children (e.g. at least 7^{21}), rather than using a traditional MCTS approach, a novel search approach is followed while traversing through the tree to *optimize the design of hypervectors* to achieve lower energy and computational resources while increasing the accuracy. The proposed approach is evaluated using two representative health applications where wearable IoT devices can be easily put to use: human activity recognition, and fetal state diagnosis.

2 Background and Overview

This section presents the general architecture and the operations both in HDC and Monte Carlo Tree Search (MCTS) algorithms, as illustrated in Figure 1.

2.1 Hyperdimensional Computing (HDC) Overview

HDC Training: The first step in HDC is to encode the input sample points to high-dimensional space. This is done by following the steps ① Quantization and mapping and ② Construction of sample hypervectors as illustrated in

Figure 1. Assume that there are S training samples and each training sample x_s for $s \in \{1, \dots, S\}$ is an N -dimensional vector $x_s = \{x_s^1, x_s^2, \dots, x_s^N\}$, where $x_s^n \in \mathbb{R}$ is the value of feature f_n in this sample and $f_n \in \mathbb{R}$ corresponds to the n^{th} feature in the feature space F . The encoding starts with quantizing the feature values into M uniform levels using the minimum (f_n^{\min}) and the maximum (f_n^{\max}) values of each feature obtained from the training set. To map the input samples to high-dimensional space, first each quantization level is assigned to a *level hypervector* $L \in \{-1, 1\}^D$ or $L \in \{0, 1\}^D$ where D is the dimensionality of the high-dimensional space and is typically ≥ 1000 . The minimum quantization level of each feature f_n^{\min} is assigned a random bipolar/binary D -dimensional hypervector L_n^1 . To generate each consecutive hypervector L_n^2 to L_n^M , $b = \frac{D}{2(M-1)}$ bits are randomly flipped. This operation is denoted by $L_n^{m+1} = g(L_n^m, b) \forall m \in \{1, \dots, M\}$. Using the level hypervectors and input samples, we first decide the quantization levels for each feature of the sample point x_s . $q(x_s^n)$ returns the quantization level $m \in \{1, \dots, M\}$ that contains the value of x_s for feature f_n . Then, we simply add the corresponding level hypervectors $L_n^{q(x_s^n)}$ to obtain sample hypervectors X_s as follows:

$$X_s = \sum_{n=1}^N L_n^{q(x_s^n)} \forall s \in \{1, 2, \dots, S\} \text{ where } X_s \in \mathbb{Z}^D$$

After mapping the input sample points to high-dimensional space, training is performed by simply summing up all sample hypervectors (X_s) that have same class label (y_s for $k \in \{1, \dots, K\}$) to obtain class encoders $E_k \in \mathbb{Z}^D$ as shown in ③ Classifier encoding in Figure 1.

HDC Inference: During inference, the query samples with unknown class labels are first mapped to the high-dimensional space using the previously described encoding procedure. The mapped hypervectors are called *query hypervectors* Q . Then, a similarity metric between class encoders and the query hypervectors is used to predict the class label. Cosine similarity is the most common metric used for similarity check while Hamming distance can also be used for binary hypervectors. The class k with the highest similarity is decided as the class label of the query data point.

A detailed explanation of HDC procedures with a toy example is provided in [20].

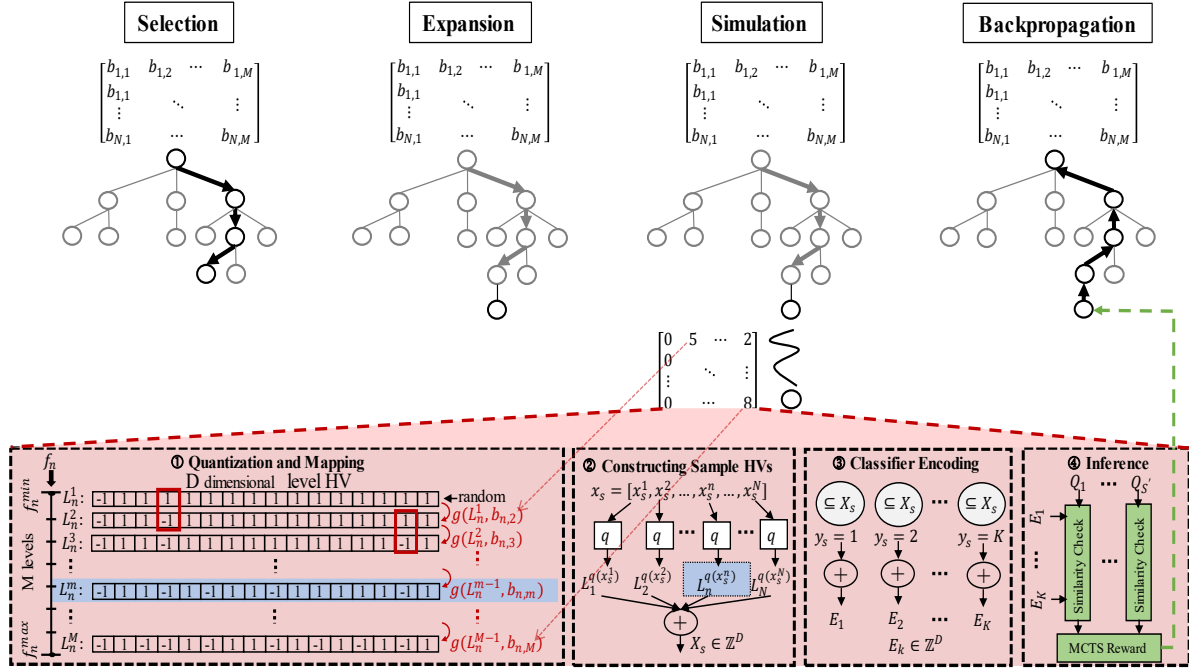


Figure 1 - Overview of HDC and MCTS Framework together. Bit values are chosen arbitrarily.

2.2 Monte Carlo Tree Search (MCTS) Overview

MCTS is a best-first tree search algorithm that combines the generality of Monte Carlo simulations based on a certain policy with the precision of a tree search [22, 23]. The algorithm iteratively builds a search tree with nodes representing the states of the given domain and the edges/links between the nodes represents the actions leading from one state to another. The algorithm starts the search at a root node which can be an empty state or a current state depending on the problem definition, and it repeats the following four steps at every iteration as shown in Figure 1.

Selection: MCTS traverses the current tree starting from the root node (v_0) using a tree policy until it reaches to a leaf node. Each node in the tree stores its own information and it is updated through backpropagation which is the final step of the algorithm. At each node, $Q(v)$, sum of all simulation rewards passed through that node and $T(v)$, number of times this node has been traversed, are stored. Using this information, the tree policy should balance the exploration-exploitation trade-off within the tree. To this end, the most commonly used tree policy, Upper Confidence Bound formula applied to trees (UCT) [24] is also adopted in this work. The UCT for the child node v_i of a parent node v is calculated as follows:

$$UCT(v_i, v) = \frac{Q(v_i)}{V(v_i)} + C \sqrt{\frac{\log(T(v))}{T(v_i)}}$$

where $C > 0$ is an exploration constant. First component of the above formula is the exploitation component, and it favors the nodes that have high average reward. The second

component on the other hand is the exploration component and it favors the nodes that have a smaller number of visits.

Expansion: If the selected node is not a terminal state and an expandable node (i.e. There are still children nodes that are not yet explored.), a child node is added to expand the tree from a set of legal actions that can be taken from the selected node. This new node's information is set to 0.

Simulation: A user-specified number of simulations based on either a uniformly sampled random legal actions or a heuristic are run from the new node until it reaches a terminal state.

Backpropagation: An average reward is calculated for the newly added node, and it is backpropagated through the selected path to the root node where each node's information is updated by simply adding the new node information to their own information.

The algorithm terminates after it reaches a user specified constraint (e.g. time constraint, memory constraint, and iteration constraint) or some other user specified condition.

3 Hypervector Design using MCTS Framework

3.1 Problem Overview

The mapping process should satisfy two conditions: (i) the hypervectors of the same class should form a cluster and (ii) different class clusters should be spread as far as possible in the high-dimensional space [20]. However, random mapping process suggested by the baseline HDC implementation does not satisfy these conditions. In baseline HDC implementation, a random hypervector is assigned to the first level of a feature and a uniform number of bits to flip between consecutive levels is decided as $b =$

$\frac{D}{2^{(M-1)}}$. At the end, it produces $N \times M$ different *level hypervectors* $L_n^m \forall n, m$ that represent the quantized features in the hyperspace. Let $B_{N \times M}$ is a matrix which includes all b_n^m values for N features and M levels. In baseline HDC implementation, all elements of this matrix equal to $\frac{D}{2^{(M-1)}}$ except the first column which is $\mathbf{0}$ since a random hypervector is assigned to the first level of each feature. In this work, a Monte Carlo Tree Search is employed to obtain an optimized B matrix. This is done by allowing the tree to choose from variable number of bits $b_n^m \forall n \in \{1, \dots, N\}$ and $m \in \{2, \dots, M\}$ while traversing through the tree and ensuring the orthogonality constraint [18] given as follows:

$$\sum_{m=1}^M b_n^m \leq \frac{D}{2} \text{ where } n \in \{1, \dots, N\}$$

We define this problem as a sequential action selection process in which the actions are increasing the b_n^m values by a fixed amount for only one level for each feature. The states in the tree correspond to different B matrices. Tree starts with a root node of an empty B matrix. It then explores the search space based on the tree policy and Monte Carlo simulations as shown in Figure 1. During these simulations, random actions are taken following the same tree policy until the simulations hits the terminal state. Using the B at the terminal state, HDC training and inference is done all over again using only the training set ($Q_s = X_s$ in Figure 1). We also employ a single-pass retraining approach to enhance the quality of class encoders. Single-pass retraining does an inference using the training dataset and updates the class encoders if there happens a misclassification for a given training sample hypervector as follows:

$$\begin{aligned} E_k &= E_k + X_s \\ E'_k &= E'_k - X_s \end{aligned}$$

where X_s has a class label $y_s = k$ but predicted as k' . It adds the misclassified sample hypervector to E_k to increase the similarity between X_s and E_k and subtracts from E'_k to decrease the similarity between X_s and E'_k . After the single-pass retraining, inference is done using the updated class encoders and training dataset. The reward obtained from the simulations is defined as:

$$MCTS_{reward} = R - R_{baseline}$$

where R represents the macro-averaged recall obtained using the new B matrix, and $R_{baseline}$ represents the macro-averaged recall obtained using the baseline HDC implementation with $D > 1000$. We use macro-averaged recall in our implementation since the total accuracy calculation suffers from imbalanced datasets. At every node, n_{MCTS} simulations run in parallel, and an average reward is returned as the expected reward of that node. In

our framework, we choose $n_{MCTS} = 100$. After the search terminates, a B matrix with variable $b_n^m \sim \forall n \in \{1, \dots, N\}, m \in \{2, \dots, M\}$ values are obtained. This new mapping yields increased accuracy compared to the baseline HDC implementation using one order of magnitude reduction in dimensionality.

3.2 Constructing the Search Tree

The tree search starts with expanding the root node to its children. This node has state v_0 which corresponds to an $N \times M$ matrix whose elements are zero. The terminal node has a state B where $\sum_{m=1}^M b_n^m = \frac{D}{2}$ where $n \in \{1, \dots, N\}$. During the search, at every branch, the algorithm can only increase the number of bits to flip ($b_n^m \forall n \in \{1, \dots, N\}, m = m'$) for a single level of each feature by an amount of γ . The remaining levels for a feature maintain their own values ($b_n^m \forall n \in \{1, \dots, N\}, m \neq m'$). The action space is huge if all features are considered together at the same branch of the tree. It becomes $(M-1)^N$ since $b_n^m \forall n, m = 1$ are all zero. To cope with the huge action space problem, we propose a novel search approach where we sequentially explore the feature space and thus reduce both the breadth and depth of the tree. Algorithm 1 describes the proposed search approach.

Sequential Feature Exploration: At every branch, we only consider a certain number of features, N' where $N' \ll N$. At the first branch of the tree, there are $(M-1)^{N'}$ children nodes arising from the root node. At every consecutive branch, the algorithm explores an additional N' features. Based on our preliminary sweep studies of this parameter ($N' \in \{1, 2, 3, 4\}$), we decided to choose $N' = 2$ considering the runtime, memory requirements, and obtained accuracy. For example, let total number of features in a dataset is $N = 4$ and number of quantization levels is $M = 4$. At the first branch, the algorithm only explores the first two features where it can increase the $b_n^m \forall n \in \{1, 2\}$ by an amount of γ where $\gamma < \frac{D}{2}$. At the second branch, the algorithm explores the third and fourth features. At the third branch, since the orthogonality condition is not yet satisfied, tree goes back to explore the first two features by increasing the $b_n^m \forall n \in \{1, 2\}$ by an additional amount of γ . The tree search basically follows this trend until it terminates due to a user specified constraint as illustrated in Figure 2. The proposed sequential feature exploration significantly reduces the action space at each branch as it becomes $(M-1)^2$.

Coarse to Finer Search: The search algorithm starts with a coarse γ value and as the tree goes deeper this γ value becomes finer in order to obtain a solution closer to an optimal one. Each node in the tree stores two additional information namely (i) *FeatureNo* which shows the features to be explored at the next branch and (ii) *ExpLevel* which shows how many times each feature is explored

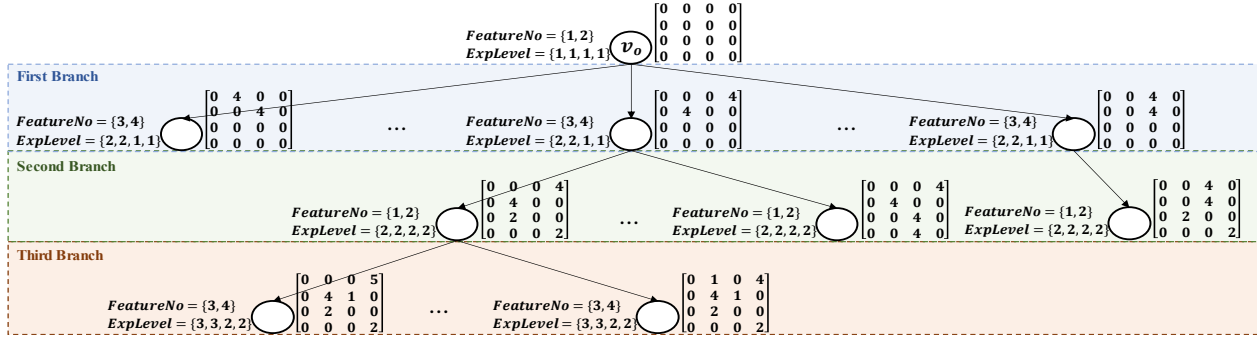


Figure 2 - Toy example for proposed novel search tree construction approach

during the tree search. Let us consider the same example where the total number of features is $N = 4$, number of quantization levels is $M = 4$, and the dimensionality is $D = 16$. $FeatureNo$ and $ExpLevel$ are $\{1,2\}$ and $\{1,1,1,1\}$ for the root node respectively as shown in Figure 2. As the features are explored, their corresponding $ExpLevel$ values are updated accordingly. For each $FeatureNo$, γ is updated as follows:

$$\gamma(FeatureNo) = \frac{D}{2 \times \alpha \times ExpLevel(FeatureNo)}$$

where α is an arbitrary constant that decides the initial γ value and is decided as $\alpha = 2$ in this work. Instead of using a fixed γ value, the proposed coarse to finer search approach significantly reduces the depth of the tree. However, the algorithm may not explore more optimal solutions due to the large γ value at initial branches. Choosing α as a large value increases the overall runtime of the algorithm while reduces the risk of missing the optimal solution during the search.

Root Switching: The obtained reward from Monte Carlo simulations is generally in the range of $[-0.15, 0.15]$. Since the search approach that we employ follows a sequential feature exploration, the effect of choosing different actions for insignificant features in the dataset may have little effect on the overall reward which itself is small. Due to the exploration constant, C , in the UCT formula, the algorithm may waste its time to explore insignificant nodes. Therefore, in contrast to traditional MCTS frameworks in the literature where the rewards are either -1 or 1 and $C = \sqrt{2}$, we choose $C = 0.001$. In addition to that, to guide the tree to search the space with new features or with finer γ values, we specify root switching conditions where the tree switch the root node to the current best node obtained so far. Current best node is decided as traversing through the current tree using the UCT formula with $C = 0$. Each node also stores the B_{MCTS} matrix which corresponds to a terminal state that gives the maximum reward among 100 simulations. Each iteration, the algorithm also finds current best node v_{best} , current best B_{MCTS} , and the best value $\frac{Q(v_{best})}{T(v_{best})}$ obtained so far. If current best node is the same for $\beta \times (M - 1)^2$ iterations, the algorithm switches the root node to current best node if B_{MCTS} can be reached from the current best node. If B_{MCTS} cannot be reached from the

current best node and the best reward (since there can be multiple nodes yielding the same reward) is the same for another $\beta \times (M - 1)^2$ iterations, the algorithm switches the root node to current best node. We choose $\beta = 4$ in our implementation.

Algorithm 1: Proposed MCTS Approach.

```

1: function MCTS ( $C, \beta, \alpha$ )
2: create root node  $v_0$  with empty  $B$  matrix
3: while user specified constraint do
4:  $v_i \leftarrow v_0$ 
5: while  $v_i$  is not a terminal node do
6:   if is not fully expanded then
7:     choose  $a \in v_i.getLegalActions(\alpha)$ 
8:      $v_i \leftarrow v_i.expand(a)$ 
9:   else
10:     $v_i \leftarrow v_i.bestChild(C)$ 
11:  end if
12: end while
13: [ $reward, B_{MCTS}$ ]  $\leftarrow v_i.Simulation()$ 
14:  $v_i.backPropagate()$ 
15:  $v_i \leftarrow v_0$ 
16:  $v_{best} \leftarrow v_i.bestChild(C = 0)$ 
17: best value  $\leftarrow \frac{Q(v_{best})}{T(v_{best})}$ 
18: if  $v_{best}$  is same for  $\beta \times (M - 1)^2 B_{MCTS}$  is reachable then
19:    $v_0 \leftarrow v_{best}$ 
20: else if best value is same for  $2 \times \beta \times (M - 1)^2$  then
21:    $v_0 \leftarrow v_{best}$ 
22: end if
23: if  $v_0$  is a terminal node then
24:   break
25: end if
26: end while

```

Preliminary Results

We perform our evaluations on two publicly available representative wearable health applications: human activity recognition [25], and cardiocography [26].

Human activity recognition (HAR) dataset provides stretch sensor and accelerometer readings of 22 subjects while they are performing eight activities: jump, lie down, sit, stand, walk, stairs up, stairs down, transition. It provides 120 comprehensive features extracted from the raw data. We choose a subset of 33 features using sequential feature selection. The data is divided randomly into 80% training and 20% test set. 33 features are extracted from 5148 accelerometer and stretch sensor readings. **Cardiocography (CTG) dataset** provides 21 features from 2126 fetal cardiocograms, which are extracted to diagnose the fetal state. The fetal state is divided into three classes: normal, suspect, and pathological. We randomly chose 80% of the data as the training set and 20% as the test set.

We compare this work with our previous approach, where an evolutionary algorithm is used to reduce the dimensionality of HDC approach, and with the baseline HDC in terms of macro-recall and total accuracy. We first apply the baseline HDC using $D = 8192$, which is shown to be sufficiently large. Then, our proposed approach is used to minimize the number of dimensions while maintaining or increasing the accuracy of the approach. As a comparison point in lower dimensions, we also compare the proposed approach with our previous work [20]. Table 1 shows the performance comparison of these approaches. The baseline HDC achieves 80% total accuracy with $D = 1024$ dimension and 87.95% total accuracy with $D = 8192$ dimension for the CTG dataset and HAR dataset respectively. Our approach reduces the number of dimensions to $D = 32$ (*i.e.* by $32 \times$) and to $D = 64$ (*i.e.* by $128 \times$) for the same datasets while increasing the accuracy and recall rates. Our proposed approach further increases the performance of the HDC approach compared to our previous work.

Table 1 - Performance comparison of different approaches

| | CTG | HAR |
|--|--|--|
| Baseline HDC (D=1024, 8192) | Macro-Recall: 76.58% Total Acc.: 80% | Macro-Recall: 84.53% Total Acc.: 87.95% |
| Prev. Work [20] (D=32, 64) | Macro-Recall: 81.33% Total Acc.: 82.88% | Macro-Recall: 85.71% Total Acc.: 86.89% |
| Current Work (D=32, 64) | Macro-Recall: 84.02% Total Acc.: 88.71% | Macro-Recall: 88.94% Total Acc.: 90.36% |

4 References

- [1] M. Capra, R. Peloso, G. Masera, M. Ruo Roch et M. Martina, «Edge computing: A survey on the hardware requirements in the internet of things world.» *Future Internet*, vol. 11, p. 100, 2019.
- [2] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan et X. Chen, «Convergence of edge computing and deep learning: A comprehensive survey.» *IEEE Communications Surveys & Tutorials*, vol. 22, p. 869–904, 2020.
- [3] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini et I. De Munari, «IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment.» *IEEE Internet of Things Journal*, vol. 6, p. 8553–8562, 2019.
- [4] S. Bhatt, F. Patwa et R. Sandhu, «An access control framework for cloud-enabled wearable Internet of Things.» chez *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, 2017.
- [5] S. Hiremath, G. Yang et K. Mankodiya, «Wearable Internet of Things: Concept, architectural components and promises for person-centered healthcare.» chez *2014 4th MOBIHEALTH*, 2014.
- [6] G. Bhat, N. Tran, H. Shill et U. Y. Ogras, «w-HAR: An Activity Recognition Dataset and Framework Using Low-Power Wearable Devices.» *Sensors*, vol. 20, p. 5356, 2020.
- [7] J. Park, G. Bhat, A. Nk, C. S. Geyik, U. Y. Ogras et H. G. Lee, «Energy per operation optimization for energy-harvesting wearable IoT devices.» *Sensors*, vol. 20, p. 764, 2020.
- [8] A. L. S. de Lima et others, «Feasibility of Large-Scale Deployment of Multiple Wearable Sensors in Parkinson's Disease.» *PLOS One*, vol. 12, p. e0189161, 2017.
- [9] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan et Y. Zhang, «Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks.» *IEEE access*, vol. 4, p. 5896–5907, 2016.
- [10] A. Ozanne et others, «Wearables in Epilepsy and Parkinson's disease-A Focus Group Study.» *Acta Neurologica Scandinavica*, vol. 137, p. 188–194, 2018.
- [11] M. Merenda, C. Porcaro et D. Iero, «Edge Machine Learning for AI-enabled IoT devices: a review.» *Sensors*, vol. 20, p. 2533, 2020.
- [12] P. Kanerva, «Hyperdimensional Computing: An Introduction To Computing In Distributed Representation With High-dimensional Random Vectors.» *Cognitive computation*, vol. 1, p. 139–159, 2009.
- [13] Y. Kim, M. Imani et T. S. Rosing, «Efficient Human Activity Recognition Using Hyperdimensional Computing.» chez *Proceedings of the 8th International Conference on the Internet of Things*, 2018.
- [14] A. Rahimi, P. Kanerva, L. Benini et J. M. Rabaey, «Efficient Biosignal Processing using Hyperdimensional Computing: Network templates for Combined Learning and Classification of ExG Signals.» *Proceedings of the IEEE*, vol. 107, p. 123–143, 2018.
- [15] A. Moin, A. Zhou, A. Rahimi, A. Menon, S. Benatti, G. Alexandrov, S. Tamakloe, J. Ting, N. Yamamoto, Y. Khan et others, «A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition.» *Nature Electronics*, p. 1–10, 2020.
- [16] S. Datta, R. A. G. Antonio, A. R. S. Ison et J. M. Rabaey, «A Programmable Hyper-Dimensional Processor Architecture for Human-Centric IoT.» *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, p. 439–452, 2019.
- [17] M. Imani, J. Messerly, F. Wu, W. Pi et T. Rosing, «A binary learning framework for hyperdimensional computing.» chez *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2019.
- [18] L. Ge et K. K. Parhi, «Classification using hyperdimensional computing: A review.» *IEEE Circuits and Systems Magazine*, vol. 20, p. 30–47, 2020.
- [19] M. Imani, C. Huang, D. Kong et T. Rosing, «Hierarchical hyperdimensional computing for energy efficient classification.» chez *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, 2018.
- [20] T. Basaklar, Y. Tuncel, S. Y. Narayana, S. Gumussoy et U. Y. Ogras, «Hypervector Design for Efficient Hyperdimensional Computing on Edge Devices.» *arXiv preprint arXiv:2103.06709*, 2021.
- [21] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis et S. Colton, «A survey of monte carlo tree search methods.» *IEEE Transactions on Computational Intelligence and AI in games*, vol. 4, p. 1–43, 2012.
- [22] H. Baier et P. I. Cowling, «Evolutionary mcts for multi-action adversarial games.» chez *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 2018.
- [23] T. Anthony, Z. Tian et D. Barber, «Thinking fast and slow with deep learning and tree search.» *arXiv preprint arXiv:1705.08439*, 2017.
- [24] L. Kocsis et C. Szepesvári, «Bandit based monte-carlo planning.» chez *European conference on machine learning*, 2006.
- [25] G. Bhat, R. Deb et U. Y. Ogras, «OpenHealth: Open Source Platform for Wearable Health Monitoring.» *IEEE Design & Test*, 2019.
- [26] J. P. Marques de Sa, J. Bernardes et D. Ayres de Campos, *Cardiotocography Data Set*, 2010.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | X |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | |
| Name | Ganapati Bhat |
| Affiliation | Washington State University |
| Address | 335 NE Spokane St, EME 506, Pullman, WA 99164 |
| E-mail | ganapati.bhat@wsu.edu |

Enabling Robust Human Activity Recognition through Generative Adversarial Networks

Dina Hussein and Ganapati Bhat

School of Electrical Engineering and Computer Science, Washington State University, Pullman WA

Abstract – Human activity recognition (HAR) using wearable devices is widely used in health monitoring, rehabilitation, and movement disorders since knowing what a person is doing at any time is critical to decide the appropriate therapy for the patients. HAR algorithms typically collect data in controlled environments and train a classifier for identifying activities at runtime. The training step generally assumes that, during runtime, the sensors are working perfectly with no missing or incorrect samples. However, real-world usage of wearable sensors introduces a number of uncertainties. For instance, due to limited communication bandwidth or limited battery energy, samples may get dropped. The dropped samples, in turn, lead to a significant reduction in the accuracy. To this end, we explore Generative Adversarial Imputation Networks (GAIN) to recover the missing samples to enable robust HAR. Specifically, if we detect any missing samples, we use GAIN to recover the samples before classifying the activities. In this paper, we perform a design space exploration of various GAIN structures to analyze the trade-off between accuracy and runtime overhead. Overall, our results show that using GAIN to recover missing samples enables 94% classification accuracy with as much as 20% missing samples.

Keywords: Wearable IoT Devices, Human Activity Recognition, Generative Adversarial Networks.

1 Introduction

Human activity recognition (HAR) has increased in popularity due to the recent development of low-power wearable devices that integrate multiple sensors, microprocessors, and communication capabilities [1–3]. HAR algorithms are being used for fitness monitoring, rehabilitation, and in knowing the activities of movement disorder patients [4, 5].

HAR algorithm development and validation typically involve four crucial steps: training data collection and labeling, activity segmentation, classifier training, and testing on new users [1, 6, 7]. The data collection is done in a controlled environment where the developers ensure that the sensors are in perfect working condition, and they are mounted in a known position. The collected data is then divided into distinct segments and labeled by one or more experts. The labeled data is passed through a feature generation algorithm to obtain features that distinguish between the activities of interest. The features and labels are input to a supervised learning algorithm to train an activity classifier. Finally, the trained classifier is used at runtime to identify the activities of one or more new users.

State-of-the-art HAR algorithms in the literature generally assume that all the sensor data is available perfectly without any missing samples. However, in real-world usage, the wearable devices can encounter scenarios where there are missing samples either due to user error, sensor malfunction, or energy limitations. The missing data, in turn, leads to a significant drop in the accuracy. For

instance, the accuracy drops from 95% to 50% with only 10% of the samples missing. Therefore, there is a strong need to develop robust HAR classifiers that provide high accuracy in the presence of missing data.

One of the common approaches to develop robust classifiers is to include examples with missing samples during training. However, including missing samples during training does not yield significant benefits in accuracy. For instance, including examples with 10% missing data in training increases the accuracy to only about 65%. Furthermore, including missing samples during training does not account for all possible scenarios at runtime. Therefore, in this paper, we develop a runtime approach to recover missing samples in HAR. Specifically, we leverage recently proposed generative adversarial imputation networks (GAIN) to recover missing data at runtime [8]. GAIN is an adaptation of the general generative adversarial networks that is well suited for recovering missing samples in time series data [9]. Starting with a baseline GAIN with a large number of weights, we perform a design space exploration of various GAIN structures to analyze the trade-off between accuracy and overhead. This is important because wearable devices are generally constrained by their limited batteries. As such it is critical to obtain highest accuracy while minimizing the execution time and energy. After recovering the missing samples with GAIN, the robust HAR classifier generates the features and identifies the activity.

We validate the robust HAR classifier using w-HAR, a publicly available HAR dataset [10]. The dataset does not

include any missing data by default. Therefore, we first use missing data in the dataset and then use GAIN to recover the missing data. We also implement the GAIN algorithm on the TI CC2652 microcontroller to measure the energy overhead [11]. Overall, the experiments show that GAIN is able to effectively recover missing data and enable accuracy activity classification.

2 HAR Background

As detailed in Section 1, robust HAR involves segmentation, feature generation and classification. In this section we provide a brief summary of each step in HAR.

Segmentation and Data Recovery: Sensors used in HAR produce streaming data at a pre-configured sampling frequency. The streaming data must be divided into distinct activity segments so that each segment contains a single activity. The segmentations step achieves this by utilizing variable-length or fixed-length segmentation algorithms [1, 6, 7]. Next, the data recovery algorithm checks for any missing samples and recovers them if it detects missing samples.

Feature Generation: The activity segments are processed by the feature generation block to calculate the features required for classification. In this paper, we re-use the features provided in the w-HAR dataset [10].

Classification: The features are finally used to identify the activity being performed by the user. Any supervised learning algorithm can be used to identify the activities. In this work, we utilize the neural network structure proposed in the w-HAR dataset.

3 Generative Adversarial Imputation Networks

Generative adversarial networks (GAN) [9] have been used recently to obtain synthetic data for a number of applications including image recognition, speech recognition, and natural language processing [12–14]. GANs are useful in generating completely new data examples, but they are not well-suited to recover missing data when partial information is available. To address this limitation, the work in [8] proposed a variation of GAN called as GAIN. Following the general structure of GANs, GAIN consists of a generator and a discriminator. The goal of the generator in GAIN is to accurately recover missing data by utilizing the available samples while the job of the discriminator is to distinguish between recovered synthetic data and observed data. The generator is trained to minimize the error between the generated data and the actual observed data. By minimizing this error, the generator can generate accurate samples for instances where the observed data is not available. The generator is also trained to maximize the misclassification error in the discriminator so that the discriminator is unable to distinguish between synthetic and observed data.

When fully trained, the GAIN generator is able to accurately recover the missing samples and the

discriminator is unable to distinguish the synthetic data. This means that the data generated by GAIN follows the distribution of the actual data. Next, we illustrate the accuracy gains achieved by GAIN when the sensor data as missing samples.

4 Experiments and Results

4.1 Experimental Setup

Wearable device: We use the TI-CC2652R as the primary processor for our wearable device model [11]. The TI-CC2652R processor consists of ARM Cortex processor which does the segmentation, data recovery, feature generation, and classification. The robust HAR classifier is implemented on the TI-CC2652 processor to measure the overhead.

HAR Dataset: We use the w-HAR dataset to perform our experiments. The dataset provides stretch sensor and accelerometer readings for 22 users while they are performing eight activities: jump, lie down, sit, stand, walk, stairs up, stairs down, and transition.

Missing data implementation: We introduce missing data for each 3 second interval of the w-HAR dataset by randomly choosing indices in the 3 second window. To evaluate the performance of GAIN for various missing data lengths, we use the following configurations of missing data percentages: {2%, 5%, 10%, 20%, and 30%}.

GAIN Structure: We use a GAIN generator neural network with two hidden layers and one output layer. We vary the number of neurons in the hidden layers to perform our design space exploration. The hidden layers in the generator use the ReLU activation. The output layer contains one unit for each sample in the input data and uses the sigmoid activation. It is important to note that the generator produces data even for observed samples. In our implementation we discard those data and use the generated samples for only the missing data.

4.2 Design Space Exploration for GAIN

In this section we vary the number of neurons in the hidden layers with 2 hidden layer and 3 hidden layer architectures and evaluate the accuracy achieved by the robust HAR classifier with GAIN. We also compare the memory requirements of each configuration to understand the overhead of GAIN. The memory requirements, in turn, correlate with the execution time and energy overheads on the wearable device. For instance, if the number of weights double, we see an increase in the execution time and energy required for using the GAIN to generate data. We use this methodology to quantify the overhead because all GAIN configurations are not feasible for the TI-CC2652R processor due to its limited memory. At the same time, this analysis provides a useful comparison point for future wearable with higher memory.

Table 1 – Details of GAIN Architectures

| | 1st Hidden Layer | 2nd Hidden Layer | 3rd Hidden Layer |
|------------------------------|----------------------------|-----------------------------|-----------------------------|
| 2 Hidden Layer Models | 4-183 Acc: 91.04%-96.8% | 4-183 Acc: 91.04%-96.8% | — |
| 3 Hidden Layer Models | 4-22 Acc: 92.37%-88.52% | 11-91 Acc: 92.37%-88.52% | 22-91 Acc: 92.37%-88.52% |

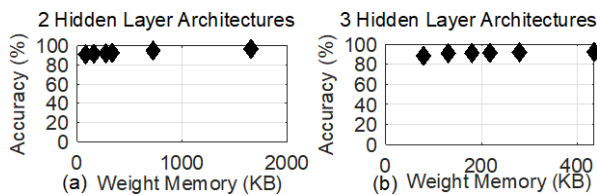


Figure 1 Comparison of accuracy achieved by HAR with GAIN for data recovery

Table 1 provides details on the hidden layer configurations we use for the GAIN. As shown in the table, for the 2 hidden layer models, we vary the number of neurons in both 1st and 2nd layers from 4 to 183. We also work with 3-layer models and vary the number of neurons in the first hidden layer from 4 to 22, while for the second hidden layer we vary from 11 to 91 and lastly, we vary the number of neurons in the third layer from 22 to 91 neurons. For each configuration, we use a uniform mix of missing data percentages and obtain the accuracy.

Figure 1 shows the overall recognition accuracy for each configuration as a function of the number of weights. Increasing the number of weights leads to higher accuracy while incurring a higher overhead. Therefore, designers must choose an architecture that provides the optimal trade-off between the accuracy, device memory constraints, and the overhead. To this end, we implement a three hidden layer architecture with 4 neurons in the first layer, 11 neurons in the second layer and 45 neurons in the third layer. We use this architecture since it fits in the memory of the TI CC2652R processor while providing a high accuracy. Our measurements on the TI CC2652R processor show that GAIN takes 9 ms for the stretch sensor and 15 ms for each accelerometer direction to recover the data. Moreover, the energy consumption is 99 μ J and 168 μ J for stretch and accelerometer, respectively. These overhead values amount to less than 10% overhead for each activity.

5 Conclusion

HAR applications are gaining popularity due to their applications in rehabilitation, movement disorders, and fitness. However, most of the HAR algorithms in use today are not robust to missing sensor values. To improve the robustness of HAR algorithms, this paper explored the use of GAIN to recover missing samples. Our design space exploration showed that GAIN is a promising method to recover missing samples in HAR and maintain high recognition accuracy.

6 References

- [1] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Commun. Surveys & Tut.*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [2] A. J. Espay and others, "Technology in Parkinson's Disease: Challenges and Opportunities," *Movement Disorders*, vol. 31, no. 9, pp. 1272–1282, 2016.
- [3] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill, and U. Y. Ogras, "Online Human Activity Recognition Using Low-Power Wearable Devices," New York, NY, USA, 2018. doi: 10.1145/3240765.3240833.
- [4] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity Recognition Using Cell Phone Accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.
- [5] W. Maetzler, J. Klucken, and M. Horne, "A Clinical View on the Development of Technology-Based Tools in Managing Parkinson's Disease," *Movement Disorders*, vol. 31, no. 9, pp. 1263–1271, 2016.
- [6] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "A Survey of Online Activity Recognition Using Mobile Phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.
- [7] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A Comparative Study on Human Activity Recognition Using Inertial Sensors in a Smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, 2016.
- [8] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: Missing Data Imputation using Generative Adversarial Nets," *arXiv:1806.02920 [cs, stat]*, Jun. 2018, Accessed: Nov. 13, 2021. [Online]. Available: <http://arxiv.org/abs/1806.02920>
- [9] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [10] G. Bhat, N. Tran, H. Shill, and U. Y. Ogras, "w-HAR: An Activity Recognition Dataset and Framework Using Low-Power Wearable Devices," *Sensors*, vol. 20, no. 18, p. 5356, Sep. 2020, doi: 10.3390/s20185356.
- [11] "CC2652R SimpleLink™ Multiprotocol 2.4 GHz Wireless MCU." Texas Instruments, 2018. [Online]. Available: <https://www.ti.com/lit/ds/symlink/cc2652r.pdf?ts=1>

- 611278259614&ref_url=https%253A%252F%252Fwww.ti.com%252Fproduct%252FCC2652R
- [12] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Oct. 2017, pp. 2764–2773. doi: 10.1109/ICCV.2017.299.
- [13] Md. A. Haidar and M. Rezagholizadeh, "TextKD-GAN: Text Generation Using Knowledge Distillation and Generative Adversarial Networks," in *Advances in Artificial Intelligence*, vol. 11489, M.-J. Meurs and F. Rudzicz, Eds. Cham: Springer International Publishing, 2019, pp. 107–118. doi: 10.1007/978-3-030-18305-9_9.
- [14] A. Sriram, H. Jun, Y. Gaur, and S. Satheesh, "Robust Speech Recognition Using Generative Adversarial Networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Apr. 2018, pp. 5639–5643. doi: 10.1109/ICASSP.2018.8462456.

FedGroup: Efficient Federated Learning via Decomposed Similarity-Based Clustering

Moming Duan*, Duo Liu*, Xinyuan Ji†, Renping Liu*, Liang Liang*, Xianzhang Chen*, Yujuan Tan*

*College of Computer Science, Chongqing University, Chongqing, China

†School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

Abstract—Federated Learning (FL) enables the multiple participating devices to collaboratively contribute to a global neural network model while keeping the training data locally. Unlike the centralized training setting, the non-IID and imbalanced (statistical heterogeneity) training data of FL is distributed in the federated network, which will increase the divergences between the local models and the global model, further degrading performance. In this paper, we propose a novel clustered federated learning (CFL) framework FedGroup, in which we 1) group the training of clients based on the similarities between the clients' optimization directions for high training performance; 2) construct a new data-driven distance measure to improve the efficiency of the client clustering procedure. 3) implement a newcomer device cold start mechanism based on the auxiliary global model for framework scalability and practicality. FedGroup can achieve improvements by dividing joint optimization into groups of sub-optimization and can be combined with FL optimizer *FedProx*. The convergence and complexity are analyzed to demonstrate the efficiency of our proposed framework. We also evaluate FedGroup and FedGrouProx (combined with *FedProx*) on several open datasets and made comparisons with related CFL frameworks. The results show that FedGroup can significantly improve absolute test accuracy by +14.1% on FEMNIST compared to *FedAvg*, +3.4% on Sentiment140 compared to *FedProx*, +6.9% on MNIST compared to *FeSEM*.

Index Terms—Federated Learning, Distributed Machine Learning, Neural Networks.



1 INTRODUCTION

Federated Learning (FL) [1], [2], [3], [4], [5] is a promising distributed neural network training approach, which enables multiple end-users to collaboratively train a shared neural network model while keeping the training data decentralized. In practice, a FL server first distributes the global model to a random subset of participating clients (e.g. mobile and IoT devices). Then each client optimizes its local model by gradient descent based on its local data in parallel. Finally, the FL server averages all local models' updates or parameters and aggregates them to construct a new global model. Unlike the traditional cloud-centric learning paradigm and the distributed machine learning frameworks based on Parameter Server [6], there is no need to transfer private data over the communication network during the FL training. With the advantage of privacy-preserving, Federated Learning is currently the most attractive distributed machine learning framework. Nevertheless, due to the FL server does not have the authority to access the user data or collect statistical information, some data preparation operations such as balancing and outlier detection be restricted. Therefore, the high statistical heterogeneity is a challenging problem in federated learning [5].

In this paper, we propose an efficient and accurate clustered federated learning framework FedGroup, which clusters clients into multiple groups based on a new decomposed data-driven measure between their parameter updates. In each communication round, each active client only contributes its local optimization result to the corresponding group model. The framework still maintains an auxiliary server to address the cold start issues of new de-

vices. To improve the performance of high-dimension low-sample size (HDLSS) parameter updates clustering, we use a novel data-driven measure of cosine dissimilarity called Euclidean distance of Decomposed Cosine similarity (EDC), which can also avoid the concentration phenomenon of ℓ_p distances in high dimensional data clustering [19]. Furthermore, by combining FedGroup with the federated optimizer *FedProx* [20], FedGroup can be revised as FedGrouProx, which be explored in our experiments.

2 FEDGROUP AND FEDGROUPROX

2.1 Framework Overview

Before we go into more detail, we first show the general overview of the FedGroup. The federated training procedure of FedGroup is shown in Fig. 1. FedGroup contains one auxiliary server, a certain amount of groups, and multiple clients, each of them maintains the latest model and the latest update for this model. Each client and group have a one-to-one correspondence, but there are also possible to have a group without clients in a communication round (empty group) or a client is not in any groups (e.g. a newcomer joins the training later). The newcomer device uses a cold start algorithm to determine the assigned group (we call the unassigned client is cold), we will explain the details of the cold start algorithm later. For the dataset, each client c_i has a training set and a test set according to its local data distribution $p_{data}^{(c_i)}$.

As shown in the Fig. 1, FedGroup has three model transmission processes, including intra-group aggregation (①), inter-group aggregation (②), and optimization gradient upload (③). First, the auxiliary server determines the initial

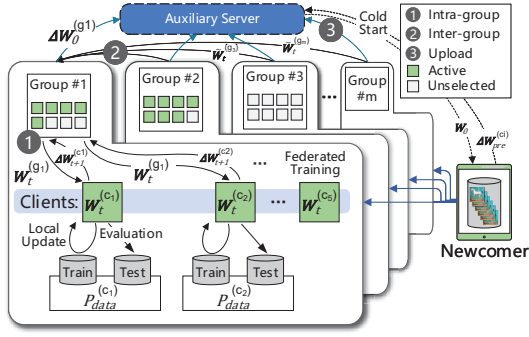


Fig. 1. An overview of FedGroup.

Algorithm 1 Group Cold Start

Input: Clients set \mathcal{C} , number of group m , global initial model w_0 , pre-training scale hyperparameter α .

Output: Groups set \mathcal{G} , set of group parameters $\mathcal{W}_0^{(G)}$, set of group updates $\Delta\mathcal{W}_0^{(G)}$.

```

1: procedure GROUP COLD START
2:    $S \leftarrow$  Server selects a random subset of  $\alpha * m$  clients.
3:   Server broadcasts  $w_0$  to all selected clients.
4:   for each client  $c_i$  in  $S$  parallelly do
5:      $\Delta w_0^{(c_i)} \leftarrow \text{flatten}(\text{ClientUpdate}(i, w))$ .
6:    $\Delta\mathcal{W} \leftarrow [\Delta w_0^{(c_1)}, \Delta w_0^{(c_2)}, \dots, \Delta w_0^{(c_{\alpha * m})}]$ .
7:   if MADC then:
8:      $\mathcal{M} \leftarrow K(\Delta\mathcal{W}, \Delta\mathcal{W})$ . //Ref. Eq. (2)
9:     Proximity matrix  $\mathcal{M}_p \leftarrow \text{Calculate MADC}(\mathcal{M})$ . //Ref. Eq. (3)
10:     $[g_1.\text{clients}, \dots, g_m.\text{clients}] \leftarrow \text{Hierarchical Clustering}(\mathcal{M}_p, m)$ .
11:   if EMD then:
12:      $V \leftarrow \text{truncated SVD}(\Delta\mathcal{W}^T, m)$ .
13:     Distance matrix  $\mathcal{M}_d \leftarrow K(\Delta\mathcal{W}, V^T)$ . //Ref. Eq. (4)
14:      $[g_1.\text{clients}, \dots, g_m.\text{clients}] \leftarrow \text{K-Means++}(\mathcal{M}_d, m)$ .
15:   for  $g_j$  in  $\mathcal{G} \leftarrow [g_1, \dots, g_m]$  do
16:      $w_0^{(g_j)} \leftarrow \text{Average}([\Delta w_0^{(c_i)}, \forall c_i \in g_j.\text{clients}])$ .
17:      $\Delta w_0^{(g_j)} \leftarrow w_0^{(g_j)} - w_0$ .
18:    $\mathcal{W}_0^{(G)} \leftarrow [w_0^{(g_1)}, \dots, w_0^{(g_m)}]$ , and  $\Delta\mathcal{W}_0^{(G)} \leftarrow$ 
19:      $[\Delta w_0^{(g_1)}, \dots, \Delta w_0^{(g_m)}]$ .
20:   return  $\mathcal{G}, \mathcal{W}_0^{(G)}, \Delta\mathcal{W}_0^{(G)}$ 
    
```

model and optimization direction of each group through clustering. Then, each group federally trains on a certain set of clients based on their training data to optimize its group model, and evaluates the group model on the same set of clients based on its test data. Specifically, each group broadcasts its model parameters to their clients and then aggregates the updates from these clients using *FedAvg* in parallel, we call this aggregation procedure the intra-group aggregation. After all federated trainings of groups complete, the models are aggregated using a certain weight, which we call inter-group aggregation. Note that, unlike the previous FedAvg-based frameworks [2], [12], [20], [29], [30], the optimization gradients in the server is not broadcast to all clients or groups, we only maintain this gradients for the cold start of newcomers.

2.2 Group Cold Start

For the choice of clustering measure, a heuristic way is to use the ℓ_2 distance between models, which has miserable performance as we will mention in Section 3. The loss value can be used as a surrogate for the proximity of difference domains [31], but the huge computational overhead involved in model inferencing. The cosines similarity between the gradients calculated by backpropagation or the updates

Algorithm 2 FedGroup and FedGrouProx

Input: Clients set $\mathcal{C} \leftarrow \{c_1, c_2, \dots, c_n\}$, groups set $\mathcal{G} \leftarrow \{g_1, g_2, \dots, g_m\}$, initial group parameters set $\mathcal{W}_0^{(G)} \leftarrow \{w_0^{(g_1)}, w_0^{(g_2)}, \dots, w_0^{(g_m)}\}$, $g_j.\text{clients} \leftarrow \{c_i | c_i \text{ is in group } g_j, \forall c_i \in \mathcal{C}\}$, initial model parameters w_0 , number of communication rounds T , number of selected clients per round K , inter-group learning rate η_g , proximal hyperparameter μ .

Output: Updated group model parameters $\mathcal{W}_T^{(G)}$.

```

1: procedure FEDGROUP PROX TRAINING
2:    $\mathcal{W}_1^{(G)} \leftarrow \{w_1^{(g_1)}, w_1^{(g_2)}, \dots, w_1^{(g_m)}\} \leftarrow$  initialized  $\mathcal{W}_0^{(G)}$  by  $w_0$ .
3:   for each communication round  $t = 1, 2, \dots, T$  do
4:      $S_t \leftarrow$  Server selects a random subset of  $K$  clients.
5:     for each group  $g_j$  in  $\mathcal{G}$  parallelly do
6:        $S_t^{(g_j)} \leftarrow \{c_i | c_i \in g_j.\text{clients}, \forall c_i \in S_t\}$ .
7:        $\tilde{w}_{t+1}^{(g_j)} \leftarrow \text{IntraGroupUpdate}(S_t^{(g_j)}, w_t^{(g_j)})$ .
8:        $\tilde{\mathcal{W}}_{t+1}^{(G)} \leftarrow \{\tilde{w}_{t+1}^{(g_1)}, \tilde{w}_{t+1}^{(g_2)}, \dots, \tilde{w}_{t+1}^{(g_m)}\}$ .
9:     for each group  $g_j$  in  $\mathcal{G}$  parallelly do
10:       $\mathcal{W}_{t+1}^{(G)} \leftarrow \text{InterGroupAggregation}(\tilde{\mathcal{W}}_{t+1}^{(G)}, \eta_g)$ .
11:   function INTRAGROUPUPDATE( $S_t, w_t$ )
12:     if  $S_t$  is  $\emptyset$  then return  $w_t$ . //Empty group.
13:     IF FedGroup:  $w_{t+1} \leftarrow \text{FedAvg}(S_t, w_t)$ . //Ref. Algorithm ??
14:     IF FedGrouProx:  $w_{t+1} \leftarrow \text{FedProx}(S_t, w_t, \mu)$ . //Ref. [20].
15:     return  $w_{t+1}$ 
16:   function INTERGROUPAGGREGATION( $\tilde{\mathcal{W}}_{t+1}, \eta_g$ )
17:     for each group parameter  $\tilde{w}_{t+1}^{(g_j)}$  in  $\tilde{\mathcal{W}}_{t+1}$  parallelly do
18:        $\Delta \tilde{w}_{t+1}^{(g_j)} \leftarrow \eta_g \sum_{l \neq j} \frac{\tilde{w}_{t+1}^{(g_l)} - \tilde{w}_{t+1}^{(g_j)}}{\|\tilde{w}_{t+1}^{(g_l)} - \tilde{w}_{t+1}^{(g_j)}\|}$ , then  $w_{t+1}^{(g_j)} \leftarrow \tilde{w}_{t+1}^{(g_j)} + \Delta \tilde{w}_{t+1}^{(g_j)}$ .
19:     return  $\mathcal{W}_{t+1} \leftarrow \{w_{t+1}^{(g_1)}, w_{t+1}^{(g_2)}, \dots, w_{t+1}^{(g_m)}\}$ 
    
```

of model parameters is an alternative measure. The cosine similarity between the updates of any two clients c_i and c_j is defined by:

$$\mathcal{S}(i, j) \triangleq \frac{\langle \Delta w_t^{(c_i)}, \Delta w_t^{(c_j)} \rangle}{\|\Delta w_t^{(c_i)}\| \|\Delta w_t^{(c_j)}\|}, \quad (1)$$

and the pairwise cosine similarity matrix $\mathcal{M} \subset \mathbb{R}^{n \times n}$ can be written with cosine similarity kernel K as follows:

$$\mathcal{M} = K(\Delta\mathcal{W}, \Delta\mathcal{W}), \quad \mathcal{M}_{ij} = \mathcal{S}(i, j). \quad (2)$$

The computational complexity of calculating \mathcal{M} is $O(n^2 d_w^2)$, n and d_w are the number of clients and number of parameters, respectively. We assume the all parameters updates Δw_t are flattened row vectors, so $\Delta w_t \subset \mathbb{R}^{1 \times d_w}$. In general, the number of parameters d_w is huge and $d_w \gg n$ (HDLSS), which make the pairwise cosine similarity-based clustering methods computationally inefficient. Unlike ℓ_p distance, the expectation of \mathcal{S} asymptotically remains constant as dimensionality increases [27], which is friendly to the clustering in high dimensional data. Unfortunately, \mathcal{S} is not suitable for low-dimensional situations because the variance of it is $O(1/d_w)$. Therefore, we extend the data-driven method MADD [19] to our cosine similarity-based case, so we can reduce the observation bias by using the mean of residuals of \mathcal{S} , fox example:

$$\text{MADC}(i, j) = \frac{1}{n-2} \sum_{z \neq i, j} |\mathcal{S}(i, z) - \mathcal{S}(j, z)|. \quad (3)$$

The above dissimilarity measure is based on the Mean of Absolute Differences of pairwise Cosine similarity, so we call it MADC. However, MADC and \mathcal{M} have the same computation complexity and both are proximity measures, which means they cannot be applied for the efficient Euclidean distance-based clustering algorithms. These motivate us to develop a variant of MADC called Euclidean

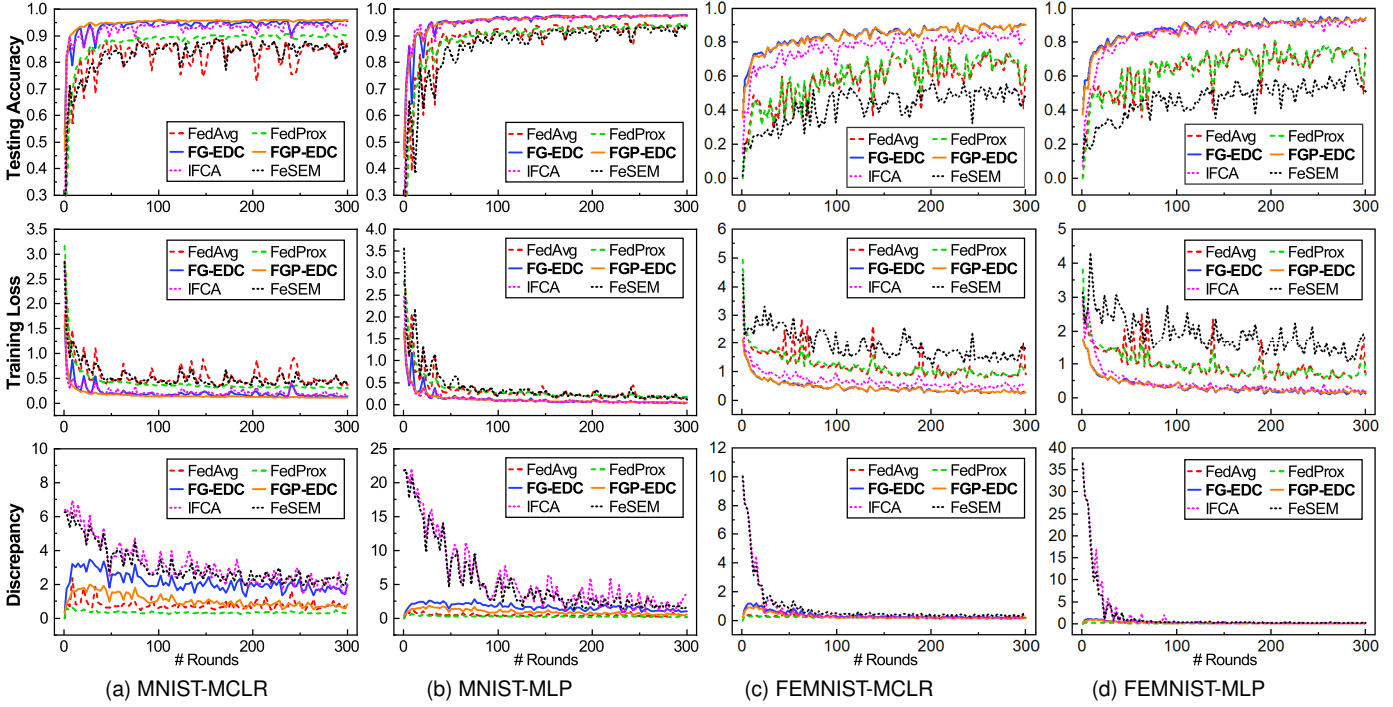


Fig. 2. Evaluation results on MNIST ($m = 3$) and FEMNIST ($m = 5$). Top: test accuracy; Middle: mean training loss; Bottom: discrepancy between clients and server (FedAvg and FedProx) or weighted discrepancy between clients and groups (FedGroup, FedGrouProx, IFCA, FeSEM).

distance of Decomposed Cosine similarity (EDC), which is defined by:

$$EDC(i, j) = \frac{1}{m} \sqrt{\sum_{v \in \mathbf{V}} (\mathcal{S}(i, v) - \mathcal{S}(j, v))^2}, \text{ or}$$

$$EDC(i, j) = \frac{1}{m} \|K(\Delta \mathbf{w}_i^{(c_i)}, \mathbf{V}^T) - K(\Delta \mathbf{w}_j^{(c_j)}, \mathbf{V}^T)\|, \quad (4)$$

$$\mathbf{V} = SVD(\Delta \mathbf{W}^T, m), \mathbf{V} \subset \mathbb{R}^{d_w \times m}.$$

Instead of calculating the pairwise similarity, EDC first decomposes the updates of models into m directions by using truncated Singular Value Decomposition (SVD) algorithm [32], so then only the similarities between the updates and these directions will be calculated. It is worth noting that the complexity of truncated SVD is only $O(2m^2 d_w)$ for $d_w \gg m$ and hence the computational complexity of EDC is $O(m^2 d_w^2)$. Some previous works (i.e. [13], [17]) calculate the pairwise cosine similarity based on all participants, so n is usually hundreds or thousands. Furthermore, [13], [17] leverage the hierarchical clustering strategies, which are recursively and computationally expensive.

To determine the optimization goals of each group, FedGroup clusters the parameter updates directions of clients into m groups using K-Means++ [33] algorithm based on EDC. The main advantage of the clustering approach is that it is unsupervised, and we can divide the global optimization function into m sub-optimization functions regardless of whether there have incongruent optimization goals. Moreover, training all clients to cluster their updated directions is not communication-friendly and practically achievable. In other words, FedGroup performs a low-dimensional embedding of the local updates matrix $\Delta \mathbf{W}$, following by K-Means++ clustering. Our calculation of client clustering is following the calculation of the similarity matrix, which means the above calculations only require

one round of communication. We call the combination of the above two processes as group cold start, and the details are shown in Algorithm 1. For comparison, we also provide the MADC version of FedGroup, which is clustered using the hierarchical strategy with the complete linkage.

After the server performs the group cold start, the optimization direction of group j , which is measured by $\Delta \mathbf{w}_0^{(g_j)}$ is determined, and the αm clients participating in this process are assigned. We leverage the centric means of groups to measure the clustering validity index like within-cluster sum-of-squares criterion. The optimization gradients will be uploaded to the auxiliary server for cold start of the newcomers.

2.3 Client Cold Start

As described before, the group cold start algorithm selects a random subset (αm) of the clients for pre-training, so the remaining clients ($n - \alpha m$) are cold clients and are not in any groups. Since the federated training network is dynamic, the new devices can join the training at any time, so we need to classify newcomers according to the similarity between their optimization goals and groups'. Our client cold start strategy is to assign clients to the groups that are most closely related to their optimization direction, as shown below:

$$g^* = \operatorname{argmin}_j \frac{-\cos(\langle \Delta \mathbf{w}_0^{(g_j)}, \Delta \mathbf{w}_{pre}^{(i)} \rangle) + 1}{2}. \quad (5)$$

Suppose the newcomer i joins the training network in round t , then the $\Delta \mathbf{w}_{pre}^{(i)}$ is the pre-training gradient of newcomer base on the global initial model \mathbf{w}_0 . We schedule the newcomer i to group g^* to minimize the normalized cosine dissimilarity. With this mechanism, FedGroup does not need to broadcast all groups' models for clustering.

3 EVALUATION

The experimental results in specific rounds are shown in Fig. 2. The accuracy and discrepancy curves illustrate that adding the proximal term can reduce the divergence caused by the heterogeneous data and make the training more stable. Although the training procedure of FedGrouProx is more stable than that of FedGroup, there is not enough evidence to suggest that adding the proximal term is significantly helpful in improving accuracy. Compare to IFCA and FeSEM, the discrepancy of FedGroup remains at a low level in the initial round, which is attributed to the effective group cold start strategy. The FedAvg and FedProx achieve low discrepancy by maintaining the consensus global model, but this will degrade the accuracy, which is supported by the obtained results. FedGroup also shows a significant improvement in convergence speed compared to FedAvg, which is helpful to reduce the communication consumption of the FL systems. By the way, the rescheduling methods of IFCA and FeSEM involve additional overheads in each round, such as broadcasting group models and calculating loss.

REFERENCES

- [1] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.
- [2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
- [3] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan et al., "Towards federated learning at scale: System design," in *Proceedings of the 2nd SysML Conference*, 2019.
- [4] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.
- [5] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [6] M. Li, D. G. Andersen, J. W. Park, A. J. Smola, A. Ahmed, V. Josifovski, J. Long, E. J. Shekita, and B.-Y. Su, "Scaling distributed machine learning with the parameter server," in *Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2014, pp. 583–598.
- [7] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.
- [8] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Citeseer, Tech. Rep., 2009.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. IEEE, 2015.
- [10] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-i.i.d. data," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, pp. 1–14, 2019.
- [11] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," in *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.
- [12] M. Duan, D. Liu, X. Chen, Y. Tan, J. Ren, L. Qiao, and L. Liang, "Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications," in *Proceedings of the IEEE 37th International Conference on Computer Design (ICCD)*. IEEE, 2019, pp. 246–254.
- [13] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, pp. 1–13, 2020.
- [14] F. Sattler, K.-R. Müller, T. Wiegand, and W. Samek, "On the byzantine robustness of clustered federated learning," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8861–8865.
- [15] M. Xie, G. Long, T. Shen, T. Zhou, X. Wang, and J. Jiang, "Multi-center federated learning," *arXiv preprint arXiv:2005.01026*, 2020.
- [16] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 19 586–19 597.
- [17] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–9.
- [18] Y. Zhang, M. Duan, D. Liu, L. Li, A. Ren, X. Chen, Y. Tan, and C. Wang, "CSAFL: A clustered semi-asynchronous federated learning framework," *arXiv preprint arXiv:2104.08184*, 2021.
- [19] S. Sarkar and A. K. Ghosh, "On perfect clustering of high dimension, low sample size data," *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*, vol. 42, no. 9, pp. 2257–2272, 2019.
- [20] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of the 3rd SysML Conference*, 2020.
- [21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [22] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending mnist to handwritten letters," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 2921–2926.
- [23] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N project report, Stanford*, 2009.
- [24] M. Duan, D. Liu, X. Chen, R. Liu, Y. Tan, and L. Liang, "Self-balancing federated learning with global imbalanced data in mobile systems," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 32, no. 1, pp. 59–71, 2020.
- [25] L. Li, M. Duan, D. Liu, Y. Zhang, A. Ren, X. Chen, Y. Tan, and C. Wang, "FedSAE: A novel self-adaptive federated learning framework in heterogeneous systems," *arXiv preprint arXiv:2104.07515*, 2021.
- [26] J.-w. Lee, J. Oh, S. Lim, S.-Y. Yun, and J.-G. Lee, "Tornadoaggregate: Accurate and scalable federated learning via the ring-based architecture," *arXiv preprint arXiv:2012.03214*, 2020.
- [27] M. Radovanović, A. Nanopoulos, and M. Ivanović, "On the existence of obstinate results in vector space models," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 186–193.
- [28] L. Liu, J. Zhang, S. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *Proceedings of the IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.
- [29] V. Smith, C.-K. Chiang, M. Sanjabi, and A. Talwalkar, "Federated multi-task learning," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4427–4437.
- [30] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2017, pp. 1175–1191.
- [31] M. Mohri, G. Sivek, and A. T. Suresh, "Agnostic federated learning," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 4615–4625.
- [32] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," in *Linear Algebra*. Springer, 1971, pp. 134–151.
- [33] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," Stanford, Tech. Rep., 2006.
- [34] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., "Tensorflow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.
- [35] O. Shamir, N. Srebro, and T. Zhang, "Communication-efficient distributed optimization using an approximate newton-type method," in *Proceedings of the 31th International conference on machine learning (ICML)*, 2014, pp. 1000–1008.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Dr. |
| Name | Kihan Park |
| Affiliation | Department of Mechanical Engineering, University of Massachusetts Dartmouth |
| Address | 285 Old Westport Road, Dartmouth, MA 02747, USA |
| E-mail | <i>kihan.park@umassd.edu</i> |

Low-Cost Shape Sensing Technique for Soft Robotic Actuators

Rafsan Al Shafatul Islam Subad, Liam B. Cross, Kihan Park*

Department of Mechanical Engineering, University of Massachusetts Dartmouth
285 Old Westport Road, Dartmouth, MA 02747, USA

*Corresponding Author: kihan.park@umassd.edu

Abstract – Estimating the shape of a soft-robotic actuator can play a crucial role to reckon the shape of the object that is being handled by the end-effector. In this paper, we articulate a novel, simple, and low-cost solution for reconstructing the 3D shape of a soft robotic gripper utilizing just two commercial-grade piezo-resistive flex sensors for each finger. We also implement force sensors to gather force data exerted by the gripper. Our proposed methodology to reconfigure shape demonstrates a promising standpoint for approximating the shape of soft-robotic fingers and replacing the shape sensing technology of the soft-robotic actuator in the future in terms of simplicity and cost-effectiveness.

Keywords: Soft-Robotics, Shape Sensing Algorithm, Automation.

1 Introduction

Soft robotics has emerged as a new growing field of study in recent days. It has revolutionized the area of robotics, with the potential to transform society and industries that demand compliant robot interfaces [1]–[4]. Soft robotic actuators are steadily drawing the interest of researchers, and industries that are partially or predominantly dependent upon robots due to their flexibility in target object handling, safe human-machine interaction, adaptability to wearable devices, facile actuation, simple gripping system, so on. Hence, they possess tremendous prospects in automation, space industries, the medical sector, underwater environment, daily life household applications, etc [5]–[8]. Generally, end-effectors employed in most of these cases have been governed by human remote-control. In some of the scenarios described above, multi-modal tactile sensing may be required to approximate the object it's handling (without any camera vision) not just to offer extra information to the operator, but also to improve the tasks that a robot can accomplish autonomously [9]. For example, factory line automation assembly may take advantage of a soft robotic end-effector equipped with multi-modal sensing capability to detect whether it is handling the right object, or if the shape and the size of the object are perfect enough in the production line. For this, a low-cost sensing technology might be a game changer.

However, due to their particular role in applications, many common sensors used in hard robots become outdated in the case of compliance and morphology to soft robotics. As a result, there is now a major research effort worldwide focused on developing state-of-the-art

numerical models, and stretchable electronics such as electronic skin, screen-printed curvature sensors, and chemical and biological-based sensors to facilitate tactile and shape sensing for robotics hands [10]. Up to date, fiber-brag optical fiber-based sensors are being utilized mostly to determine the shape of the robotic actuator [11]–[13]. However, they are quite expensive. To make shape sensing technology widely available, researchers around the world are working on developing contemporary numerical models, methods, and sensors.

In this work, we report a simple and cost-effective approach to estimate the shape of a soft robotic actuator. We employ two commercial-grade piezo-resistive flex sensors, and a developed novel algorithm to determine the shape of the robotic finger. Shape of the robotic gripper has been regenerated and compared with physical images for single finger, general tri-actuator, and object grasping cases. Force sensors have also been utilized in each finger to evaluate the force exerted by the actuators on the object while grasping.

2 Proposed Method

We propose implementing flex sensors to approximate the shape of the fingers of a soft robotic hand. It should be noted that the curvature of a finger can be recognized precisely by employing as many flex sensors as possible. However, using more flex sensors would create problems handling the electrical circuitry and make the numerical model complex to estimate the shape. Thus, for simplicity and low-cost solution, we employed two piezo-resistive flex sensors potted in 10A durometer silicone to reproduce a 2-part compound radius of curvature.

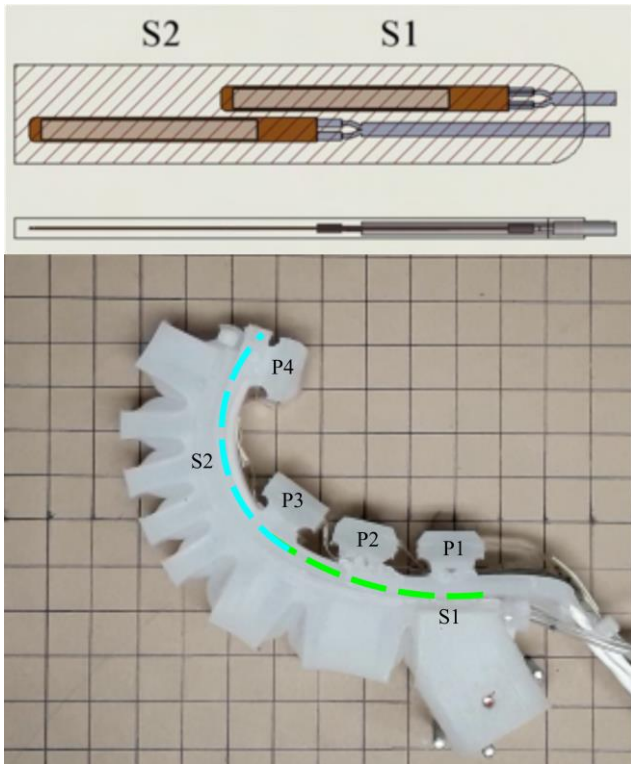


Figure 1: Positions of flex sensors S_1 and S_2 (top) and force sensors P_1 , P_2 , P_3 , and P_4 (bottom).

Figure 1 (top) demonstrates the positions of the flex sensors placed inside the finger. It also shows that the two serially positioned flex sensors in silicone molding are good enough to approximate the curvature of the soft robotic finger.

2.1 Calibration

To generate the curvature of an actuator, we first calibrate each flex sensor having resistance data (converted from voltage) with the change of radius of curvature of a finger. The voltage outputs from each of the two embedded flex sensors were recorded for the fully open to close position of a finger at intervals of 1 second within a period from 0 to 32 seconds and converted into resistance values.

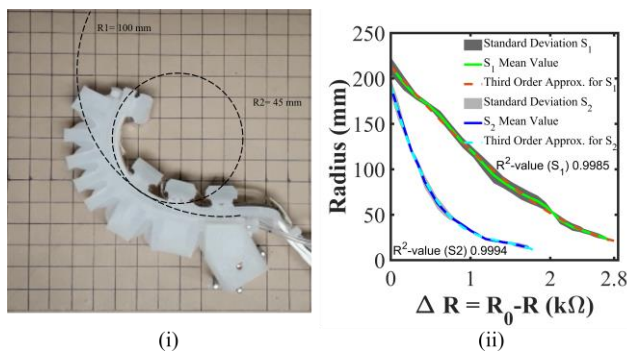


Figure 2. Calibration technique for shape sensing.

Next, we use a secondary drawing tool to gauge the sizes of the two radii as shown for example in Figure 2(i) for each 1-second-image. In order to verify the repeatability of the calibration, we have conducted the test four times and calculated the mean value and the standard deviation. The collected dataset from the four trials is demonstrated in Figure 2(ii). It can be assessed from the figure that the change in resistance value with the change of radius value is quite consistent. Both the curve for S_1 and S_2 have been fitted to third-degree polynomial equations as below:

- $R_1 = 2.147 * S_1^3 + 4.909 * S_1^2 - 99.19 * S_1 + 213.4$
- $R_2 = -50.64 * S_2^3 + 217.2 * S_2^2 - 327.7 * S_2 + 192.4$

R^2 -value of the curve for S_1 and S_2 are 0.9985 and 0.9994, respectively which ensure goodness of fit. Calibration equations are generated for each of the flex sensors used since they are different from each other.

2.2 Algorithm to Regenerate Shape

We have utilized a voltage divider circuit and Arduino Mega to get the voltage values from the flex sensors and converted them into resistance values.

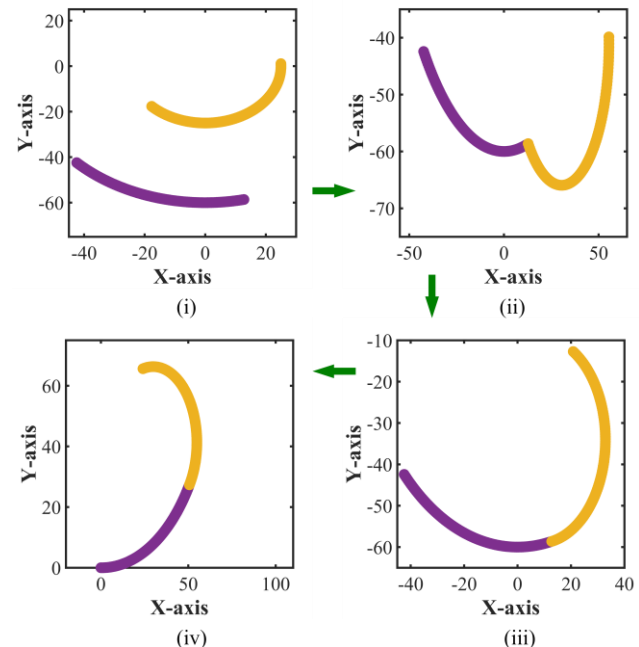


Figure 3. Visual demonstration of the algorithm to regenerate the shape of the robot finger.

The calibration equation is now set in MATLAB to produce the radius value from the change in resistance. With the radius values, we first plot two simple radii of curvature. Next, we combine them with the help of co-ordinate transformation and make the initial point of the S_2 -curve tangent to the endpoint of the S_1 -curve. Now again, implementing co-ordinate transformation, we shift the initial point of the S_1 -curve either to the origin (in case

of 2D) or to the hinge point of the robotic finger according to the physical design in the coordinate system in the 3D general case. The algorithm to generate the shape of a finger is presented step by step in Figure 3. Moreover, we employ a rotational matrix to rotate the finger orientation in a general 3D case according to the servo motor positions (discussed in 3.2 Tri-actuator case) [14].

3 Results and Discussions

3.1 Single Actuator Case

In order to verify the actuator’s shape sensing performance, we compared reconstructed curvature with the actuator’s full range of actual physical motion demonstrated in Figures 5, and 4 accordingly. To capture the fully open to close position of the actuator, a video has been recorded with a time span of 32 seconds and images taken from the video at a 4-second interval (Figure 4). Respective reconstruction of the shape of the finger has been depicted in Figure 5. It should be noted that we have plotted the zero-second figure manually (assuming two radius data far greater than the usual) since the radius would be theoretically infinite in that case.

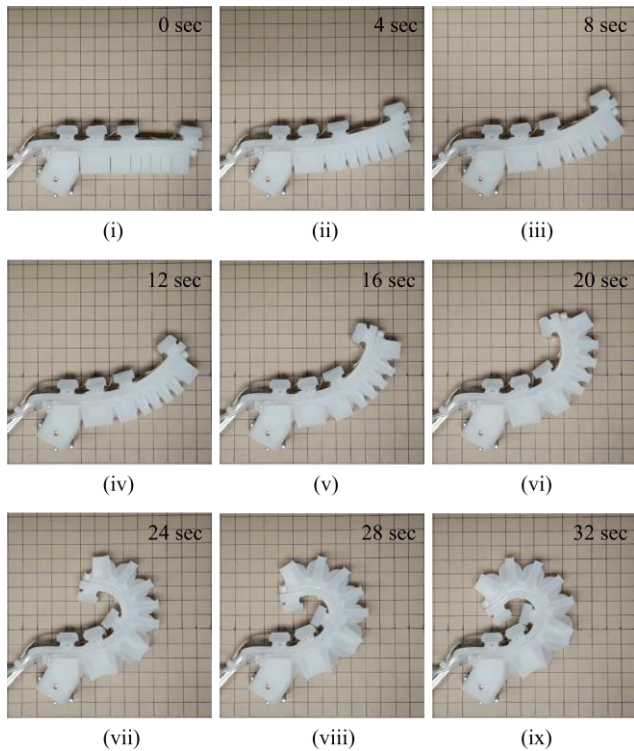


Figure 4. Time-stamped physical image of a single actuator from fully open to close position.

The consistent approximation of the shape regeneration of the finger exhibits a promising outlook to implement this model in future fieldwork as an accurate and effective solution to tactile shape sensing. To provide further verification of the claim, approximated curvature has been

superimposed onto the corresponding physical configuration of the end-effector for 4, 20, and 32 seconds (Figure 6). It is evident from those figures that the proposed model can regenerate the shape of the finger almost accurately.

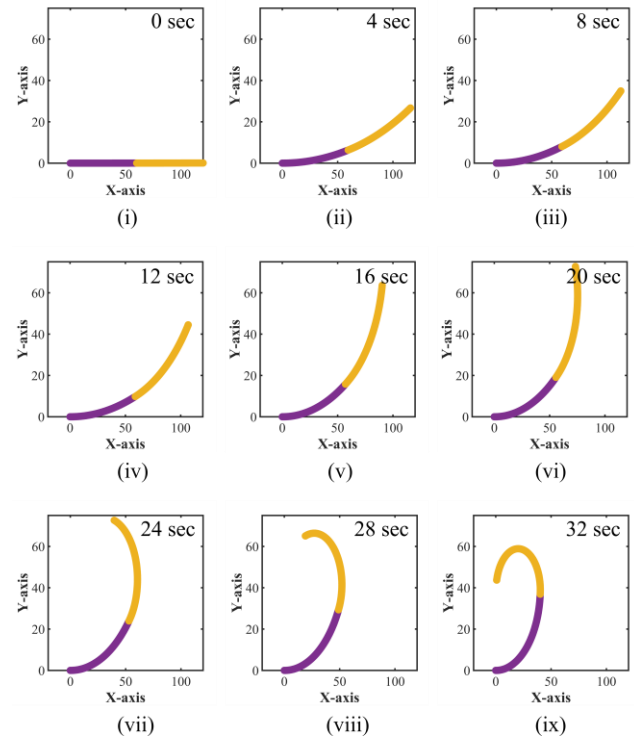


Figure 5. Corresponding regeneration of the shape of the actuator.

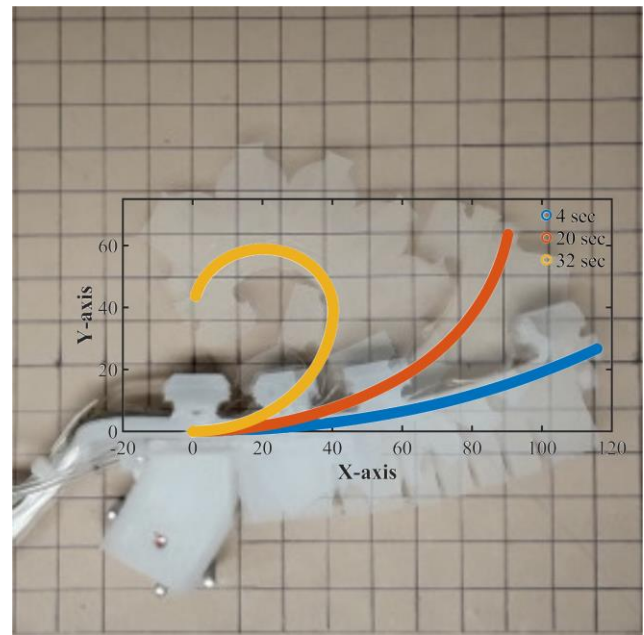


Figure 6. Comparison of the reconstructed shape with the actual shape of the finger.

3.2 Tri-actuator Case

Our robotic hand is equipped with three fingers. There are servos positioned at the hinge point of each actuator. And with the help of the servo, the actuator can rotate roughly around 60° on each side (120° total) (Figure 7). So, it is also important to represent the rotational capability for the reconstructed finger shape for implementing it in general cases.

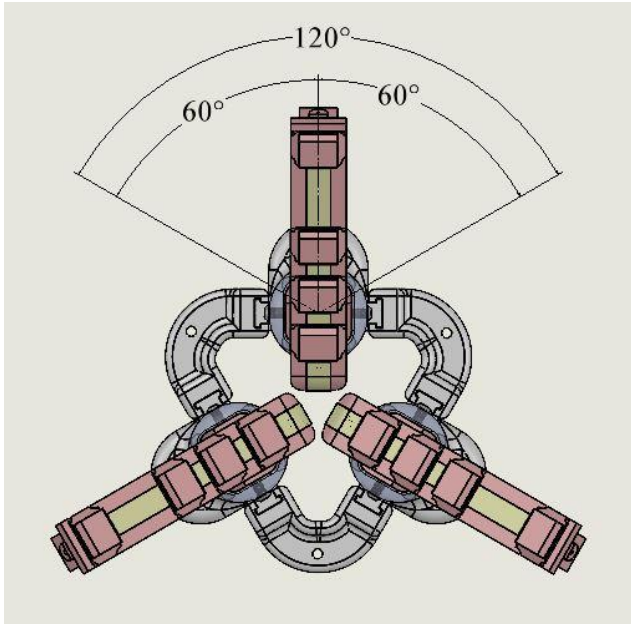


Figure 7. The angular range of the single actuator.

By using translation, and a rotational matrix, we can rotate the constructed shape of the finger according to the servo positions. Figure 8 demonstrates the different shapes and rotational positions of the fingers and corresponding reconstruction.

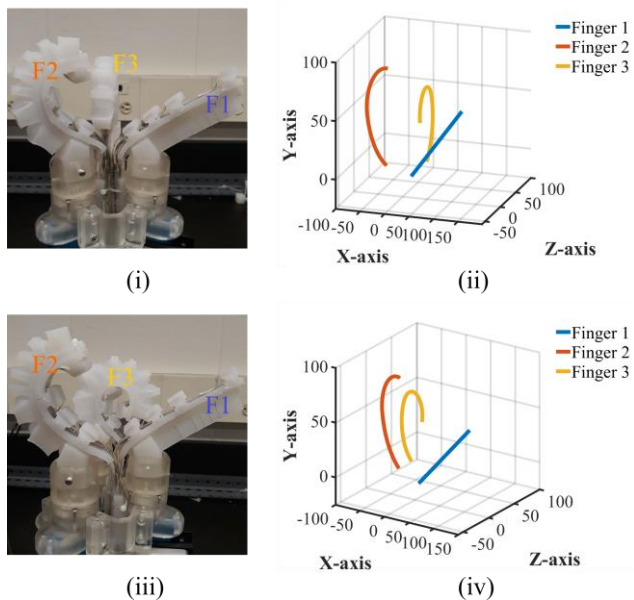


Figure 8. Reproducing different positions and shapes of the robotic fingers.

3.3 Object Grasping

There are three types of objects grasping mode in our soft robotic hands such as Y, T, and U mode for (Figure 8 (i), (iii), and (ii) accordingly). Y and T modes are for spherical-shaped objects e.g., apple, tennis ball, and U mode is for cylindrical-shaped object grasping.

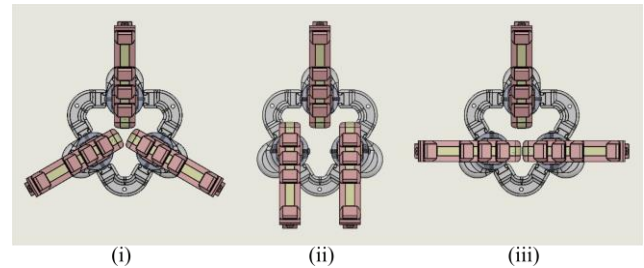


Figure 8. Top views of three grasping modes (i) Y, (ii) U, (iii) T mode.

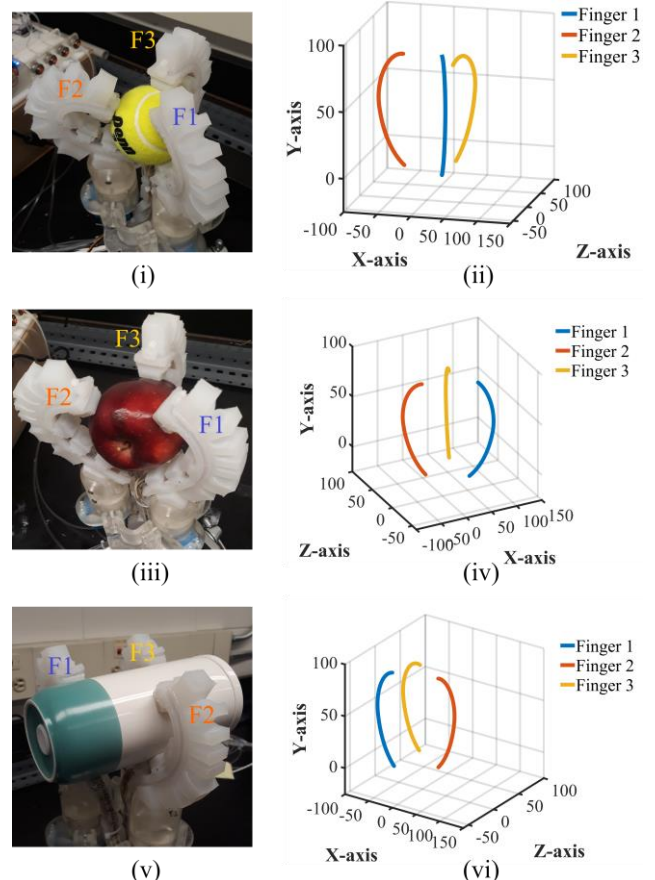


Figure 9. Different object grasping modes of the robotic hand and the corresponding shape of the fingers.

Data has been taken from each flex sensor of three fingers. And figures of the shape of the individual finger have been reconstructed manipulating the data. Figure 9 (i) shows the Y grip of the robotic hand for a tennis ball, (ii)

shows the T grip for an apple, and (iii) shows the U grip for a cylindrical-shaped object.

3.4 Force Measurement

In addition to the shape sensing, we collected force values exerted on the objects (apple, tennis ball, and cylindrical shaped object) while grasped by the three-fingered soft robotic actuator. Force sensors positions are demonstrated in Figure 1 (bottom). It shows that the P₁ and P₄ are the two most important sensors in the system for grasping objects in 3D. In our previous study (currently under review for IEEE RA-L), it was demonstrated that the other two sensors P₂, and P₃ can provide important information for a smaller object while grasped by a single actuator in 2D motion. We plan to add one more force sensor at the top of the finger to get data for pinch grasp as well. In the future scope of the study, we aspire to build a model to predict the shape of the object in real-time handled by the gripper combining the shape and force data.

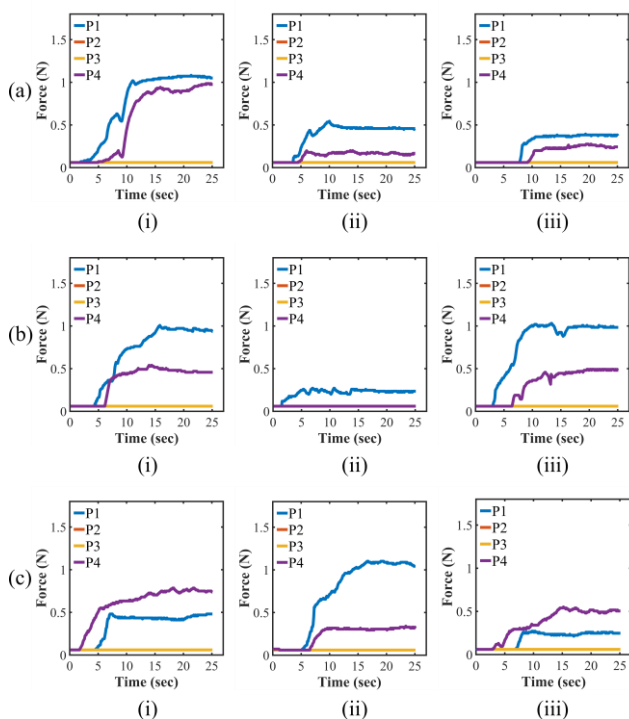


Figure 10. Transient force values while grasping (a) an apple, (b) a tennis ball, and (c) a cylindrical shaped object for (i) finger 1, (ii) finger 2, and (iii) finger 3 accordingly.

4 Conclusions

In recapitulation, we have demonstrated a novel, basic method to reconstruct the shape of a soft-robotic gripper in general 3D coordinate utilizing only two piezo-resistive flex sensors for each finger. The method is proved to be pragmatic and cost-effective. We also take force data from the finger. In the future scope of the study, we hope to combine shape and force data and develop an

algorithm to predict the shape of the objects manipulated by the gripper. Such a study would help towards multi-modal sensing-based object recognition for industries essentially contingent on robotic platform.

Acknowledgments

This research work is supported by UMass Dartmouth's Marine and Undersea Technology (MUST) Research Program funded by the Office of Naval Research (ONR) under Grant No. N00014-20-1-2170.

References

- [1] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft Robotic Grippers," *Advanced Materials*, vol. 30, no. 29, p. 1707035, 2018, doi: 10.1002/adma.201707035.
- [2] G. M. Whitesides, "Soft Robotics," *Angewandte Chemie International Edition*, vol. 57, no. 16, pp. 4258–4273, 2018, doi: <https://doi.org/10.1002/anie.201800907>.
- [3] S. Kim, C. Laschi, and B. Trimmer, "Soft robotics: a bioinspired evolution in robotics," *Trends in Biotechnology*, vol. 31, no. 5, pp. 287–294, May 2013, doi: 10.1016/j.tibtech.2013.03.002.
- [4] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, Art. no. 7553, May 2015, doi: 10.1038/nature14543.
- [5] C. Lee *et al.*, "Soft robot review," *Int. J. Control Autom. Syst.*, vol. 15, no. 1, pp. 3–15, Feb. 2017, doi: 10.1007/s12555-016-0462-3.
- [6] Z. Jing, L. Qiao, H. Pan, Y. Yang, and W. Chen, "An overview of the configuration and manipulation of soft robotics for on-orbit servicing," *Sci. China Inf. Sci.*, vol. 60, no. 5, p. 050201, Apr. 2017, doi: 10.1007/s11432-016-9033-0.
- [7] M. Cianchetti, C. Laschi, A. Menciassi, and P. Dario, "Biomedical applications of soft robotics," *Nat Rev Mater*, vol. 3, no. 6, pp. 143–153, Jun. 2018, doi: 10.1038/s41578-018-0022-y.
- [8] R. A. S. I. Subad, L. B. Cross, and K. Park, "Soft Robotic Hands and Tactile Sensors for Underwater Robotics," *Applied Mechanics*, vol. 2, no. 2, Art. no. 2, Jun. 2021, doi: 10.3390/applmech2020021.
- [9] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands — Review," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, Dec. 2015, doi: 10.1016/j.robot.2015.07.015.
- [10] Y. Liu, R. Bao, J. Tao, J. Li, M. Dong, and C. Pan, "Recent progress in tactile sensors and their applications in intelligent systems," *Science Bulletin*, vol. 65, no. 1, pp. 70–88, Jan. 2020, doi: 10.1016/j.scib.2019.10.021.
- [11] G. Sun, Y. Wu, H. Li, and L. Zhu, "3D shape sensing of flexible morphing wing using fiber Bragg grating

- sensing method," *Optik*, vol. 156, pp. 83–92, Mar. 2018, doi: 10.1016/j.ijleo.2017.10.159.
- [12] T. L. T. Lun, K. Wang, J. D. L. Ho, K.-H. Lee, K. Y. Sze, and K.-W. Kwok, "Real-Time Surface Shape Sensing for Soft and Flexible Structures Using Fiber Bragg Gratings," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1454–1461, Apr. 2019, doi: 10.1109/LRA.2019.2893036.
- [13] H. Wang, R. Zhang, W. Chen, X. Liang, and R. Pfeifer, "Shape Detection Algorithm for Soft Manipulator Based on Fiber Bragg Gratings," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 6, pp. 2977–2982, Dec. 2016, doi: 10.1109/TMECH.2016.2606491.
- [14] I. M. Di Napoli and C. M. Harwood, "Design and benchmarking of a robust strain-based 3D shape sensing system," *Ocean Engineering*, vol. 201, p. 107071, Apr. 2020, doi: 10.1016/j.oceaneng.2020.107071.

Distance Teaching and Learning for Artificial Intelligence in Engineering Education at University

Jaehyun Lee*

MathWorks
521 Teheran-ro Gangnam-gu, Seoul, Korea

*jaehyunl@mathworks.com

Abstract –Data-driven artificial intelligence affects modeling technologies in specific fields such as computer science and across traditional engineering fields. In addition, the current global pandemic situation is accelerating the online-based work and education environment. Therefore, online-based artificial intelligence education is becoming increasingly important in university education. This paper presents artificial intelligence as a workflow and proposes virtual laboratory-based education as a primary discipline.

Keywords: Artificial Intelligence, Virtual Laboratory, Distance Teaching and Learning.

1 Introduction

Artificial intelligence has been the most popular field in the industry, and its importance is rapidly increasing due to its applications in various engineering fields. However, according to a recent survey of 3,000 companies conducted by Gartner, only 4% of the 50% of Artificial intelligence plans were implemented.[1] This report intimates that artificial intelligence must overcome many obstacles to be effective in the industry. Among these difficulties, the most crucial task for companies is to secure and develop human resources to perform end-to-end artificial intelligence processes.[2]

Universities are also establishing specialized curricula from introductory theory courses to advanced application courses or graduate courses for education on artificial intelligence. The specialized education focuses on designing and analyzing various models from data. However, artificial intelligence education in the traditional engineering fields should focus on comprehensive applied education that should be implemented repeatedly in the real environment through data acquisition by experiment and verification in analysis and modeling. In addition, the recent transition to online-based education should also be considered.

This paper introduces a virtual experiment-based and overall workflow-oriented education method to incorporate artificial intelligence education into traditional engineering fields.

1.1 Workflows in Artificial Intelligence

The definition of Artificial Intelligence coined in the 1950s and still in use is "the capability of a machine to

imitate intelligent human behavior." When people think of Artificial Intelligence nowadays, they almost always mean machine learning: training a machine to learn the desired behavior.[3] In traditional programming, writing a program that processes data to produce the desired output can be evaluated. Figure 1 shows the overall classifications of Artificial Intelligence.

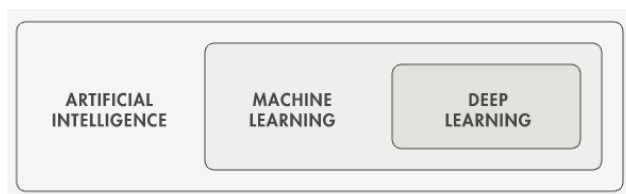


Figure 1. Configurations of Artificial Intelligence

With machine learning, the steps are reversed, for instance, feeding in both data and the desired output, and the computer writes the program. Machine learning programs (or, more accurately, models) are mainly black boxes. They can generate the desired output, but they are not composed of a sequence of operations like a traditional program or algorithm.

There is much movement today about a specialized type of machine learning called deep learning. Deep learning uses neural networks. (The term "deep" refers to the number of layers in the network—the more layers, the deeper the network.) One key advantage of deep learning is that it removes manual data processing steps and extensive domain knowledge required for other techniques. To put the key terms into context, think of machine learning and deep learning as ways of achieving Artificial Intelligence—they are the most common techniques applied today.

However, an Artificial Intelligence model is just one part of the complete workflow for developing a fully engineered system.



Figure 2. Workflows in AI for Engineering

2 Virtual-Laboratory for AI

In order to incorporate artificial intelligence into the traditional engineering field, the overall so-called End-to-End workflow, such as Figure 2, need then be carried out, especially in the engineering field, throughout experiments. However, securing a laboratory and conducting classes are costly and constrained. This paragraph introduces how to efficiently prepare and execute the entire workflow through a virtual laboratory environment in simulation.

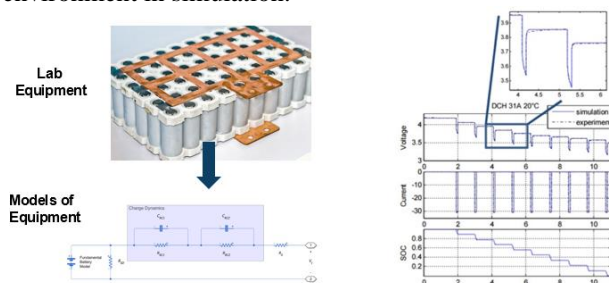


Figure 3. Building Virtual Lab based on Physical Experiment

2.1 Physical Modeling based on Experiment

The advantages of virtual laboratories over physical laboratories are cost reduction and increased accessibility. When a virtual laboratory is prepared based on data from a physical laboratory, the experiment can be conducted without space and time constraints, and data can be secured.

2.2 Data Analytics using Common Models

Data-based artificial intelligence requires a repetitive modeling process rather than a one-way process. To this end, to conduct multi-dimensional analysis and rapid prototyping using standard models that have already been developed.

2.3 Implementation onto a real experiment

Models based on virtual laboratories can be applied to the engineering field only when executable codes are produced. Most engineering fields educate these basic programming, but it takes a lot of time and effort to apply artificial intelligence model codes to various hardware. Therefore, autogenerated code can overcome obstacles to

applying artificial intelligence by reviewing the effectiveness of the current model.

2.4 Monitoring signals in Virtual Lab.

Using a virtual laboratory, various internal values that cannot be obtained through a physical laboratory sensor can be identified using variables. Internal physical monitoring also helps improve engineering expertise, which allows us to learn insights into more implementable practical artificial intelligence applications.

2.5 Distance Teaching and Learning

Due to the recent pandemic, many universities are accelerating their transition to online classes. In order to provide communication and immediate feedback effects, which are the most significant advantages of on-site education, in non-face-to-face education, more interactive educational materials and cloud-based tools should be integrated into lectures. The process of students trying to learn about theoretical classes directly on the web browser and checking the results through simple virtual experiments has a significant impact on encouraging students to participate in classes during online classes actively.

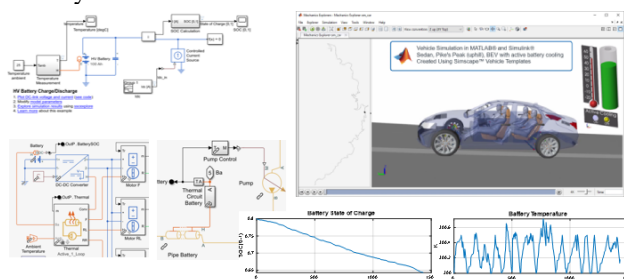


Figure 4. Virtual Experiment in Online using Web Browser

3 Conclusions

This paper introduced workflows for utilizing virtual laboratories to propagate artificial intelligence education into the engineering field and online-based education methods.

Acknowledgements

This paper has been supported by MathWorks

References

- [1] "The Real Truth of Artificial Intelligence." Presented at Gartner Data & Analytics Summit, March 2018.
- [2] "AI and ML Development Strategies, Motivators and Adoption Challenges," Gartner Research Note, published 19 June 2019
- [3] <https://doi.org/10.1016/j.bushor.2018.08.004>

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Analytical Model for Circuit Parameter Estimation of Permanent Magnet Synchronous Motor for EV traction |
| Name | Min-Ro Park |
| Affiliation | Interactive Robotics R&D Division, Korea Institute of Robotics & Technology Convergence |
| Address | 30 Haean-ro, 1106beon-gil, Buk-gu, Pohang-si, Gyeongbuk, Pohang 37553, Korea |
| E-mail | minro@kiro.re.kr |

Analytical Model for Circuit Parameter Estimation of Permanent Magnet Synchronous Motor for EV traction

Min-Ro Park*

Interactive Robotics R&D Division, Korea Institute of Robotics & Technology Convergence
30 Haean-ro, 1106beon-gil, Buk-gu, Pohang-si, Gyeongbuk, Pohang 37553, Korea

*Corresponding Author: minro@kro.re.kr

Abstract - This paper proposes an analytical model for the parameter estimation of the permanent magnet synchronous motor (PMSM) considering the magnetic nonlinearity. The proposed analytical model consists of a simple equivalent magnetic circuit (EMC) and equations. At this time, the simple EMC compensates for the limitation of the equation, with which is difficult to consider the magnetic nonlinearity. From the proposed EMC, the stator reluctance is calculated, which reflects nonlinearity for parameter calculation. Then the parameters are calculated mathematically through the proposed equation, which reflect the calculated stator reluctance. Finally, the proposed analytical model is verified via finite element analysis (FEA). As a result, the maximum errors of the back electromotive force and the inductance are smaller than 5%.

Keywords: Back electromagnetic force, inductance, permanent magnet synchronous motor.

1 Introduction

As the permanent magnet synchronous motor (PMSM) has good power density and power factor, it has been under study for many years over various fields [1], [2]. When designing the PMSM, it is an important step to accurately estimate the circuit parameters of the PMSM. The reason is that motor performances such as torque and efficiency are determined by circuit parameters that are resistance, back electromotive force (EMF) and inductance. The parameters are hard to estimate, however, due to the nonlinearity of the magnetic material.

Various researches have been conducted to calculate the circuit parameters of the PMSM. In [3], [4], the parameters were estimated using finite element analysis (FEA). However, this method is time-consuming due to the computational time as well as the pre-processing time to create the simulation model. In [5-7], the parameters were estimated using the equivalent magnetic circuit (EMC). The EMC shown in these papers, however, is too complicated. These complex EMCs are not intuitive and difficult to apply to various models. In addition, the conventional analytical models do not consider the nonlinearity of the magnetic material [8-11].

Thus, this paper proposes an analytical method for estimating the parameter and performance of the electric motor considering the nonlinearity of the magnetic material. The proposed analytical model consists of the simple EMC and equation. First, the reluctance is calculated from the simple EMC considering the nonlinearity of the magnetic material. Next, the circuit

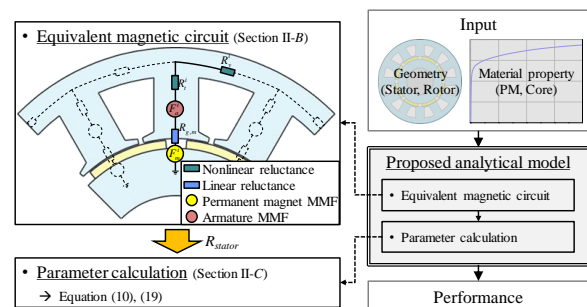


Fig. 1. Calculation process using analytical model.

parameters are calculated through the proposed equation reflecting the calculated reluctance. Also, the characteristics of the electric motor are evaluated through the calculated parameters. Finally, the proposed analytical model is verified by comparison with FEA.

2 Analytical Model considering Nonlinearity

In this paper, an analytical model that combines the EMC and equations is proposed to calculate the circuit parameters of electric motor. The process of calculating the parameters using the proposed analytical model is shown in Fig. 1.

First, a simple EMC is proposed, which distributes the magneto-motive force (MMF) of the field and armature to each tooth. The advantages of the proposed EMC are as follows.

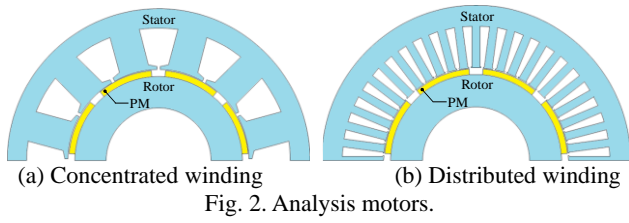


Fig. 2. Analysis motors.

Table 1. Dimension and Material of Analysis Motors

| Item | Concentrated winding | Distributed winding | Unit |
|--|----------------------|---------------------|------|
| Pole/Slot ($2p/S$) | 8 / 12 | 8 / 48 | - |
| Stator diameter (D_s) | 260 | | mm |
| Rotor diameter (D_r) | 154 | | mm |
| Axial length (L_{stk}) | 40 | | mm |
| Air-gap length (g) | 0.8 | | mm |
| Series turn number per phase (N_s) | 144 | | mm |
| PM thickness (t_m) | 5 | | mm |
| Pole angle ($2\alpha_m/p$) | 36 | | ° |
| Core material | 35PN230 | | - |

1) As shown in Fig. 1, as the MMF and the reluctance of the stator, air-gap, and permanent magnet (PM) are respectively distributed to each tooth in the proposed EMC, the proposed method can be applied not only to integer slot winding but also to fractional slot winding.

2) As the PM and armature MMFs according to the current and current phase angle are distributed to each tooth, the reluctance of each tooth and yoke can be calculated considering self- and cross-saturation.

The proposed EMC calculates the reluctance considering the nonlinearity of the magnetic material.

Second, the inductance and back EMF, which are circuit parameters, are calculated to reflect the reluctance calculated in the previous step.

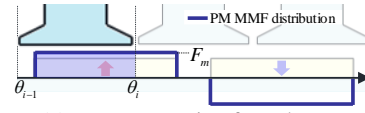
Finally, the performance of the motor is estimated using the calculated parameters.

2.1 Analysis motors

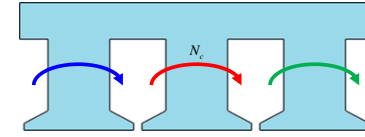
The analysis motors of the surface-mounted permanent magnet (SPM) type are shown in the Fig 2. These motors are the concentrated winding (CW) with 3 phases, 8 poles, 12 slots and the distributed winding (DW) with 3 phases, 8 poles 48 slots. The dimensions and material property for analysis are shown in the Table I. The proposed analytical model is used to calculate the circuit parameters and performance of the analysis motors. Also, the proposed analytical model is verified by FEA.

2.2 Simple equivalent magnetic

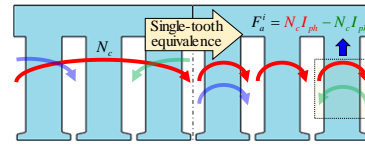
As described above, the feature of the proposed EMC converts all the MMFs into the MMF per tooth. The advantage of this method is that the EMC is simply



(a) Magnetomotive force by PM



(b) Magnetomotive force by concentrated winding



(c) Magnetomotive force by distributed winding

Fig. 3. Magnetomotive force applied to each tooth.

configured regardless of the pole-slot combination. In other words, unlike the conventional method of dividing the air-gap reluctance into a large number of parts considering the relative positions of the rotor and stator, the proposed EMC can be configured simply.

First, the method of calculating the MMF per tooth is shown in Fig. 3. The MMF of the PM applied to the i -th tooth is calculated by integrating the MMF distribution in the section corresponding to the tooth as shown in (1) and (2).

$$F_m = \frac{B_r}{\mu_0 \mu_{rec}} t_m \quad (1)$$

$$F_m^i = \frac{1}{\theta_i - \theta_{i-1}} \int_{\theta_{i-1}}^{\theta_i} F_m(\theta) d\theta \quad (2)$$

where B_r is the residual induction of the PM, μ_{rec} is the recoil permeability of the PM, μ_0 is the permeability of the vacuum, and t_m is the thickness of the PM.

Also, the armature MMF applied to the i -th tooth is calculated by summing the MMF of the coils wound around the i -th tooth. At this time, in case of the distributed winding (The slot pitch of the coils is 2 or more.), each coil is converted into a single-tooth coil as shown in Fig. 3 (c). The armature MMF applied to the i -th tooth is expressed as follow:

$$F_{ph} = \pm N_c I_{ph} \quad (3)$$

$$F_a^i = \sum_{ph=A}^C F_{ph} \quad (4)$$

where, N_c is the turn number of the armature coil, I_{ph} is the armature phase current, F_{ph} is the MMF of the armature coil.

Second, the reluctances of the stator teeth, yokes, air-gap, and PM are evaluated considering the hexahedron shape [12]. The reluctances of the iron core are calculated

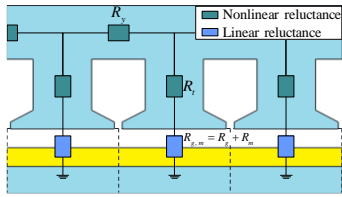


Fig. 4. Air-gap and PM reluctances equalized to the stator teeth.

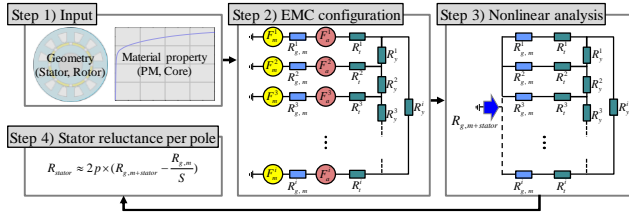


Fig. 5. Considered nonlinearity through proposed EMC.

considering the magnetic nonlinearity according to the load conditions. The reluctances of the air-gap and PM are converted to reluctance per stator tooth as shown in the Fig. 4. The permanent magnet is assumed to be full pitch only when calculating the reluctance of the permanent magnet. Because the recoil permeability of the permanent magnet and the relative permeability of the air are almost the same. The proposed EMC is composed of the MMF and reluctance converted to each tooth.

The proposed EMC is used to calculate the stator reluctance per pole (R_{stator}) considering the magnetic nonlinearity of the stator. The process is shown in the Fig. 5 and described as follow:

Step 1) The geometry and material property of the electric motor are input.

Step 2) The proposed EMC is constructed.

Step 3) The nonlinear analysis is used to estimate the reluctance for each part [12]. Also, the total reluctance ($R_{g,m+stator}$) of the electric motor is calculated.

Step 4) Using (5), the stator reluctance per pole (R_{stator}), which takes into account the magnetic nonlinearity, is calculated.

$$R_{stator} \approx 2p \times \left(R_{g,m+stator} - \frac{R_{g,m}}{S} \right) \quad (5)$$

where, p is the pole-pair number, S is the slot number, R_g and R_m are the reluctance per tooth of the air-gap and PM, respectively.

To verify the proposed EMC, the magnetic flux densities obtained from the proposed EMC and FEA are compared in Fig. 6. Compared with the FEA results, an about 5% error is shown, and the magnetic flux density tendency according to the load condition coincides with the FEA result. The proposed EMC is thus reasonable. Therefore, the nonlinear reluctances considering the magnetic nonlinearity are calculated from the proposed EMC. In the next section, the calculated stator reluctance R_{stator} is used to calculate the parameters taking into account magnetic nonlinearity.

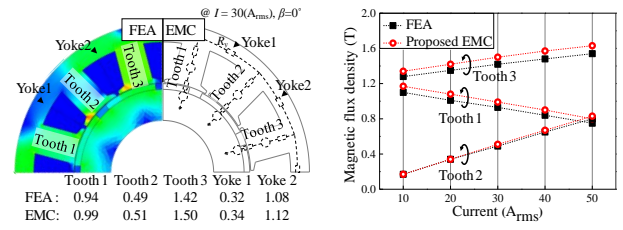
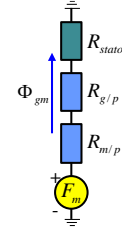
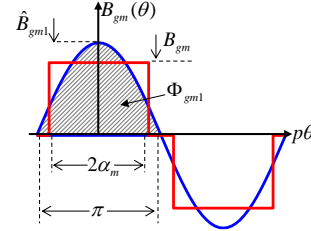


Fig. 6. Magnetic flux distribution according to load condition.



(a) Equivalent magnetic circuit per pole



(b) Field distribution

Fig. 7. Field distribution of magnetic loading.

2.3 Circuit parameter considering nonlinearity

In order to evaluate the correct characteristics of the electric motor, the circuit parameters must be calculated accurately. In particular, the flux linkage by field and inductance are affected by the nonlinearity of the magnetic material. Therefore, the circuit parameters are calculated through an equation that reflects the previously estimated stator reluctance.

First, the flux linkage is derived as follows through Fig. 7. The fundamental component (Φ_{gm1}) of the air-gap flux per pole (Φ_{gm}) is obtained from the MMF of PM (F_m) and distribution of the air-gap flux density.

$$R_{g/p} = \frac{g}{\mu_0 \alpha_m D_r L_{stk} / p}, \quad R_{m/p} = \frac{t_m}{\mu_0 \mu_{rec} h_m L_{stk}} \quad (6)$$

$$B_{gm} = \frac{\Phi_{gm}}{A_g} = \frac{F_m}{R_g + R_m + R_{stator}} \frac{1}{\alpha_m D_r} \quad (7)$$

$$B_{gm1}(\theta) = \hat{B}_{gm1} \cos p\theta, \quad \hat{B}_{gm1} = \frac{4}{\pi} B_{gm} \sin \alpha_m \quad (8)$$

$$\Phi_{gm1} = \frac{D_r L_{stk}}{2} \int_{-\pi/2p}^{\pi/2p} B_{gm1}(\theta) d\theta = \frac{4}{\pi} \frac{D_r L_{stk}}{p} B_{gm} \sin p\alpha_m \quad (9)$$

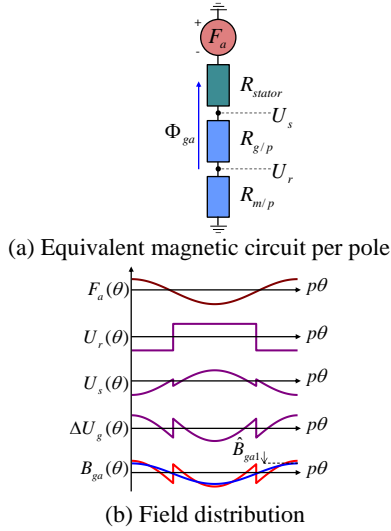


Fig. 8. Field distribution of electric loading.

where, $R_{g/p}$ and $R_{m/p}$ are the reluctance per pole of the air-gap and PM, respectively, g is the effective air-gap length, D_r is the rotor diameter, L_{stk} is the stack length, α_m is half the pole angle in electric angle, h_m is the width of the PM, B_{gm1} is the fundamental component of the air-gap flux density.

Therefore, the flux linkage by field (Ψ_m) and back EMF (E) are expressed by (10), (11).

$$\Psi_m = k_w N_s \Phi_{gm1} = \frac{4}{\pi} k_w N_s \frac{D_r L_{stk}}{p} \sin \alpha_m \frac{F_m}{R_g + R_m + R_{stator}} \frac{1}{\alpha_m D_r} \quad (10)$$

$$E = \omega_e \Psi_m \quad (11)$$

where, ω_e is the electric angular velocity.

Second, the inductance is derived from Fig. 8 and the following equation. The 3 phase armature MMF per pole is shown in (12).

$$F_a(\theta) = \hat{F}_a \cos p\theta, \quad \hat{F}_a = \frac{3}{\pi} \frac{k_w N_s}{p} \hat{I} \quad (12)$$

where, k_w is the winding factor, N_s is the series turn number per phase.

Also, the magnetic scalar potentials (U) of the rotor and stator are shown in (13), (14).

$$U_r = \frac{-R_m}{R_m + R_g + R_{stator}} F_{a(av)} \quad (13)$$

$$U_s = \frac{R_{stator}}{R_m + R_g + R_{stator}} F_{a(av)} - F_a(\theta) \quad (14)$$

where, the subscripts r and s of the magnetic scalar potentials (U) are rotor and stator, respectively, $F_{a(av)}$ is the average value of the 3 phase armature MMF per pole.

The distribution of the air-gap flux density is obtained from (12)-(14) and the fundamental component of this distribution is calculated using the Fourier series expansion.

$$B_{ga}(\theta) = \frac{\mu_0}{g} (U_r - U_s) \quad (15)$$

$$B_{ga1}(\theta) = \hat{B}_{ga1} \cos p\theta \quad (16)$$

$$\begin{aligned} \hat{B}_{ga1} &= \frac{4p}{\pi} \int_0^{\frac{\pi}{2p}} B_{ga}(\theta) \sin p\theta d\theta \\ &= \frac{3}{\pi} \frac{\mu_0 k_w N_s \hat{I}}{p g} \left(1 - \frac{8}{\pi^2} \frac{R_m + R_{stator}}{R_m + R_g + R_{stator}} \right) \end{aligned} \quad (17)$$

The fundamental component (Φ_{ga1}) of the air-gap flux per pole is expressed by (18).

$$\begin{aligned} \Phi_{ga1} &= \frac{D_r L_{stk}}{2} \int_{-\frac{\pi}{2p}}^{\frac{\pi}{2p}} B_{ga1}(\theta) d\theta \\ &= \frac{3}{\pi} \frac{\mu_0 k_w N_s \hat{I}}{p^2 g} D_r L_{stk} \left(1 - \frac{8}{\pi^2} \frac{R_m + R_{stator}}{R_m + R_g + R_{stator}} \right) \end{aligned} \quad (18)$$

Thus, the flux linkage by armature (λ) and inductance (L) are as follow.

$$\lambda = k_w N_s \Phi_{ga1} \quad (19)$$

$$\begin{aligned} L &= \frac{\lambda}{\hat{I}} \\ &= \frac{3}{\pi} \frac{\mu_0 D_r L_{stk}}{g} \left(\frac{k_w N_s}{p} \right)^2 \left(1 - \frac{8}{\pi^2} \frac{R_m + R_{stator}}{R_m + R_g + R_{stator}} \right) \end{aligned} \quad (20)$$

The circuit parameters (10) and (20) are calculated considering the saturation of the magnetic material through the stator reluctance obtained from the EMC. In section III, the proposed analytical method is validated through the FEA.

3 Result and vaidation

Using the equations proposed above, the circuit parameters that determine the performance of the electric motor are calculated. The proposed equation reflects the nonlinearity of the magnetic material through the stator reluctance calculated from simple EMC. The current is input in 10 Arms intervals from 10 Arms to 50 Arms, and the back-EMF and inductance calculated for each current are compared with the result calculated from FEA.

The results obtained from the proposed equations are shown in Fig. 9. Compared to the FEA results, it can be observed that the trends for each parameter are identical according to the load condition. The maximum errors of the back EMF and the inductance are smaller than 5%, which are acceptable.

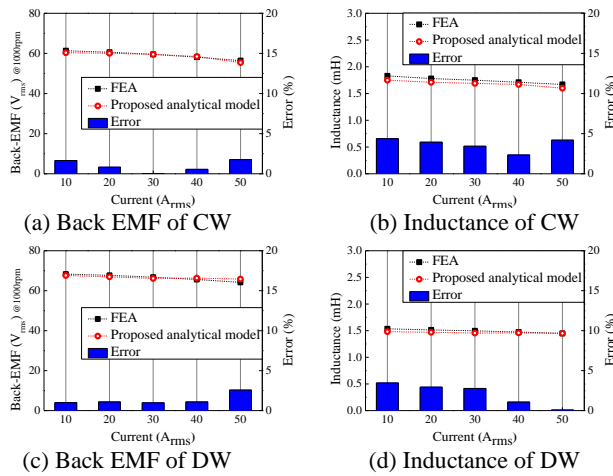


Fig. 9. Circuit parameter.

4 Conclusions

This paper proposes an analytical model for calculating the parameters of the PMSM considering the magnetic nonlinearity. First, a simple EMC is proposed to consider the magnetic nonlinearity. The stator reluctance for the parameter estimation is calculated from the proposed EMC. Second, the parameter equations reflecting the calculated stator reluctance are derived. In addition, the proposed analytical model combining the EMC and the equations is verified via FEA.

References

- [1] H. J. Kim, D. Y. Kim and J. P. Hong, "Structure of Concentrated-Flux-Type Interior Permanent-Magnet Synchronous Motors Using Ferrite Permanent Magnets," *IEEE Trans. Magn.*, vol. 50, no. 11, pp. 1-4, Nov. 2014.
- [2] M. Park, H. Kim, Y. Choi, J. Hong and J. Lee, "Characteristics of IPMSM According to Rotor Design Considering Nonlinearity of Permanent Magnet," *IEEE Trans. Magn.*, vol. 52, no. 3, pp. 1-4, Mar. 2016.
- [3] W. Zhao, X. Wang, C. Gerada, H. Zhang, C. Liu and Y. Wang, "Multi-Physics and Multi-Objective Optimization of a High Speed PMSM for High Performance Applications," *IEEE Trans. Magn.*, vol. 54, no. 11, pp. 1-5, Nov. 2018.
- [4] M. Park, J. Jung, D. Kim, J. Hong and M. Lim, "Design of High Torque Density Multi-Core Concentrated Flux-Type Synchronous Motors Considering Vibration Characteristics," *IEEE Trans. Ind. Appl.*, vol. 55, no. 2, pp. 1351-1359, Mar.-Apr. 2019.
- [5] G. Xu, G. Liu, S. Jiang and Q. Chen, "Analysis of a Hybrid Rotor Permanent Magnet Motor Based on Equivalent Magnetic Network," *IEEE Trans. Magn.*, vol. 54, no. 4, pp. 1-9, Apr. 2018.
- [6] A. Hemeida and P. Sergeant, "Analytical Modeling of Surface PMSM Using a Combined Solution of Maxwell's Equations and Magnetic Equivalent Circuit," *IEEE Trans. on Magn.*, vol. 50, no. 12, pp. 1-13, Dec. 2014.
- [7] G. Liu, L. Ding, W. Zhao, Q. Chen and S. Jiang, "Nonlinear Equivalent Magnetic Network of a Linear Permanent Magnet Vernier Machine With End Effect Consideration," *IEEE Trans. Magn.*, vol. 54, no. 1, pp. 1-9, Jan. 2018.
- [8] S. Teymoori, A. Rahideh, H. Moayed-Jahromi and M. Mardaneh, "2-D Analytical Magnetic Field Prediction for Consequent-Pole Permanent Magnet Synchronous Machines," *IEEE Trans Magn.*, vol. 52, no. 6, pp. 1-14, Jun. 2016.
- [9] H. Chen, D. Li, R. Qu, Z. Zhu and J. Li, "An Improved Analytical Model for Inductance Calculation of Interior Permanent Magnet Machines," *IEEE Trans. Magn.*, vol. 50, no. 6, pp. 1-8, Jun. 2014.
- [10] K. Shin, H. Park, H. Cho and J. Choi, "Analytical Calculation and Experimental Verification of Cogging Torque and Optimal Point in Permanent Magnet Synchronous Motors," *IEEE Trans. Magn.*, vol. 53, no. 6, pp. 1-4, June 2017.
- [11] K. Shin, H. Cho, K. Kim, K. Hong and J. Choi, "Analytical Investigation of the On-Load Electromagnetic Performance of Magnetic-Geared Permanent-Magnet Machines," *IEEE Trans Magn.*, vol. 54, no. 11, pp. 1-5, Nov. 2018.
- [12] J. Sim, D. Ahn, D. Kim and J. Hong, "Three-Dimensional Equivalent Magnetic Circuit Network Method for Precise and Fast Analysis of PM-Assisted Claw-Pole Synchronous Motor," *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 160-171, Jan.-Feb. 2018.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Surrogate Model Based Design Optimization of Ultra-High-Speed Motor Stator Shape to Minimize Loss |
| Name | Sun-Yong Shin*, Ji-Hyeon Lee*, Dong-Min Kim**, Myung-Seop Lim*** |
| Affiliation | Department of Automotive Engineering (Automotive-Computer Convergence), Hanyang University* Department of Automotive Engineering, Honam University** Department of Automotive Engineering, Hanyang University*** |
| Address | 222 Wangsimni-ro, Sungdong-gu, Seoul 04763, Korea |
| E-mail | myungseop@hanyang.ac.kr |

Surrogate Model Based Design Optimization of Ultra-High-Speed Motor Stator Shape to Minimize Loss

Sun-Yong Shin*, Ji-Hyeon Lee*, Dong-Min Kim**, Myung-Seop Lim***

Department of Automotive Engineering (Automotive-Computer Convergence), Hanyang University*

Department of Automotive Engineering, Honam University**

Department of Automotive Engineering, Hanyang University***

222, Wangsimni-ro Seongdong-gu, Seoul, Republic of Korea

*Myung-Seop Lim: myungseop@hanyang.ac.kr

Abstract - In this paper, to determine the stator shape that minimizes the loss of the ultra-high-speed permanent magnet synchronous motor, an optimal design based on the Kriging surrogate model was performed. Design variables that have a statistically significant effect on loss were selected through Analysis of variance. Sampling points were determined by Latin Hypercube Sampling for the selected design variables, and finite element analysis was performed at the determined sampling points to generate a Kriging surrogate model. Then, the validity of the generated Kriging surrogate model was verified using the cross-validation method. Optimal stator shape that minimizes loss was obtained by performing optimization to minimize loss using a validated Kriging surrogate model.

Keywords: Ultra-high-speed, Design variables, Sampling points, ANOVA, Kriging surrogate model

1 Introduction

Recently, Ultra-high-speed motors using permanent magnets, which are used for various purposes, are light, small in volume, and have high efficiency, so related research is being actively conducted. The types of losses in Ultra-high-speed permanent magnet synchronous motors including iron loss, copper loss, eddy current loss, and wind loss are greatly affected by the stator shape. There are many cases where optimization techniques are applied to reduce the loss of such Ultra-high-speed motors [1], [2]. In particular, there is a trend to use a Response Surface Model (RSM) to replace Finite Element Analysis (FEA) of motor performance or to use genetic algorithms as an optimization algorithm. RSM is a method of minimizing the difference between the FEA results at the sampling points and the predicted value using a surrogate model. It is somewhat less accurate when predicting. Also, in the case of genetic algorithms, there is a disadvantage that it takes a long time to optimize.

Therefore, in this paper, an optimal design based on the Kriging surrogate model was performed to determine the stator shape that minimizes the loss of the Ultra-high-speed motor. Design variables of stator necessary for the optimal design were selected through Analysis of variance (ANOVA), and a surrogate model for Kriging was created using these design variables and FEA results [3]. The validity of the generated Kriging surrogate model was verified using the cross-validation method, and a sensitivity-based optimization technique was used to minimize the loss [4]. As a result of the optimal design, the loss was reduced by 13.2% compared to the initial model.

2 Sensitivity Analysis and Optimal Design

2.1 Loss of Initial Model

The initial model was a three-phase, two-pole, six-slot, Surface-mounted Permanent Magnet Synchronous Motor (SPMSM), and the rotor was composed of a sleeve, permanent magnet, and a hollow shaft. In this paper, as the losses of the initial model, iron loss from the stator core and hollow shaft of the rotor, copper loss from the stator winding, eddy current loss from the rotor sleeve and permanent magnet, and wind loss from the air gap were considered.

2.2 Sensitivity Analysis Using Table of Orthogonal Array

Fig. 1 shows the six stator design variables and the table of orthogonal array for the sensitivity analysis. The design variables were tooth tip thickness (X1), shoe height (X2), slot opening width (X3), lower fillet radius of the slot (X4), upper fillet radius (X5), and yoke thickness (X6). A 3-level table of orthogonal array was used as the design of the experiment (DOE), and the motor loss according to the design variable at each sampling point was calculated using FEA. Sensitivity analysis using ANOVA was performed to select design variables that significantly affect the loss of the initial model. As shown in Fig. 2, the ratio of the intergroup variance (SSTr: sum of treatment) and residual (SSE: sum of error) of the design variable is defined as a measurement of ANOVA, and the p -value (p -value: probability value) was calculated F-

distribution. In this case, the measurement of ANOVA can be calculated as in Equation (1).

$$F_o = \frac{\frac{SSTr}{N_1 - 1}}{\frac{SSE}{(N_2 - 1) - q \cdot (N_1 - 1)}} \quad (1)$$

In this case, N_1-1 is the degree of freedom of each design variable, N_2-1 is the degree of freedom for the total number of experiments, and q is the number of design variables.

In this paper, design variables X1, X2, and X3 with p -values less than 0.05 were selected as loss-sensitive variables [5].

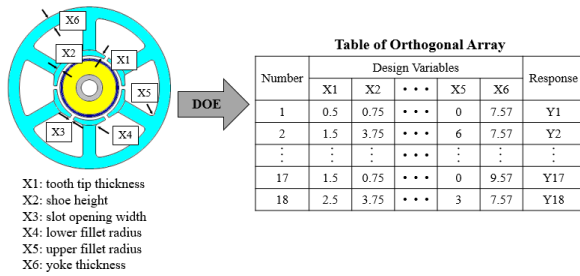


Figure 1. Design variables and Table of orthogonal array

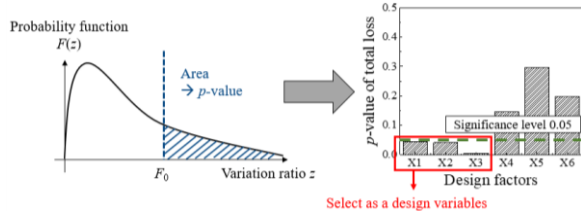


Figure 2. p -value graph of design variables according to F-distribution

2.3 Kriging Surrogate Model

As shown in Fig. 3, the Kriging surrogate model can increase the accuracy because the predicted value is equal to the FEA results at the sampling point compared to the RSM. In this paper, a Kriging surrogate model is created using the three design variables determined to have a significant effect on loss through sensitivity analysis. To this end, sampling points were extracted using Latin Hypercube Sampling (LHS), which considers all experimental areas without repetition with a few numbers of the experiments. The motor loss at each sampling point was calculated through FEA, and based on the FEA results, a Kriging surrogate model for the motor loss was created using a Gaussian correlation function.

A cross-validation method was used to verify the validity of the generated Kriging surrogate model, and it was confirmed that the model had a relatively small error of 1.9% based on the normalized root mean square error.

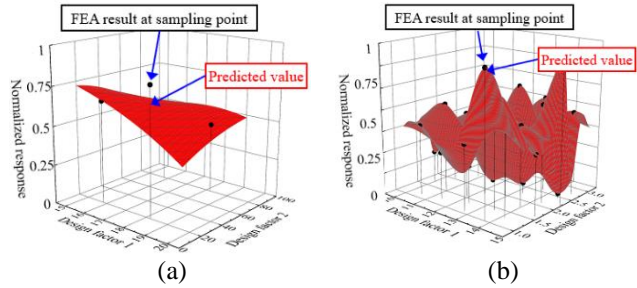


Figure 3. (a) RSM Example (b) Kriging Surrogate Model Example

2.4 Optimal Design of stator shape

Using the Kriging surrogate model created in Section 4, this section applies the sensitivity-based optimization technique to determine the stator shape to minimize loss. In order to minimize the loss while satisfying the output of 20kW or more, the objective function and constraints were defined as in Equation (2).

$$\begin{aligned} & \text{minimize } \hat{Y}_{loss} \\ & \text{subject to } \hat{Y}_{output} \geq 20kW \end{aligned} \quad (2)$$

Here, Y_{loss} is the motor loss predicted using the Kriging surrogate model, and Y_{output} is the motor output predicted using the Kriging surrogate model.

Fig. 4 shows the loss according to tooth tip thickness and shoe height, tooth tip thickness and slot opening width, respectively. It was found that the loss decreased as the tooth tip thickness decreased, the shoe height decreased, and the slot opening width decreased. Since the constraints were satisfied in all design area, the optimal design point was selected as the point with the minimum loss as shown in Fig. 4. Fig. 5 compares the loss and efficiency of the initial model and the optimal model. Compared to the initial model, it can be seen that the overall loss of the optimal model was reduced by 13.2%, and thus the motor efficiency increased by 0.57%p.

3 Conclusions

In this paper, to reduce the loss of the Ultra-high-speed permanent magnet synchronous motor, the optimal design of the stator shape was carried out. Iron loss, copper loss, eddy current loss, and wind loss were considered as losses occurring in the Ultra-high-speed permanent magnet synchronous motor, and a table of orthogonal array and ANOVA were used to identify design variables that significantly affect the total motor loss. Through this, among various design variables of the stator, tooth tip thickness, shoe height, and slot opening width were selected as design variables that had a significant effect on motor loss. Sampling points were determined using LHS for the selected design variables, and motor loss values at each sampling point were calculated using FEA. Through this, a Kriging surrogate model for motor loss was created, and the validity of the model generated through the cross-validation method was verified. The

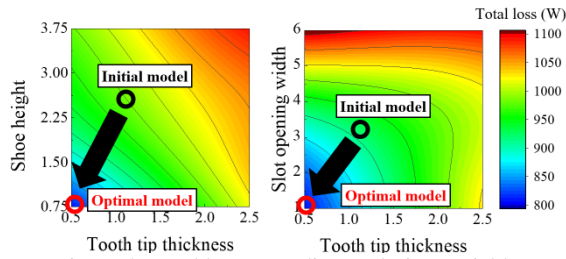


Figure 4. Total loss according to design variables

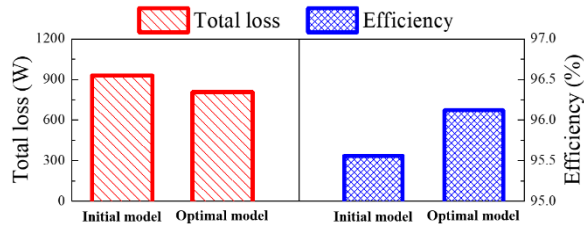


Figure 5. Comparison of Total loss and Efficiency of the initial model and optimal model

generated Kriging surrogate model was used to calculate the motor loss, which is the objective function, and the motor output, which is the constraint, in the optimal design. By applying the sensitivity-based optimization technique, it was possible to obtain an optimal model that satisfies more than the target output and minimized the loss.

Acknowledgements

This work was supported by the R&D program of MOTIE/KEIT. [20010986, Development of High-Power and High-Durability Air Compressor Technology for Medium and Heavy-Duty Commercial Vehicles]

References

- [1] Jin-Cheol Park, Nyeon-Han Hong, Sung-Woo Hwang, Seung-Hee Chai, Jung-Pyo Hong, Design for Reducing Field Induced Voltage and Torque ripple of WFSM for ISG using Response Surface Methodology, KIEE Annual Summer Conference & Exhibition, 806-807, 2015
- [2] Mingjie Li, Fuzhen Xing, Byung-il Kwon, Optimized design of a v-type consequent-pole motor for reducing permanent magnet usage, KIEE Annual Autumn Conference & Exhibition, 193-195, 2019
- [3] Saekyeol Kim, Su Kil Cho, Junyong Jang, Tae Hee Lee, Selection of Correlation Coefficients for Accurate Kriging Surrogate Model, KSME Autumn Conference & Exhibition, 936-937, 2014
- [4] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P. K. Tucker, Surrogate-based analysis and optimization, Prog. In Aerospace Sciences, vol. 41, no. 1, 1-8, 2005
- [5] Jae-Hyun Kim, Dong-Min Kim, Hyeon-Jin Park, Jung-Pyo Hong, Kyu-Sung Choi, Hyun-Sup Yang, Kyung-Seok Cho, Optimum Design for Efficiency of Ultra-high-speed Permanent Magnet Synchronous Motor for

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | x |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Comparative Study on the Optimal Design of Linear Oscillating Actuator using Kriging and Deep Neural Network |
| Name | Peel-Joong Kim*, Soo-Hwan Park**, Myun-Seop Lim**(Corresponding Author) |
| Affiliation | Department of Automotive Engineering (Automotive-Computer Convergence), Hanyang University* Department of Automotive Engineering, Hanyang University** |
| Address | 222, Wangsimni-ro Seongdong-gu, Seoul, Republic of Korea |
| E-mail | myungseop@hanyang.ac.kr |

Comparative Study on the Optimal Design of Linear Oscillating Actuator using Kriging and Deep Neural Network

Peel-Joong Kim, *Soo-Hwan Park, **Myung-Seop Lim**

Department of Automotive Engineering (Automotive-Computer Convergence), Hanyang University*
 Department of Automotive Engineering, Hanyang University**
 222, Wangsimni-ro Seongdong-gu, Seoul, Republic of Korea
 Myung-Seop Lim: myungseop@hanyang.ac.kr

Abstract – Linear Oscillating Actuator(LOA) is being used as a core motor for home appliances. Therefore, the optimal design of LOA is required to reduce unit price and improve performance of home appliances. The results of surrogate model-based optimal design depend on how well the surrogate model simulates the objective function. Therefore, in this paper, surrogate models are generated using kriging method and deep neural network(DNN), and normalized root mean square error(NRMSE) of each surrogate model and finite element analysis(FEA) results of optimal designed LOA models are compared. And by comparing each surrogate model generating time, this paper presents an surrogate model creation method suitable for LOA optimal design.

Keywords: Deep Neural Network, Kriging, Linear Oscillating Actuator, NRMSE

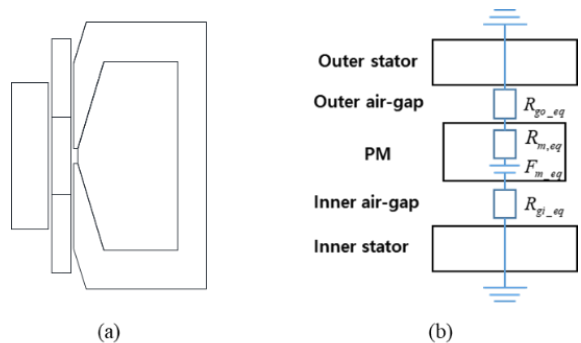
1 Introduction

The linear compressor [1], which is more efficient than the rotary compressor, is a key component of the home appliance industry. Recently, due to the rise in unit prices caused by the price of rare earth permanent magnets, the optimal shape of a linear oscillating actuator(LOA), which reduces magnet usage and performs above the same level, is required. The results of surrogate model-based optimal design depend on how well the surrogate model simulated the actual objective function. Therefore, it is necessary to consider an appropriate surrogate model generating method.

In this paper, the normalized root mean square error(NRMSE) of each surrogate model generated by kriging method [2] and deep neural network(DNN) is compared, and the predicted value of the surrogate model and the objective function value derived through finite element analysis(FEA) are compared. Additionally, the model generating time was compared. In addition, the main performance indicators of LOA, such as back-EMF constant, flatness of back-EMF, electromagnetic loss, and permanent magnet usage, which are the main reduction targets, were selected as objective functions.

figure. 1. [3] 5 Design variables that determine magnetic equivalent circuit components are chosen. Permanent magnet(PM) length and thickness, outer stator slot area and outer stator tooth angle which determines the shape of outer stator that is highly saturated were chosen for design variables. Additionally, magnet pole ratio that is known as important variable for LOA performance is also chosen as design variables. [4] And relationships between variables and magnetic equivalent circuit [5] components are shown in equation (1).

$$R_g = \frac{l_g}{\mu_0 l_m \pi D_g}, R_m = \frac{t_m}{\mu_0 \mu_c l_c \pi D_m}, F_m = \frac{B_r t_m}{\mu_0 \mu_c}, F_e = A_{slot} f f J \quad (1)$$



2 Surrogate model and analysis is conditions

2.1 LOA Model and design variables

3-poles 1-slot topology was adopted as shown in

Figure 1. LOA figure and equivalent magnetic circuit (a) LOA model figure, (b) Equivalent magnetic circuit

2.2 Optimization conditions

In order for the surrogate model to simulate the object functions properly, the experimental points must be evenly distributed within the design area. Therefore 300, 600, 900, and 1200 experimental points were extracted, using optimal latin hypercube sampling(OLHD). The main performance index of LOA were chosen as the objective function. The back-EMF constant, back-EMF flatness and electromagnetic loss expressed by the sum of iron loss and copper loss, and the permanent magnet usage, which is the main reduction target for unit price reduction. The weights for optimization were determined larger for more important object functions and the performances of reference model were set as constraints for optimization. Table 1 shows the analysis conditions and table 2 shows conditions for optimization.

Table 1. Analysis conditions.

| Parameter | Unit | Value |
|-----------------|------|-------|
| Rated Frequency | Hz | 60 |
| Stroke | Mm | 11 |

Table 2. Optimization conditions.

| Parameter | Weight | Criterion |
|-----------------------|--------|-----------|
| Back-EMF Constant | 0.2 | Maximize |
| Back-EMF Flatness | 0.1 | Minimize |
| Electro-magnetic Loss | 0.2 | Minimize |
| PM Volume | 0.5 | Minimize |

2.3 Surrogate model

In order to perform the optimal design, surrogate model that predicts the value of nonlinear objective functions is required. As surrogate model, kriging method [6], which is widely used as surrogate model for nonlinear functions, and DNN were selected and compared. The hyper parameter of DNN were tuned for each objective functions through searching the hyper parameters that minimize NRMSE. Number of layers and number of nodes for each layers of DNN were tuned and the results are shown in table 3.

Table 3. The results of hyper parameter tuning

| Parameter | Learning Rate | Node(layer) | | |
|----------------------|---------------|-------------|----|----|
| | | 1 | 2 | 3 |
| Back-EMF Constant | 0.001 | 36 | 36 | 24 |
| Flatness | 0.001 | 24 | 36 | 36 |
| Electromagnetic loss | 0.01 | 24 | 36 | 24 |

3 Result of the comparison

3.1 Comparison of NRMSE and model generating time of surrogate models

NRMSE of each number of data was extracted to compare the performance of each surrogate models. 90% of total data was used to generate model and 10% was used to validate models. As shown in Figure 2, NRMSE of two surrogate models reduce as the number of data increases. Model for each objective function with NRMSE less than 5% was generated with number of data over 600. Model generating time of kriging method increases sharply as the number of data rises. But DNN method can limit generating time by batch training.

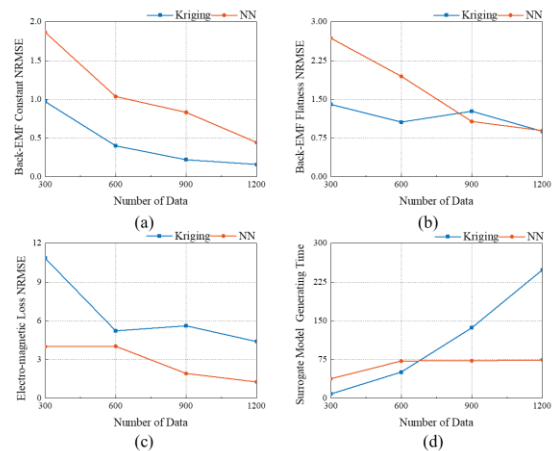


Figure 2. Comparison of NRMSE and model generating time for each objective functions

(a) Back-EMF constant NRMSE, (b) Back-EMF Flatness NRMSE, (c) Electro-magnetic loss NRMSE, (d) Time consumption

3.2 Comparison of optimization results

The same optimization process using sequential quadratic programming (SQP) algorithm was performed to the surrogate models of 1200 data [7]. And the figures of the optimized LOAs are shown in Figure 3. As shown in Figure 4 the performance of optimized LOA by surrogate models are improved compared to the reference model. However, even though the NRMSE of the surrogate model is all below 5 %, when compare prediction of surrogate models and FEA results, error was much higher than expected and this was noticeable for back-EMF flatness. This is because NRMSE does not accurately represent the NRMSE value of a specific design variable because it is the average value of total error. Due to this error, the performance of the optimal design model decreased compared to the prediction of the surrogate model. However, as a result, the two surrogate models resulted in an optimal figure that reduced permanent magnet usage

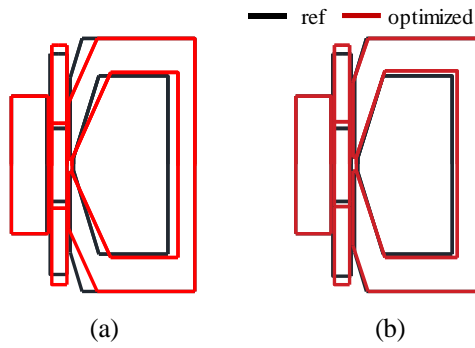


Figure 3. Optimized LOA
 (a) Back-EMF constant NRMSE, (b) Back-EMF Flatness NRMSE, (c) Electro-magnetic loss NRMSE, (d) Time consumption

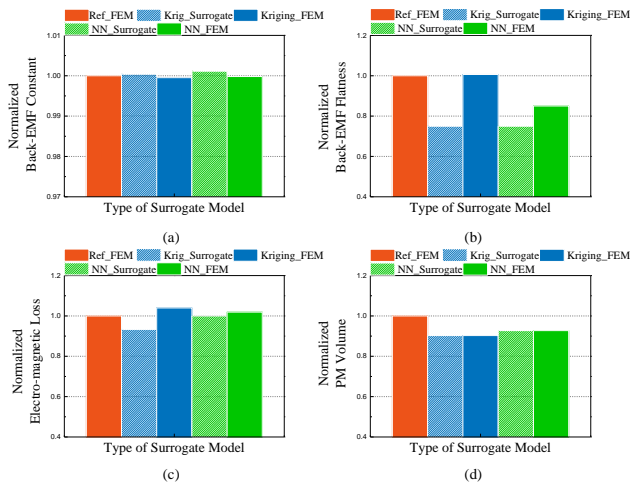


Figure 4. Comparison of surrogate model prediction and FEA results
 (a) Back-EMF constant, (b) Back-EMF Flatness, (c) Electro-magnetic loss, (d) Model generating time

While performing above the same level with reference model.

4 Conclusions

In this paper, the optimal design of LOA was performed by generating surrogate models by kriging method and DNN, and the NRMSE of each surrogate models and the objective functions were compared. In both kriging and DNN, the NRMSE tended to decrease as the number of data increased. But size relationship was different for each objective functions. However, it was confirmed that the model creation time of the kriging method increases exponentially as the number of data increases. Therefore, when the number of variables increases, when more experimental points are required,

DNN will be advantageous in terms of surrogate model generating time. As a result of comparing the objective function values through FEA, it was confirmed that the NRMSE of the alternative model does not represent the error in specific design variables. And through the two surrogate model generating methods, each optimal figure that satisfies the design conditions was derived.

References

- [1] Wang, D.Howe "Design Optimization of Short-Stroke Single-Phase Tubular Permanent Magnet Motor for Refrigeration Applications", IEEE Transactions on Industrial Electronics, 10.1109/TIE.2009.2025710,327 – 334, 19 June 2009
- [2] Bin Xia Ziyang Ren "Selecting proper kriging surrogate model for optimal design of electromagnetic problem" 9th IET international Conference on Computation in Electromagnetics, 10.1049/cp.2014.0226, 31 March-1 April 2014
- [3] X.chen Z.Q.Zhu, D.Howe "Comparative Study of Alternative Permanent Magnet Linear Oscillating Actuators" , 2008 international Conference on Electrical Machines and Systems, 10467577, 17-20 Oct.2008
- [4] X.chen Z.Q.Zhu "Analytical Determination of Optimal Split Ratio of E-Core Permanent Magnet Linear Oscillating Actuators", IEEE Transactions on Industry Applications,10.1109/TIA.2010.2090840,32,09November 2010
- [5] Liang Yan, Juanjuan Peng, "Magnetic field modeling of a linear permanent magnet motor with two isolated movers",IEEE,10.1109/CGNCC.2014.7077312,8,10August.2014
- [6] So-Yeon IM, "Kriging Surrogate Model-Based Design of an Ultra-High Speed Surface-Mounted Permanent-Magnet Synchronous Motor Considering Stator Iron Loss and Rotor Eddy Current Loss" IEEE TransactionsonMagnetics,10.1109/TMAG .2021.3080119, 1-1, 13 May 2021.
- [7] Yuehua Gao, "A Sequential Optimization Method Based on Kriging Surrogate Model", 2011 Fourth international Joint Conference on Computational Sciences and Ootimization, 10.1109/CSO.2011.56, 15- 19 April 2011

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | X |
| Oral Preferred | |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Professor |
| Name | Peter Xu |
| Affiliation | Department of Mechanical and Mechatronics Engineering, University of Auckland |
| Address | Auckland 1010, New Zealand |
| E-mail | p.xu@auckland.ac.nz |

A Ring-shaped Soft Robotic Actuator for Mimicking Gastric Motility

Shahab Kazemi, Weiliang Xu*

Department of Mechanical and Mechatronics Engineering, University of Auckland
Auckland 1010, New Zealand

p.xu@auckland.ac.nz

Abstract – Food and pharmaceutical industries demand *in-vitro* setups because they are cheaper and more efficient compared to *in-vivo* tests. Gastric simulators are *in-vitro* soft biomimetic modules reproducing the motility function of the stomach (so-called peristalsis). However, conventional open-loop approaches to controlling gastric simulators are not efficient when dealing with certain diseases or conditions that affect the stomach motility function. In this paper, we propose a closed-loop controller reproducing the peristalsis in healthy or disturbed stomachs. The proposed controller is implemented on a ring-shaped soft pneumatic actuator (RiSPA), representing the lower part of the stomach (antrum). The RiSPA is also used as the primary actuation system of an anatomically realistic gastric simulator. The experimental results show that the RiSPA can successfully imitate the peristalsis under the proposed closed-loop controller.

Keywords: Closed-loop control, Gastric simulator, Soft actuator, Peristalsis, Pneumatic actuators.

1 Introduction

Gastric simulators (GSs) have been developed to investigate the change of the food or drug in the human stomach, applicable for food and pharmaceutical industries. Peristaltic contractions are a series of wave-like contractions that break down the food particles and push them towards the small intestine. Soft robotic gastric simulators are the latest generation of GSs, mimicking the stomach peristaltic contractions in an anatomically realistic manner [1-2]. Physical characteristics of the peristalsis, such as frequency and pace, depend on the stomach's healthy or pathological conditions [3]. A soft robotic GS must be able to control its actuators to reproduce the peristalsis in both the healthy and pathological conditions. Yet, the compliant nature of actuators embedded in soft robotic GSs makes them suitable only for open-loop approaches [1]. A new soft robotic GS (SoRSS) is developed, on which various closed-loop control methods can be implemented. It comprises several ring-shaped soft pneumatic actuators (RiSPAs) as the primary actuation system. In this paper, we present the RiSPA and its model. Then, a closed-loop controller is designed and implemented on the RiSPA to imitate the peristalsis.

2 The RiSPA's Dynamic Modelling

The RiSPA concept is inspired by the lower part of the stomach known as the antrum. Also, it is the primary actuation mechanism for the SoRSS to generate contractive movements (see Fig. 1A). The RiSPA, comprises five silicon bellows (air-springs) enclosed in a PDMS frame [4]. The bellows' mechanism is when the pneumatic pressure applies, the air pressurizes the hollow chamber of the bellows, leading to expansion in longitudinal axes. A soft layer (made of Ecoflex 0030) connects all five bellows. A range sensor is installed at the bottom of each

bellows to measure its top part's displacement (see Fig. 1B). The RiSPA's frame has a 52 mm inner diameter, 20 mm width, and 15 mm thickness, similar to the dimensions of the antrum. The soft layer has a 50 mm diameter, 30 mm width, and 1 mm thickness. The bellows can deliver a maximum displacement of 20 mm. To model the RiSPA, each bellows assumed as a set of a mass-spring-damper (MSD) [4]. The hardness of the soft layer (00-30) is too small compared to bellows' (60 Shore A), and its 1 mm profile is negligible. Thus, the effects of coupling were ignored, and the model was considered decoupled. According to the Euler-Lagrange mechanics, the system dynamics can be obtained via the Lagrangian method. The dynamic model is written as the matrix form,

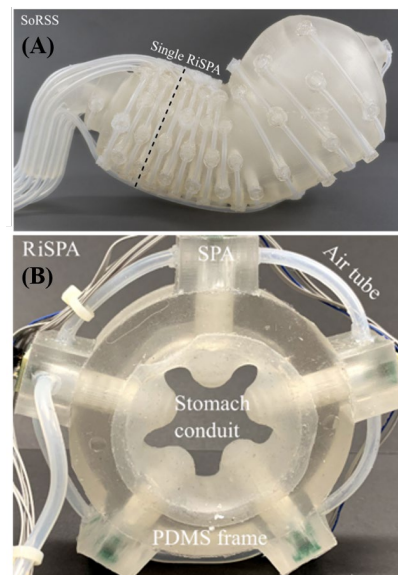


Figure 1. The SoRSS and the RiSPA setups. (A) SoRSS comprises of several RiSPAs (B) The sensor and the bellows (10 mm diameter) are embedded in the PDMS frame (bellows base).

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{A} &= \begin{bmatrix} \mathbf{0}_{5 \times 5} & \mathbf{I}_{5 \times 5} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{0}_{5 \times 5} \\ \mathbf{M}^{-1} \end{bmatrix}, \end{aligned} \quad (1)$$

where \mathbf{M} , \mathbf{K} , and \mathbf{C} are diagonal matrices of concentrated mass, spring stiffness, the damping coefficient, respectively. Also, $\mathbf{x}(t)$ and $\mathbf{u}(t)$ are the state variable and input vectors.

If $\mathbf{A} \in \mathcal{R}^{10 \times 10}$ and $\mathbf{B} \in \mathcal{R}^{10 \times 5}$ are constant time-invariant matrices, the dynamic model is a linear time-invariant (LTI) system, suitable for linear state feedback methods. On the contrary, if any component of MSDs is nonlinear or time-varying, a nonlinear state feedback method will be required.

3 Control Design and Implementation

To design a controller suitable for the dynamic model of the RiSPA, the pair of (\mathbf{A}, \mathbf{B}) must be stabilizable for all $t \in \mathcal{R}^+$. The control gain \mathbf{k} is defined in a way that $\mathbf{u}(t) = \mathbf{k}(\mathbf{x}(t) - \mathbf{x}_{reference}(t))$ minimizes the general quadratic performance index given by

$$J(\cdot) = \frac{1}{2} \left[\mathbf{x}^T(t_f) \mathbf{S} \mathbf{x}(t_f) + \int_0^{t_f} (\mathbf{x}^T(t) \mathbf{Q} \mathbf{x}(t) + \mathbf{u}^T(t) \mathbf{R} \mathbf{u}(t)) dt \right], \quad (2)$$

where \mathbf{Q} , \mathbf{S} and \mathbf{R} (state and input weighting matrices) are real symmetric positive semi-definite and positive definite matrices, respectively. To minimize $J(\cdot)$ for all $\mathbf{x}(0)$, \mathbf{k} must satisfy the optimal necessary condition

$$\mathbf{k} = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P}. \quad (3)$$

If the time of the control (t_f) is of importance, the state feedback can control the RiSPA in any time interval, suitable for pathological conditions where physical characteristics of the peristalsis is disturbed. In this case, \mathbf{P} is calculated from the differential equation

$$-\dot{\mathbf{P}}(\mathbf{x}(t)) = \mathbf{P}(\mathbf{x}(t))\mathbf{A} + \mathbf{A}^T \mathbf{P}(\mathbf{x}(t)) - \mathbf{P}(\mathbf{x}(t))\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T \mathbf{P}(\mathbf{x}(t)) + \mathbf{Q}. \quad (4)$$

Yet, if the time of the control is not important, Eq. (2) changes in a way that $t_f \rightarrow \infty$ and $\mathbf{S} = 0$. In this case,

$\dot{\mathbf{P}}(\mathbf{x}(t)) = 0$ and Eq. (4) turns into an algebraic equation.

The implementation process initiates in the LabView interface. The control signals are sent to the pneumatic valves through the MyRIO board. With RiSPA actuating, range sensors send a feedback signal to the LabVIEW through MyRIO and generates the displacement error. Subsequently, the error is fed to the controller, and the control loop is closed. In Figure 2, a peristalsis trajectory, based on the physiological ligature [3], was given to the RiSPA for the tracking problem. The peristalsis frequency is 0.05 Hz, and the amplitude and offset were chosen 4 and 5 mm to be consistent with the RiSPA's physical limitations.

Figure 2 illustrates that substituting the control gain from Eq. (3) in the control law, the RiSPA could track the peristalsis trajectory successfully. In Figure 2, the root-mean square errors (RSMEs) of tracking the peristalsis trajectory were calculated 1 mm and 1.2 mm for the differential (Eq. (4)) and its algebraic version, respectively. These amounts of errors are minor comparing to RiSPA's dimensions.

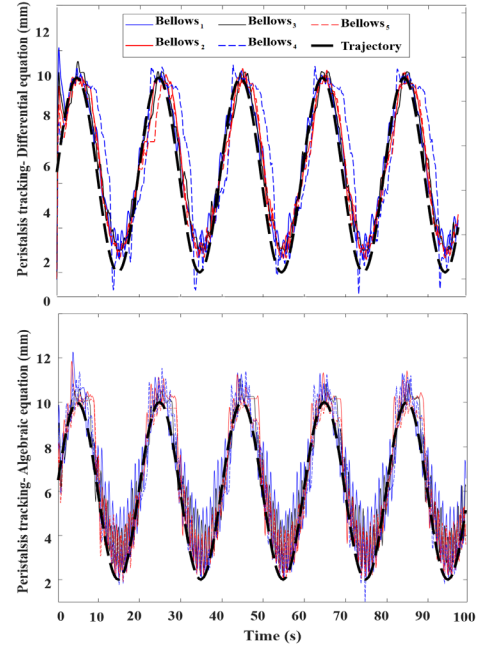


Figure 2. The experimental results of applying the controller on the RiSPA in both the differential and algebraic approaches.

4 Conclusions

In this paper, we designed and implemented an optimal closed-loop control on the RiSPA to reproduce the peristalsis. It was shown that the RiSPA could track a prescribed peristalsis trajectory successfully with a negligible error. The control method proposed could reproduce the peristalsis in both the healthy and disturbed stomach.

References

- [1] Y. Dang, et al., "SoGut: A Soft Robotic Gastric Simulator", *Soft Robotics*, Vol. 8, No. 3, pp. 273-283, June 2021.
- [2] L. Chen, et al., "Gastric emptying and morphology of a 'near real' in vitro human stomach model (RD-IV-HSM)", *Journal of Food Engineering*, Vol. 183, pp. 1-8, August 2016.
- [3] K. Schulze, "Imaging and modelling of digestion in the stomach and the duodenum", *Neurogastroenterol Motil*, Vol. 18, No. 3, pp. 172-183, March 2006.
- [4] R. Hashem, et al., "A Biologically Inspired, Ring-Shaped, Soft Pneumatic Actuator for Large Deformations", *Soft Robot*. In Press, 2021 DOI: 10.1089/soro.2021.0013.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Ph. D. |
| Name | Hanguen Kim |
| Affiliation | Seadronix Corp. |
| Address | 3rd, 20, Teheran-ro 20-gil, Gangnam-gu, Seoul, Republic of Korea |
| E-mail | hank05@seadronix.com |

Real-Time Vision-based Segmentation Algorithm for Autonomous Vessels

Hanguen Kim*, Donghoon Kim

Seadronix Corp.

3rd Floor, 20, Teheran-ro 20-gil, Gangnam-gu, Seoul, Republic of Korea

*Corresponding Author: hank05@seadronix.com

Abstract - It is very important to recognize obstacles in the marine environment. In the case of autonomous vessels, most researchers use a vision sensor as the fundamental sensor. In this paper, we propose a real-time vision-based segmentation algorithm based on deep learning. The proposed algorithm can be used to classify obstacles into a relatively high-resolution input image at the pixel level. In conclusion, it can be seen that very high performance was achieved while guaranteeing real-time.

Keywords: Image Segmentation, Computer Vision, Deep Learning, Autonomous Vessel.

1 Introduction

It is very important to recognize obstacles in the marine environment, and various sensors such as radar and automatic identification systems (AIS) are used due to environmental conditions. In the case of autonomous vessels, besides the existing radar and AIS, obstacles are recognized using sensors such as cameras and LiDAR. Various algorithms were proposed by many researchers using vision sensors. In this paper, we propose a real-time vision-based segmentation algorithm based on deep learning. Unlike other obstacle detection/recognition algorithms, the proposed algorithm can be used to classify obstacles into a relatively high-resolution pixel-level input image.

The remaining section is organized as follows. Section 2 describes the proposed algorithm. Section 3 compares the performance and computational speed of the proposed algorithm to that of other segmentation models. Finally, Section 4 sets out the conclusions and describes future research and development plans.

2 Proposed Method

Conventional segmentation algorithms based on FCN [1] have achieved impressive performance on different benchmarks. The DeepLabV3 [2] adopted a dilated spatial pyramid pooling module to capture multi-scale context. However, most approaches require high computational costs because of the high-resolution function and complexity of network connections. To address this issue, there are two main components to designing effective segmentation methods, such as lightweight base

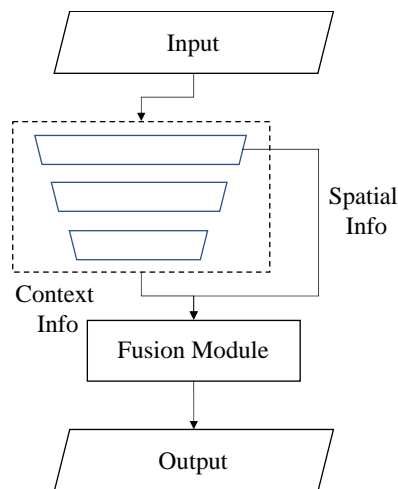


Figure 1. Structure of proposed model.

architecture and multi-branch architecture. By improving the BiSeNet V2 [3] architecture, our proposed deep learning model has two paths for low-level details and high-level contextual information, separately as shown in Figure 1.

Table 1. Results of mean intersection of union (mIoU) evaluation.

| Model | mIoU(%) | FPS | Params |
|----------------------|-------------|--------------|---------------|
| DeepLabV3 + ResNet50 | 51.1 | 12.51 | 39.64M |
| BiSeNetV2 | 47.4 | 102.60 | 3.63M |
| Ours | 56.3 | 66.98 | 16.08M |

3 Performance evaluation

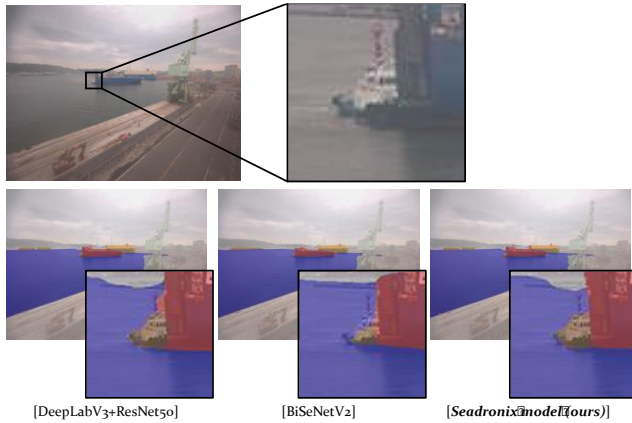


Figure 2. Comparison of inference results for each model based on evaluation dataset.

The computational speed and the mean intersection of union (IoU) analysis was calculated for the 1000 valuation images in 1280x960. These results are shown in Table 1. Our proposed model has approximately 10% more performance than DeepLabV3 and approximately 5 times faster computational speed. Compared to BiSeNetV2, the computational speed is 2 times slower, but the performance is about 18% higher. In conclusion, it can be seen that very high performance was achieved while guaranteeing real-time.

4 Future works

In the future, the proposed model will be improved for commercialization. In addition, an image enhancement technique will be incorporated into pre-processed images so that the model can be used both day and night in all weather conditions. The model will also be enhanced so that the distances to obstacles can be estimated to track and avoid them. Additionally, various technologies will be developed to enhance the power ratio and efficiency of the obstacle recognition system.

Acknowledgements

This research was a part of the project titled 'Development of Autonomous Ship Technology', funded by the Ministry of Oceans and Fisheries, Korea.

References

[1] Long, J., Shelhamer, E., and Darrell T., "Fully convolutional networks for semantic segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2015, pp. 3431-3440.

[2] Liang-Chieh C., George P., Florian S., and Hartwig A., "Rethinking atruous convolution for semantic image segmentation", *arXiv preprint arXiv:1706.05587*, 2017.

[3] Yu C., Gao C., Wang J., Yu G., Shen C., and Sang N., "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation", *International Journal of Computer Vision*, pp. 1-18, 2021.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Ph. D. |
| Name | Hanguen Kim |
| Affiliation | Seadronix Corp. |
| Address | 3rd, 20, Teheran-ro 20-gil, Gangnam-gu, Seoul, Republic of Korea |
| E-mail | hank05@seadronix.com |

Multi-sensor Fusion-based Vessel Monitoring System using Artificial Intelligence

Hanguen Kim*, Donghoon Kim

Seadronix Corp.

3rd Floor, 20, Teheran-ro 20-gil, Gangnam-gu, Seoul, Republic of Korea

*Corresponding Author: hank05@seadronix.com

Abstract - This paper proposes a multi sensor fusion-based vessel monitoring system. Berthing a vessel requires an accurate estimate of the relative distance between the quay wall and the vessel. However, operating large vessels near the port is a complex and challenging procedure. To deal with this problem, we develop a multi sensor fusion-based vessel monitoring system composed of a camera, LiDAR, differential global positioning system receiver, and inertial measurement unit, together with an algorithm for estimating the distance between the ship and berth. The proposed system provides highly accurate estimates and demonstrates robust performance under adverse weather conditions compared to a conventional berthing aid system.

Keywords: Sensor Fusion, Computer Vision, Deep Learning, Autonomous Vessel.

1 Introduction

Many accidents occur when vessels approach ports, as the pilot cannot accurately identify the vessel's surrounding environment or the port environment. Accidents in ports result in significant financial damage and injuries, leading to significant risks in port operations. To deal with the problem, an approach based on a pair of 1D laser distance sensor-based berthing aid system (BAS) is currently remains the most widely used approach in actual ports [1-2]. However, as the laser distance sensors are fixed at specific locations in the port and cannot be moved, there is a limit to the sizes of ships that can be recognized.

To solve this problem, we propose a multi-sensor vessel monitoring system; this system is more economic and overcomes the drawbacks of conventional BAS based on laser distance sensors. The proposed system provides precise and accurate measurements of the distances between the port and the ship, regardless of the environmental conditions and illumination at the time of berthing. The proposed system includes a camera, a LiDAR, a differential GPS receiver (DGPS) and a unit of inertia measurement (IMU).

The rest of this paper is organized as follows. Section 2 provides a description of the proposed systems. Section 3 presents the results of field tests, depending on the different environmental conditions. Finally, a conclusion is presented in Section 4.



Figure 1. Proposed multi-sensor system. Each sensor module consists of a wide-angle camera, LiDAR, DGPS, IMU, embedded computer, and local area network.

Table 1. System Specifications.

| Measuring item | Distance from ship to the berth |
|----------------|--------------------------------------|
| Field of view | 150m (longitudinal) x 300m (lateral) |
| Sampling rate | 1 Hz |
| IP code | IP 67 |

2 Proposed System

The sensor module of the proposed system is composed of a wide-angle camera for image capture, LiDAR for capture a 3D range data, DGPS for precise positioning, IMU for attitude measurement, embedded computer for software implementation as shown in Figure 1. Each sensor module is enclosed in a waterproof case to prevent corrosion of seawater. The sensor modules are installed in a fixed structure near the berth and at a certain

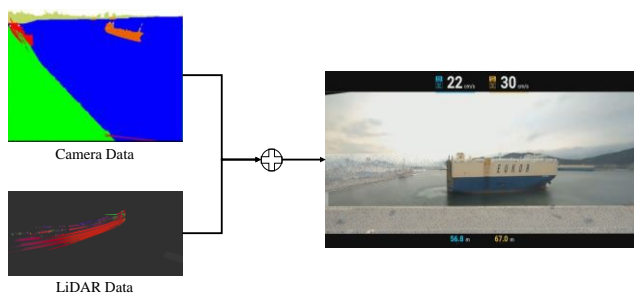


Figure 2. Overview of the fusion software.

height in relation to sea level. Sensor module positions are designed for manual or automatic adjustment. Table 1 presents the main specifications of the proposed sensor module, such as the measuring element, field of view, sampling rate and international protection (IP) code. As the vessel's maximum approach speed is less than 1m/s at the time of berthing, the sampling rate for estimating distance near the port generally requires 1 Hz. Additionally, the proposed system can provide live video data to pilots and ground crews via a cloud server at a sampling rate of 10 Hz. The IP67 rating indicates that the system can offer comprehensive protection against dust and water penetration.

3 Field test validation

The sensor-specific data of the proposed system are finally calculated through the fusion software to calculate the distance and speed between the vessel and the quay and the distance between the vessels. As shown in Figure 2, the fusion software provides the distance and speed of the vessel by fusing the information estimated from the image and LiDAR. Image data uses a deep learning model to estimate vessel shape information, and in the case of LiDAR data, an algorithm to estimate the vessel model has been applied. The verification of the algorithm was performed at the Ulsan port in Korea. In conclusion, the exact information can be seen to be produced in various environments, as shown in Figure 3.

4 Conclusion

This paper proposed a vessel monitoring system based on multi-sensor fusion system. Fusion software estimates the distance between the vessel and the berth. To evaluate the performance of the proposed system, field tests were conducted in various environments using actual vessels at Ulsan Port, Korea. From the field test results, it can be found that the performance of the proposed system was greater than that of the conventional BAS capable of accurate distance measurements in an ideal environment. In the future, the reliability of the system will be improved by extending the installation site.



Figure 3. Field test validation results

Acknowledgements

This research was a part of the project titled 'Development of Autonomous Ship Technology', funded by the Ministry of Oceans and Fisheries, Korea.

References

- [1] Yu Y., Zhao B., Zhu H., and Yang L., "Berthing support system using laser and marine hydrometeorological sensors", in *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2017, pp. 69-72.
- [2] Perkovič M., Gucma L., Bilewski M., Muczynski B., Dimc F., Luin B., Peter V., Vivien L., and Milan B., "Laser-based aid systems for berthing and docking", *Journal of Marine Science and Engineering*, vol. 8, no. 5, pp. 1-20, 2020.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | X |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|--|
| Title | Development of Military Tactical Training Simulator Using LVC-based Drone |
| Name | Il-yeop Ahn |
| Affiliation | Korea Electronics Technology Institute |
| Address | Seongnam, South Korea |
| E-mail | iyahn@keti.re.kr |

Development of Military Tactical Training Simulator Using LVC-based Drone

Sungwook Jung, Jonghong Park, Sung-chan Choi, Wonseok Jung, and Il-yeop Ahn*

Autonomous IoT Research Center
Korea Electronics Technology Institute (KETI)
Seongnam, South Korea

*Corresponding Author: iyahn@keti.re.kr

Abstract - In this paper, we propose a tactical training simulator platform using LVC-based drones. Using a small, portable platform, it is possible to access and control real and virtual drones in a single training system, and practical drone tactics training is possible with integrated training support of real maneuver training and virtual simulation training. Real-time communication and data transmission/reception are possible through interworking with the IoT platform using the digital twin relay engine, and finally, the possibility of future use is suggested.

Keywords: Live-Virtual-Constructive, Digital twin, Unmanned aerial vehicle

1 Introduction

LVC simulation technology is a system that simulates battle by providing an integrated simulation interlocking environment by interworking the models of Live, Virtual, and Constructive, and composing it into one large synthetic battlefield. Live, Virtual, and Constructive models have been selected and utilized according to the purpose of combat training and analysis in the military field. However, with only a single model, there were limitations in the training environment, high cost, a sense of disparity from practice, and the accuracy of the analysis. As a means to resolve the limitations of real-world flight training and to meet new training demands according to changes in operational concepts, the LVC-based drone mission training system using advanced M&S (Modeling & Simulation) and communication technology is drawing attention. Accordingly, domestic and foreign defense-related research institutes are conducting research and development to provide training and effectiveness analysis environments through LVC integration.

In this paper, we propose a digital twin relay engine using oneM2M-based IoT platform for real-time LVC-based drone training simulation. This relay engine connects real drones and virtual drones in real-time and transmits data in real-time to the GCS for integrated control, enabling not only the trainee but also the commander level to control the entire training.



Figure 1. LVC-based drone training platform

2 Proposed Method

What this LVC-based drone system needs are a digital twin relay engine system and a real-time communication network that can manage each individual. To this end, this research team defines UAVs as objects and successfully provides user-friendly services through an open IoT ecosystem using Mobius, an Internet of Things (IoT) platform based on the oneM2M global standard. The Mobius platform is a good framework that can centrally manage multiple unmanned aerial vehicles based on the cloud and has the advantage of easy data management and easy access by users [1-2].

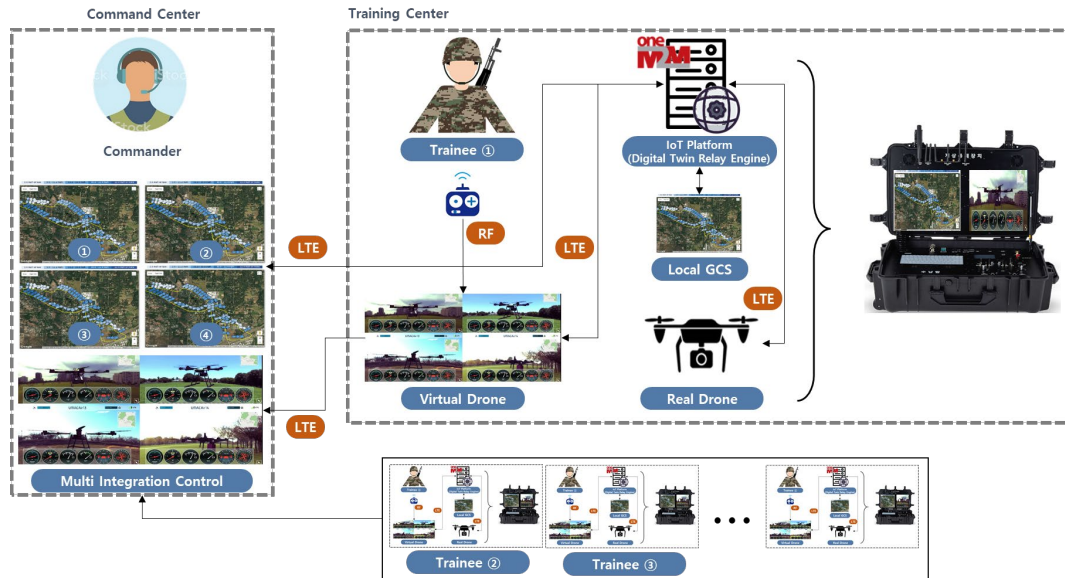


Figure 2. Overall system architecture

Figure 1 shows the final overall prototype proposed in this paper. Trainee carries box-shaped training equipment and a drone and can perform flight training and various mission training at the training center. A high-performance GPU is installed in the training equipment to operate a virtual simulator, and since digital twin relay engine software is running internally, it is possible to transmit and receive data and control the drone system using the separate protocols of real drones and virtual drones. The relay engine is connected to the IoT platform to manage the aircraft data and training missions for each trainee at the platform stage. The platform uses an LTE network to enable multi-training control of the commander.

The biggest issue in this platform is the digital twin relay engine data transmission delay time. In general, the human-recognizable delay time is about 200 ms, so the engine operates with an operation delay time of 200 ms or less, and has a processing speed of 1500 tps or more of data processing performance based on a single server.

3 Conclusions

In this paper, we propose an LVC-based drone training simulator using a digital twin relay engine. This platform expects the following effects:

1) Use as a tactical training simulator using drones in the defense field: a) In a virtual tactical training environment, it is possible to use the command system and operational operation training of tactical training using drones for each required army and squadron. b) It can be used to conduct integrated operation training for reconnaissance and attack missions using drones as assets by linking with the wargame model in operation in the military and other M&S systems. c) It can be used as a

simulator to verify drone technology with the operational performance required by the military.

2) Utilization of digital twin drone ground operation system in various mission service fields: a) Through the introduction of digital twin technology, it is possible to establish a ground operation system that can be used stably by demanding organizations when operating various mission services. b) Based on the proven platform technology based on the data interworking protocol of multiple heterogeneous drones and international standards, it can be used in mission services that require data management of multiple heterogeneous drones or require a simultaneous operation.

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Science and ICT/Ministry of Trade, Industry and Energy/Ministry of Land, Infrastructure and Transport (Grant 21DPIW-C164990-01).

References

- [1] S. Choi, I. Ahn, J. Park, and J. Kim, "Towards real-time data delivery in oneM2M platform UAV management system," in *Proc. of Int'l Conf. on Electronics, Information, and Communication (ICEIC)*, 2019, pp.1-3.
- [2] J. Kim, J. Lee, J. Kim, and J. Yun, "M2M service platform: Survey, issues, and enabling technologies," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 61-76, 2013.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|------------------|---|
| Invited Talk | X |
| Oral Preferred | |
| Poster Preferred | |
| No Preference | |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|-------------|--|
| Title | Professor |
| Name | Haruichi Kanaya |
| Affiliation | Department of Electronics, Kyushu University |
| Address | Fukuoka-shi, Fukuoka 819-0395, JAPAN |
| E-mail | kanaya@ed.kyushu-u.ac.jp |

| | |
|-------------|--|
| Title | CEO |
| Name | Osamu Takiguchi |
| Affiliation | ALSENS Inc. |
| Address | Sagamihara-shi, Kanagawa 252-0131, JAPAN |
| E-mail | takiom@mua.biglobe.ne.jp |

| | |
|-------------|---|
| Title | Technical Staff |
| Name | Hatsune Ogra |
| Affiliation | Oita Prefecture Agriculture, Forestry and Fisheries Department, Livestock Industry Technology Office |
| Address | Oita-shi, Oita 870-8501, JAPAN |
| E-mail | ogura-hatsune@pref.oita.lg.jp |

| | |
|--------------------|--|
| Title | Chief |
| Name | Hisayoshi Eto |
| Affiliation | Livestock Research Division, Oita Prefectural Agriculture, Forestry and Fisheries Research Center |
| Address | Taketa-shi, Oita 878-0201, JAPAN |
| E-mail | eto-hisayoshi@pref.oita.lg.jp |

| | |
|--------------------|--|
| Title | Associate Professor |
| Name | Hideyuki Takahashi |
| Affiliation | University Farm, Faculty of Agriculture, Kyushu University, |
| Address | Taketa-shi, Oita, 878-0201, JAPAN |
| E-mail | takahashi@farm.kyushu-u.ac.jp |

Ear Tag Type Sustainable Sensing Platform for livestock

Haruichi Kanaya^{1*}, Osamu Takiguchi², Hatsune Ogra³, Hisayoshi Eto⁴, Hideyuki Takahashi⁵

¹Department of Electronics, Kyushu University, Fukuoka-shi, Fukuoka 819-0395, JAPAN

²ALSENS Inc., Sagamihara-shi, Kanagawa 252-0131, JAPAN

³ Oita Prefecture Agriculture, Forestry and Fisheries Department, Livestock Industry Technology Office, Oita-shi, Oita 870-8501, JAPAN

⁴Livestock Research Division, Oita Prefectural Agriculture, Forestry and Fisheries Research Center, Taketa-shi, Oita 878-0201, JAPAN

⁵University Farm, Faculty of Agriculture, Kyushu University, Taketa-shi, Oita, 878-0201, JAPAN

*kanaya@ed.kyushu-u.ac.jp

Abstract - Obtaining the vital information from livestock in real time is quite important from the viewpoint of reducing the effort of farmers. Moreover, small size, non-invasive devices or surface mount devices should be need to reduce the burden on livestock during measurement. In this research, we fabricated a sustainable sensing platform that measures the temperature of the ear surface in real time by integrating a small solar panel, a temperature and humidity sensors, and a BLE (Bluetooth low energy) module. The size of this system was reduced to attach on the ear tag. Vital data from livestock and environmental data are transmitted to the tablet which work by solar power, and we can observe all data in real time.

Keywords: Ear Tag, Sustainable Sensing Platform, vital information, livestock.

1 Introduction

Recently, obtaining the vital information from livestock in real time is quite important from the viewpoint of not only finding livestock estrus, calving and illness but also reducing the effort of farmers. In particular, if changes in body temperature during the conception period of cattle can be grasped quickly and accurately, it can be expected that the calving interval will be shortened by improving the conception rate, leading to an improvement in productivity. On the other hand, small size, non-invasive devices or surface mount devices should be need to reduce the burden on livestock during measurement.

Till now, a wireless intravaginal probe [1] or ventral tail base surface temperature sensor [2] as a wearable wireless sensor in livestock especially in cattle are developed. However, these devices require batteries and require effort to replace the batteries.

In this research, we constructed a sustainable (or battery-less) sensing platform that measures the temperature of the ear surface in real time by integrating a temperature and humidity CMOS sensors on the back side, a small-size solar panel and a BLE (Bluetooth low energy) module on the top. The size of this system was reduced to attach on the ear tag. Vital data from livestock and environmental data are transmitted to the tablet which work by solar power, and we can observe all data in real time, and the data was also stored in the tablet.

2 Sustainable Sensing Platform

Fig. 1 shows the proposed ear-tag type Sustainable vital sensing platform. The system has a three-layer structure, with a solar panel on the front (15 mm x 15 mm), a BLE communication module (Cypress) underneath, and a temperature / humidity CMOS sensor on the back. All blocks such as solar panel (energy harvesting layer), sensors (sensing layer) and data transmission module (wireless telecommunication layer) are designed independently, therefore, we are free to replace the harvesters or sensors and so on. As an energy source, it is also possible to use electromagnetic waves for wireless communication system [3].

The sensing plane is placed on the surface of the ear. It has been downsized (30 mm x 50 mm x 10 mm, 14 g) in consideration of mounting on small livestock other than cattle (pigs, goats, etc.).

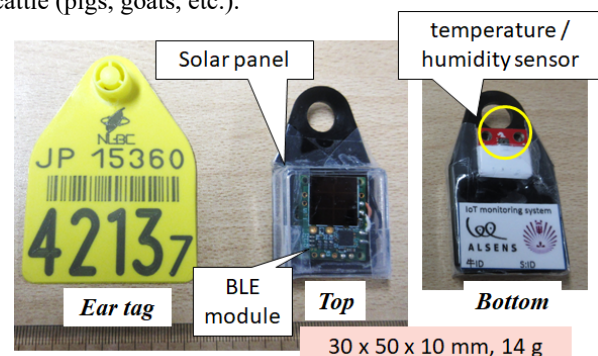


Figure 1. Photograph of the sensing platform.

Fig. 2 shows the photograph of the sensor set up. Since the sensor is smaller than the cattle's ear, it seems that there is no stress in installation.

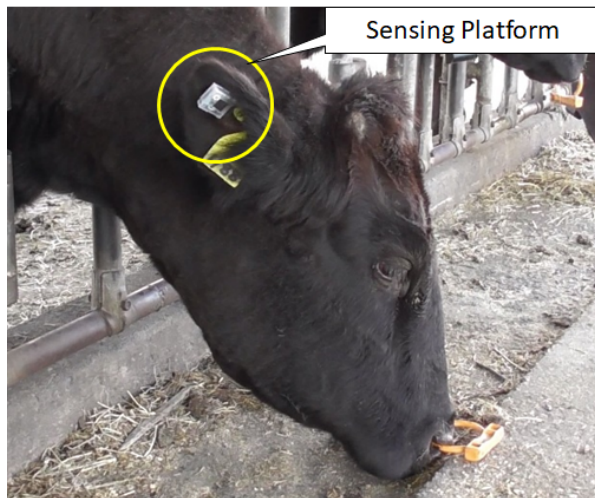


Figure 2. Ear-tag type sustainable sensing platform.

3 Measurement in the Field

The demonstration measurements were conducted at the Livestock Research Division, Oita Prefectural Agriculture, Forestry and Fisheries Research Center and University Farm, Faculty of Agriculture, Kyushu University. In the ranch, there are some grazing areas. Since livestock move between grazing areas, it is desirable to be able to easily carry the data receiver without an AC outlet. This time we chose a tablet working by the solar power as a data receiver. Fig. 3 shows the photograph of the receiver in the field. Due to its small size and no need for batteries or AC outlets, the proposed sensing system can be used to easily measure the temperature and humidity of grazing areas, paddock and barn. Currently, we can obtain the time dependence of the temperature and humidity even if the cloudy or rainy day and inside the cattle barn.



Figure 3. Photograph of the receiver in the field.

Fig. 4 shows the 24-hour results on a sunny day in last November. Since all sensing platforms have

independent IDs, it is possible to distinguish between the body surface temperature of livestock and the environmental temperature. The surface temperature of all (4) cattle's is almost constant, which is different from the behavior of the outside air temperature. It can be said that the four cattle are healthy. Moreover, by checking the data of the sensor on the east side and the sensor on the west side, you can know the time of sunrise and sunset.

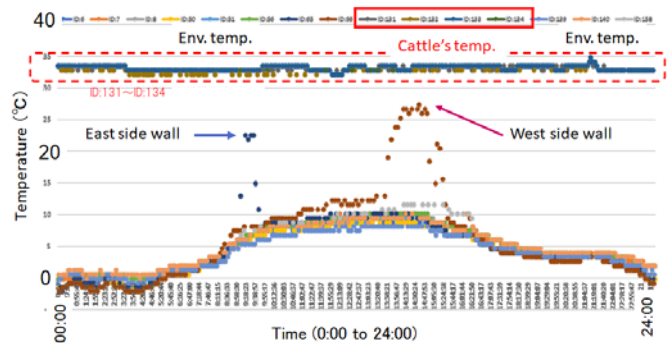


Figure 4. Time dependence of the temperatures.

4 Conclusions

Sustainable (battery-less) sensing platform which measures the body surface temperature in real time by integrating a small solar panel, a temperature and humidity sensors, and a BLE module is fabricated and tested. It operates 24 hours a day and 365 days a year without any power supply, therefore it will find estrus, parturition and illness in livestock in the near future.

Acknowledgements

This work was partially supported by the Cabinet Office (CAO), Cross-ministerial Strategic Innovation Promotion Program (SIP), an intelligent knowledge processing infrastructure, integrating physical and virtual domains (funding agency: NEDO) (JPNP18014), SCOPE, and KAKENHI (JP21K04177) JSPS.

References

- [1] L. M. Andersson, H. Okada, R. Miura, Yi Zhang, K. Yoshioka, H. Aso, T. Itoh, "Wearable wireless estrus detection sensor for cows", *Computers and Electronics in Agriculture*, Vol. 127, pp. 101-108, 2016.
- [2] R. Miura, K. Yoshioka, T. Miyamoto, H. Nogamic, H. Okada, T. Itoh, "Estrous detection by monitoring ventral tail base surface temperature using a wearable wireless sensor in cattle", *Animal Reproduction Science*, Vol. 180, pp. 50-57, 2017.
- [3] H. Kanaya, S. Tsukamoto, T. Hirabaru, D. Kanemoto, R. K. Pokharel, K. Yoshida, "Energy Harvesting Circuit on a One-Sided Directional Flexible Antenna", *IEEE Microwave and Wireless Components Letters*, Vol. 23, Issue 3, pp. 164-166, 2013.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | |
| No Preference | X |

--

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | UHF RFID Sensor Tag Comparison and Temperature Error Analysis for Covid-19 Cold Chain Logistics |
| Name | Minkyong Kim, Youchung Chung |
| Affiliation | Information and Communication Engineering |
| Address | Information and Communication Eng. Dept. Daegu University, Korea, 38453 201 Dagudae-Ro, Jinrang, Kyungsan City, Kyungbook, Korea |
| E-mail | so8sun@naver.com |

UHF RFID Sensor Tag Comparison and Temperature Error Analysis for Covid-19 Cold Chain Logistics

Minkyong Kim*, Youchung Chung,

Information and Communication Eng. Dept. Daegu University, Korea, 38453
201 Dagudae-Ro, Jinrang, Kyungsan City, Kyungbook, Korea,
youchung@daegu.ac.kr

Abstract – In order to design UHF RFID tag for Covid-19 cold chain logistics, this article investigates and compares UHF long-range passive RFID sensor tag for various perspectives. The temperature deviation of the sensor tag was measured and compared in different conditions, The sensor tags are left in the refrigerator, freezer and the room temperatures of 24 and 32°C. The errors are analyzed and investigated, and the storage temperature of Covid-19 vaccines are investigated to manage and monitor the temperature of the vaccines logistics.

Keywords : UHF temperature sensor, RFID sensor tag, Covid-19 Cold Chain

1 Introduction

Nowadays, with the advantages of low cost, the convenience of fabrication and passive, the UHF radio frequency identification (RFID) technology has promoted the development and prosperity of the Internet of Things [1]. With the rapid development of RFID technology, RFID technology has evolved from simply identifying target object information to sensing capabilities [1]. In this paper, the RFID technology is used to identify and detect Covid-19 vaccines.

Recently, due to Covid-19, Cold Chain has been urgently needed to preserve vaccines and transport them. Some vaccines must be stored at -70 °C, while others can be stored at -20 °C and usually refrigerated. In addition, some items must be kept Cold Chains that cannot be collected unless the storage temperature is detected and the storage temperature is stored at regular intervals to ensure time and temperature data are not exceeding the appropriate temperature [2].

Communication equipment with microprocessors, such as temperature sensors, is usually used in such logistics, or RFID tags are used simply. RFID has labels for HF and UHF bands, active tags that can store and log temperature information as battery support, and manual tags that measure temperature only when requested by readers [3-6].

Comparison of temperature sensor labels that can be purchased commercially in the UHF band, comparison of

battery availability with corresponding protocols, measurements, temperature ranges and frequencies, etc.

Five to ten RFM3200-AFR labels were used as UHF temperature sensor labels and the measurements were compared in room temperature and high room temperature, refrigerator and freezer. Temperature and temperature deviations of each label were measured and compared [5-6].

2 Comparison and Results

As a number of temperature sensor labels for multiple HF bands, there are active manual labels. In this paper, due to space limitations, only UHF bands are compared. EM4325V21, which can be applied to EM Microelectronics' manual active label, Magnas-S3M3E, and Magnas-S3M3D, which can measure temperature and humidity, are briefly compared in Table 1. The corresponding protocol and measurement range temperature are compared according to the presence or absence of batteries. RF Micron tags can measure temperature and humidity and recognize temperatures from -40 ~ 85 degrees Celsius. EM4325 can be passive/active and has a short recognition distance, although it has the function of saving data by time.

Using RF Micron's Magnas-S3M3E chip and RFM3200-AFR, five to ten labels tags were used at four different temperatures cases, they are measured at room temperature and high temperature, in refrigerators and in freezers, as shown in Table 2. Especially at low and high

temperatures, it is recommended to apply multiple tags and the average of the multiple tags instead of using single tag because of the error of single tag can be larger than that we expected.

Table 1. Comparison of UHF band tag chips

| Manufacturer | Product Number | Measurement Battery | Measurement range temperature | Frequency | ISO Standard |
|---------------------|-------------------|---------------------------|-------------------------------|-------------------|--------------------------------|
| RF Micron | Maguna s-S3 M3E | Temperature RSSI | -40°C ~ 85°C | 860M Hz ~ 960M Hz | 18000-6C EPC Class 1 Gen2 |
| RF Micron | Maguna s-S3 M3D | Temperature/Humidity RSSI | -40°C ~ 85°C | 860M Hz ~ 960M Hz | 1800-6C EPC Class 1 Gen2 |
| EM Micro electronic | EM 4325V2 1T P8B+ | Temperature 1.25~3.6 V | -40°C ~ 60°C | 860M Hz ~ 960M Hz | EPC Class 3 Gen2 18000 - 63.64 |

Table 2. Temperature measurements (Freezer, Refrigerator, Ambient 1, Ambient 2)

| | Freezer | Refrigerator | 24°C | 32°C |
|-----------|---------|--------------|-------|-------|
| Tag 1 | -2.4 | 14 | 25.1 | 31.8 |
| Tag 2 | -1.7 | 15.2 | 24.9 | 32.7 |
| Tag 3 | -4.2 | 15.2 | 24.9 | 30.8 |
| Tag 4 | -0.1 | 13.7 | 25 | 32.3 |
| Tag 5 | -4 | 13.2 | 23.7 | 31.9 |
| Tag 6 | -2 | 14.7 | 25.2 | 34.4 |
| Tag 7 | -3.6 | 13.5 | 24.8 | 32.6 |
| Tag 8 | -1.9 | 14.7 | 23.7 | |
| Tag 9 | -1.2 | 13.8 | 22.6 | |
| Tag 10 | -1.6 | 14.7 | 26.3 | |
| Average | -2.27 | 14.27 | 24.62 | 32.26 |
| Deviation | 1.23 | 0.68 | 0.97 | 1.02 |
| Max-Min | 4.1 | 2.0 | 2.6 | 3.4 |

Table 3. Temperature of the vaccine storage

| | AstraZeneca | Moderna/Janssen | Pfizer |
|---------------------|-------------|-----------------|-------------|
| storage temperature | 2~8°C | -25 ~ -15°C | -90 ~ -60°C |

As a result of Table 2, using the 10 labels, there are errors of 2 ~ 4 °C at ambient and errors of up to 4 °C in freezer. It can be confirmed that the error is greater at low temperatures. Therefore, it shows that at least two or more labels must be attached because only 1 label does not know the accuracy of the temperature. It is recommended to use the average of two or more attached tags instead of using the single tag because of the larger error.

Investigation and comparison of labels applicable to Covid-19 vaccines show that temperature maintenance can be checked and monitored in vaccine preservation and transportation. All vaccines except Pfizer can use the tags through Table 1 and Table 3.

3 Conclusions

In this study, the UHF band tag chip was compared, and the battery availability and supporting protocol were compared, and the physical measurement value and range of the chip were compared. Using the RFM3200-AFR label, the UHF temperature sensor label was used to measure and compare ambient, refrigerator and freezer temperature, measure and observe errors and deviations. Therefore, it is helpful to solve the problem of labeling when applying Covid-19 vaccine.

References

- [1]. Shenghui Zhi, Yuan Yao, Junsheng Yu, and Xiaodong Chen, "Design of a Passive RFID Yagi-Uda Sensor Tag Antenna" Proceedings of 2019 International Conference on Microwave and Millimeter Wave Technology, 2019,
- [2]. 코로나19 백신 보관 수송관리 지침, 질병관리청, 2021.
- [3]. NT3H2111W0FHKH Datasheet, NXP Co. (ISO 14443)
- [4]. SL2S20002F Datasheet, NXP Co. (ISO 15963, 18000-3)
- [5]. SL3S1240F Datasheet, NXp Co.
- [6]. EM4325V21 Datasheet, EM Microelectronics Co.
- [7]. Magnas-S3 M3E Datasheet, RF Micron Co.
- [8]. Magnas-S3 M3D Datasheet, RF Micron Co.
- [9]. RFM3200-AFR Tag Datasheet, RF Micron Co.

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | |
| Oral Preferred | |
| Poster Preferred | X |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|-----------------------------------|
| Title | Professor |
| Name | Jung Young-Bae |
| Affiliation | Hanbat National University |
| Address | Daejeon 34158, South Korea |
| E-mail | ybjung@hanbat.ac.kr |

Broadband Planar Array Antenna with High-Efficiency for Synthetic Aperture Radar Applications

Patrick Danuor and Young-Bae Jung*

Electronic Engineering Department, Hanbat National University
Daejeon 34158, South Korea

*Corresponding Author: ybjung@hanbat.ac.kr

Abstract - In this paper, a high efficiency broadband planar array antenna designed at X-band for synthetic aperture radar (SAR) applications is presented. The structure is composed of multiple stacked layers, an active patch, a parasitic patch and a ground plane. In order to achieve high directivity, wideband and low antenna profile regarding antenna weight, a Rohacell foam material is placed between the two dielectric substrates. The experimental results confirm that the proposed antenna achieves high gain, high efficiency with good wideband characteristics.

Keywords: Broadband, high efficiency, high gain, parasitic patch, synthetic aperture radar.

1 Introduction

An important component for earth sensing and remote monitoring systems is the synthetic aperture radar (SAR). SAR creates two-dimensional images or three-dimensional reconstruction of objects [1, 2]. SAR satellites also possess many applications pertaining to missile, pilotless aircrafts and other significant platforms.

SAR antennas are required to have high efficiency, high gain, broadband levels as well as low weight; therefore, the compact size, low cost and ease of microstrip planar antennas make them ideal for SAR antenna design [3, 4]. However, microstrip planar antennas inherently possess relatively low gain, low bandwidth and low efficiency [5, 6].

A lot of studies have been done to overcome the challenges of the microstrip planar antennas, however these methods or techniques result in complex structures which may be unsuitable for applications in SAR satellite systems.

In this paper, we propose a 12×16 planar antenna array designed using two dielectric substrates, an active patch, a parasitic patch, and a ground plane. A relatively thick Rohacell foam material is used to separate the two substrates to achieve a high gain, broadband and lightweight characteristics suitable for SAR satellite applications.

2 Proposed Method

In order to design the 12×16 planar array antenna, an element antenna is first designed as displayed in Figure

1. From the figure, the element antenna is multilayered and designed using a relatively thick Rohacell substrate (i.e. h_b), which has very low tangent loss and a dielectric constant close to that of air. The Rohacell is sandwiched between two Taconic TLY-5 substrates (i.e. h_{sub1} and h_{sub2}). Moreover, the antenna contains three copper layers, i.e., the parasitic patch, active patch and ground plane.

Multiple element antenna structures are arranged in an array configuration to form the 12×16 planar array antenna structure.

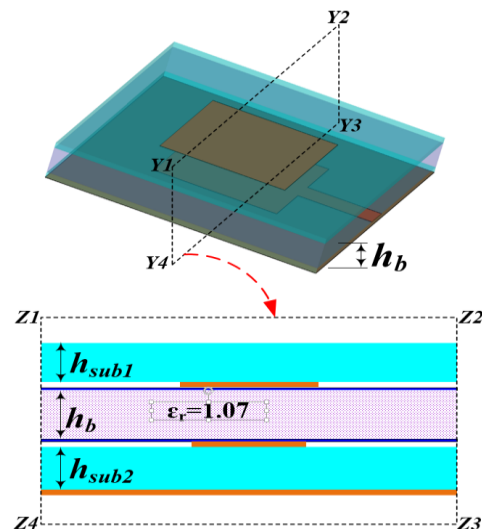


Figure 1. Proposed element antenna design

3 Conclusions

In this paper, a planar antenna array structure based on multilayered structure is presented. The antenna array, which is designed at the X-band frequency achieves high efficiency, broadband and low weight. The added Rohacell material aids the broadband operation and lightweight nature significantly. The antenna is suitable for SAR satellite applications.

Acknowledgements

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-01658, Key Technologies Development for Next Generation Satellites).

References

- [1] Kirscht, Martin, and Carsten Rinke. "3D Reconstruction of Buildings and Vegetation from Synthetic Aperture Radar (SAR) Images." MVA 1998
- [2] Elhefnawy, M.; Sri Sumantyo, J.T. A review on designing antenna arrays for long-range synthetic aperture radar. In Proceedings of the 2016 International Workshop on Recent Advances in Robotics and Sensor Technology for Humanitarian Demining and Counter-IEDs (RST); IEEE, 2016.
- [3] Wang, W.; Zhang, H.-T.; Zhang, Z.-H.; Liang, X. Broadband antenna array for SAR applications. In Proceedings of the 2014 IEEE Antennas and Propagation Society International Symposium (APSURSI); IEEE, 2014.
- [4] Akbar, P.R.; Sumantyo, J.T.S.; Saito, H. *Design of synthetic aperture radar onboard small satellite*; Japan: IEICE, 2012;
- [5] N. K. T. R, N. Kumar, M. Sreenivasan, K. G. Thomas and P. H. Rao. A multilayer X-band FMCW-SAR microstrip array. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 2016
- [6] Kirscht, Martin, and Carsten Rinke. "3D Reconstruction of Buildings and Vegetation from Synthetic Aperture Radar (SAR) Images." MVA. 1998

Please mark an "X" at the preferred presentation type. If you choose "No Preference", the decision will be made by the Program Committee.

Presentation Type

| | |
|-------------------------|----------|
| Invited Talk | O |
| Oral Preferred | |
| Poster Preferred | |
| No Preference | |

Please complete the corresponding author contact details. Others are optional. Please extend the table if you have more than four authors.

Corresponding Author

| | |
|--------------------|---|
| Title | Professor |
| Name | Sungtek Kahng |
| Affiliation | Department of Information and Telecommunication Engineering, Incheon National University |
| Address | 119, Academy-ro, Yeonsu-gu, Incheon, Republic of Korea |
| E-mail | s-kahng@inu.ac.kr |

A Study on Correlation between the Millimeter-wave Array Antenna and Digital Wireless Communication in the 5G Mobile System

Yejune Seo, Munsu Jeon, Junghyun Cho, Jiyeon Jang, Yejin Lee, and Sungtek Kahng*

Department of Information & Telecommunication Engineering, Incheon National University

Incehon 22012, Korea

*Corresponding Author: s-kahng@inu.ac.kr

Abstract - In this paper, an intuitive approach to assessing advantages of beamforming in 5G wireless communication is proposed as a novel try and practical demonstration of importance of alignment between the transmitter's and receiver's beams working in millimeter-wave frequency bands. The effects of the misalignment and alignment between beams need to be checked for, which was conducted with a horn antenna and the 4×4 Butler matrix RF-to-RF wireless connectivity between the horn and the microstrip line beamformer, concerning the changing angle of the beam from the Butler matrix, was tested, showing over 12 dB enhancement in received power. This direct electromagnetic link test was accompanied by examining 64-QAM constellations for beam-angle change.

Keywords: millimeter-wave antenna; 5G antenna; beamforming antenna; Butler matrix; 64-QAM.

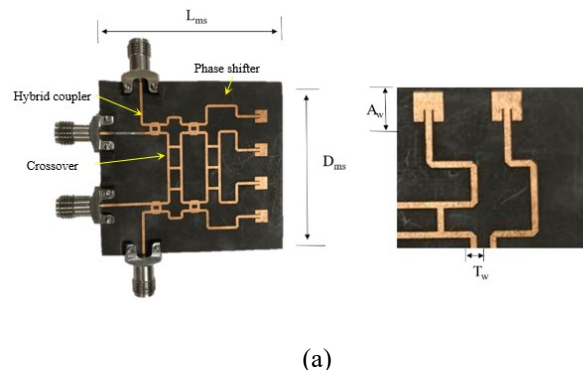
1 Introduction

The 5th generation (5G) mobile communication is featured by technological fascination such as several Gbps data transfer-rate, low latency and low interference [1–3]. In order to realize the wireless connectivity for 5G, antennas operable in millimeter-wave frequency bands are needed. Making use of millimeter-wave antennas, the wide-bandwidth and narrow beamwidth would be accomplished by designing them to be arrays whose footprint is relatively small for even the commercial wireless phone. The beamwidth becomes narrow and pointy to have higher directionality in the far-field pattern. This is so-called beamforming. In this paper, a new way that 5G wireless system developers can obtain intuitions on the quality and effects of beamforming functions is suggested. This tried-and-true verification approach comprises the design of beamforming antennas and two experimental setups. Firstly, to give the capability of beamforming and beam-tilting to the wireless connectivity tests, the 4×4 Butler matrix was designed and manufactured. Secondly, a measurement setup was devised to check RF-to-RF sensing between the horn and the microstrip Butler matrix. By changing the angle of the beam from the horn to the RX beamforming antenna, the transmission coefficient of the beam from the horn to the RX beamforming antenna was recorded for beam misalignment and alignment. Thirdly, a setup was formed to watch 64-QAM diagrams. According to the change in the angle of the beam, 64-QAM constellations were plotted. Notwithstanding, the

beam alignment led to clear pictures of I/Q symbol spots. The tests revealed that beam alignment increases the received power by over 12 dB from the beam misalignment in the RF-to-RF connection.

2 Design of the Butler Matrix Antenna as Beamforming AUT

Figure 1a is the photograph of the prototype of the microstrip-line Butler matrix antenna looking similar to the one in [4]. This beamforming antenna has four input ports and four radiating elements at the output ports made on RT5880 as the substrate with thickness of 0.25 mm. The geometrical parameters noted in Figure 1a are mentioned in the followings as Table 2.1.



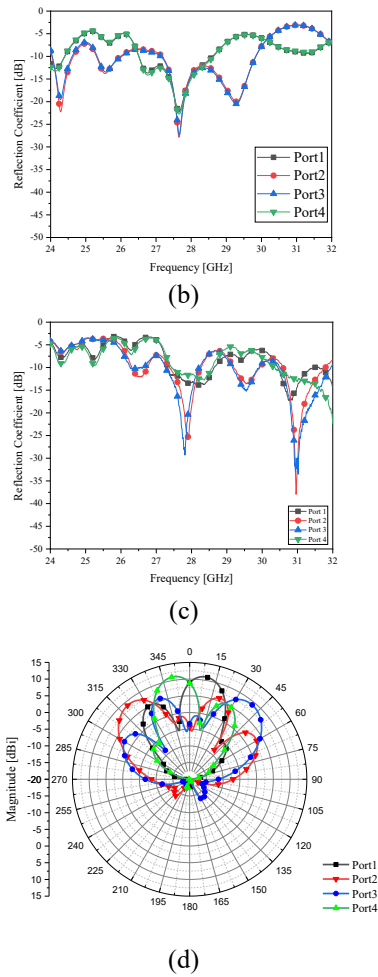


Figure 1 The microstrip-line Butler matrix (a) prototype; (b) port reflection coefficients (c) port reflection coefficients (d) beam-patterns

Table 2.1 The physical dimensions of the microstrip-line Butler matrix.

| Parameter | Value |
|-------------|----------|
| L_{ms} | 49 mm |
| D_{ms} | 51.92 mm |
| A_w | 3.48 mm |
| T_w | 0.7 mm |
| Antenna gap | 8.1 mm |

They present the impedance matching at 28 GHz as the 5G mobile frequency. The beamforming and beam-steering functions are observed in Figure 1 from EM simulation and measurement. Once fabricated, the surface of the thin substrate tends to be bent and a little deformed due to the weight of the connectors, which causes differences, i.e., unwanted back radiation and a shift in the angles of the beams from the EM simulated data. The beams range from -30° to 30° , which is adopted to the change in the beam

direction for RF-to-RF link tests and I/Q digital wireless evaluation.

RF-to-RF Test and 64-QAM Investigation

As for a TX and an RX in 5G/6G mobile communication, there are four possible scenarios of beam pointing as in the following figure. Various situations of antenna positioning and beam pointing between the horn(TX) and the beamformer(RX) are represented by Figure 2. The strongest RF link is expected in Figure 2a as the in-line beam alignment (α). As the location of the RX changes, the beam tilted by the RX catches the beam by the rotated TX horn, which means beam alignment, and results in much improved connectivity as in Figure 2d, denoted as δ_{2R} , δ_{1R} , δ_{1LR} and δ_{2L} .

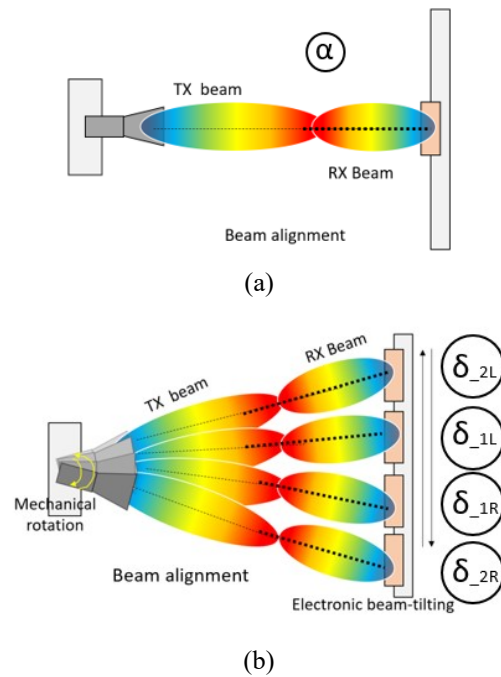
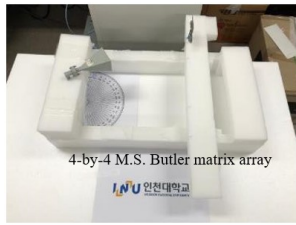
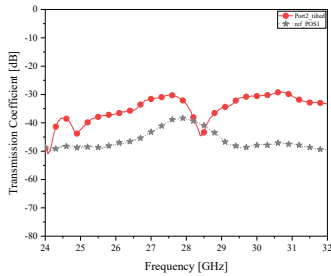


Figure 2. Directions of beams from the TX and RX antennas (a) in-line beam alignment; (b) beam alignment by cooperative TX and RX antennas.

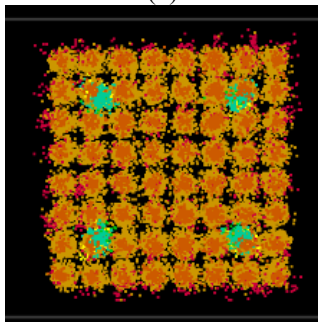
The microstrip-line Butler matrix antenna is substituted for the boresight antenna with a fixed beam. In addition, as case δ_{2L} , the -30° tilted beam is radiated to the TX horn antenna as in Figure 2, and S_{21} becomes -30 dB in Figure 7b where RF power transfer of the beamformers' beam in red is stronger than that of the non-tilting beam in gray. The increment in RF-to-RF connectivity is led to enhancement in wireless communication.



(a)



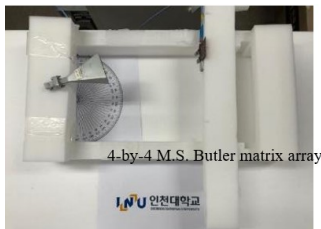
(b)



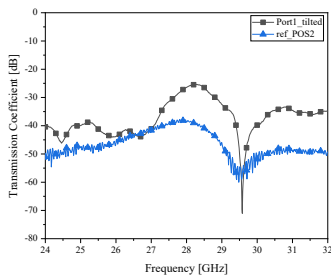
(c)

Figure 3. Beam alignment with higher tilting angle (a) in-line beam alignment (b) Transmissivity (c) 64-QAM

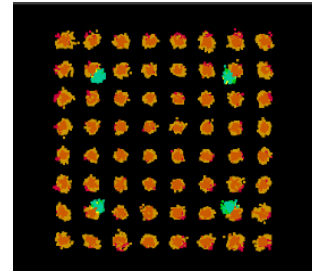
Figure 3 shows quite blurry constellations. It is inferred that the sidelobe of the RX antenna causes that. This will be mitigated by the reduced side lobe of next beam-angle.



(a)



(b)



(c)

Figure 4. Beam alignment with lower tilting angle (a) in-line beam alignment (b) Transmissivity (c) 64-QAM

Transmissivity in Figure 4 is better than that in Figure 3, which leads to very clear constellations. The lower tilting-angle tends to have a lower side-lobe level which turns out to be a higher quality of communication.

3 Conclusions

An intuitive method was suggested as a novel and practical attempt to interpret the characteristics of the beamforming antenna being forwarded to the performances of the system. Specifically, the beam-tilting and steering abilities of the TX and RX antennas are dealt with. For a horn antenna such as the TX, microstrip-line Butler matrix was built as the RX. The VNA as the RF-to-RF test setup and TRX7200 as the 64-QAM measurement apparatus were employed to measure the received power and constellations as the product of TX-to-RX electromagnetic connectivity via the beamforming antennas..

Acknowledgements

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No.PJ014762)" Rural Development Administration, Republic of Korea."

References

- [1] Shafi, M.; Molisch, A.F.; Smith, P.J. 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE J. Sel. Areas Commun.* 2017, 35, 1201–1221.
- [2] Chen, Y. Challenges and opportunities of Internet of things. In *Proceedings of the 17th Asia and South Pacific Design Automation Conference ASP-DAC, Sydney, Australia, 30 January–2 February 2012*; pp. 383–388.
- [3] Bleicher, A. The 5G phone future. *IEEE Spectr.* 2013, 50, 15–16.
- [4] Lee, C.; Park, H.; Lim, Y.; Kahng, S. A stacked array antenna able to make a very large gain and tilt the end-fire beam at the ISM-band. *Int. J. RF Microw. Comput.-Aided Eng.* 2019, 29, e21821.



ICFIC2021 VENUE

Daegu University

Daegu University was founded on the philosophies of good will and charity and with a motto of “Love, Light, and Freedom”.

Under these ideals, Daegu University has strived to provide opportunities to handicapped and socially-challenged people while simultaneously producing capable young men and women who are well-equipped to be the mainstays of a civil society and to maintain Daegu University’s founding ideals of philanthropy.

Educational Philosophy

To promote public welfare based on devotion to the welfare of humanity and the Christian Spirit.

Educational goals

- To create highly-skilled, creative, and proactive professionals
- To cultivate advanced human resources in national and social welfare
- To educate progressive and democratic citizens who will contribute to the development of society

University Mission

To cultivate creative human beings who aspire to actualize public welfare and contribute to social progress

General Information

Please visit <https://www.daegu.ac.kr/main>

Transportation

Incheon Airport

KTX Train from Incheon Airport:

Take the KTX train on the Incheon Airport Lower Level to Dongdaegu Station → Arrive at Dong Daegu Station → Take a taxi from Dong Daegu Station or Take city bus #814 or #708 to Daegu University

Limousine Bus Service from Incheon Airport:

Take the limousine bus at the Incheon Airport Arrival Floor (1F) Section 10C to Dong Daegu (departure every 40 minutes) → Arrive at the Dong Daegu Bus Transfer Center → Take a taxi from the Dong Daegu Bus Transfer Center or take city bus #708 to Daegu University

KTX Train from Seoul Station:

Take the limousine bus at the Incheon Airport Arrival Floor (1F) Section 4A or 10B to Seoul Station → Arrive at Seoul Station → Take the KTX express train to Dong Daegu Station → Arrive at Dong Daegu Station → Take a taxi from Dong Daegu Station or Take city bus #814 or #818 to Daegu University

Purchasing Bus Tickets:

The Transportation Information Counter is located next to Exit 2, 4, 9, and 13 on the Incheon Airport Arrival Floor (1F)

ICFIC

201 Daegudae-ro, Jillyang
Gyeongsan-si, Gyeongsangbuk-do, 38453, Republic of Korea
<https://icfic.daegu.ac.kr>

ICFIC

201 Daegudae-ro, Jillyang
Gyeongsan-si, Gyeongsangbuk-do, 38453, Republic of Korea
<https://icfic.daegu.ac.kr>

