

Ismail Mohd. Khairuddin · Muhammad Amirul Abdullah ·
Ahmad Fakhri Ab. Nasir · Jessnor Arif Mat Jizat ·
Mohd. Azraai Mohd. Razman · Ahmad Shahrizan Abdul Ghani ·
Muhammad Azzat Zakaria · Wan Hasbullah Mohd. Isa ·
Anwar P. P. Abdul Majeed *Editors*

Enabling Industry 4.0 through Advances in Mechatronics

Selected Articles from iM3F 2021,
Malaysia

Lecture Notes in Electrical Engineering

Volume 900

Series Editors

Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Naples, Italy

Marco Arteaga, Departament de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico

Bijaya Ketan Panigrahi, Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India

Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, Munich, Germany

Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China

Shanben Chen, Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

Rüdiger Dillmann, Humanoids and Intelligent Systems Laboratory, Karlsruhe Institute for Technology, Karlsruhe, Germany

Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China

Gianluigi Ferrari, Università di Parma, Parma, Italy

Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain

Sandra Hirche, Department of Electrical Engineering and Information Science, Technische Universität München, Munich, Germany

Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA

Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Alaa Khamis, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt

Torsten Kroeger, Stanford University, Stanford, CA, USA

Yong Li, Hunan University, Changsha, Hunan, China

Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA

Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore

Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany

Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA

Sebastian Möller, Quality and Usability Laboratory, TU Berlin, Berlin, Germany

Subhas Mukhopadhyay, School of Engineering & Advanced Technology, Massey University, Palmerston North, Manawatu-Wanganui, New Zealand

Cun-Zheng Ning, Electrical Engineering, Arizona State University, Tempe, AZ, USA

Toyoaki Nishida, Graduate School of Informatics, Kyoto University, Kyoto, Japan

Luca Oneto, Department of Informatics, Bioengineering., Robotics, University of Genova, Genova, Genova, Italy

Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi "Roma Tre", Rome, Italy

Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Gan Woon Seng, School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Joachim Speidel, Institute of Telecommunications, Universität Stuttgart, Stuttgart, Germany

Germano Veiga, Campus da FEUP, INESC Porto, Porto, Portugal

Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Beijing, China

Walter Zamboni, DIEM - Università degli studi di Salerno, Fisciano, Salerno, Italy

Junjie James Zhang, Charlotte, NC, USA

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

More information about this series at <https://link.springer.com/bookseries/7818>

Ismail Mohd. Khairuddin ·
Muhammad Amirul Abdullah ·
Ahmad Fakhri Ab. Nasir ·
Jessnor Arif Mat Jizat ·
Mohd. Azraai Mohd. Razman ·
Ahmad Shahrizan Abdul Ghani ·
Muhammad Aizzat Zakaria ·
Wan Hasbullah Mohd. Isa ·
Anwar P. P. Abdul Majeed
Editors

Enabling Industry 4.0 through Advances in Mechatronics

Selected Articles from iM3F 2021, Malaysia

 Springer

Editors

Ismail Mohd. Khairuddin
Universiti Malaysia Pahang
Pekan, Malaysia

Muhammad Amirul Abdullah
Universiti Malaysia Pahang
Pekan, Malaysia

Ahmad Fakhri Ab. Nasir
Faculty of Computing
Universiti Malaysia Pahang
Pekan, Malaysia

Jessnor Arif Mat Jizat
Faculty of Manufacturing and Mechatronic
Engineering Technology
Universiti Malaysia Pahang
Pekan, Malaysia

Mohd. Azraai Mohd. Razman
Faculty of Manufacturing Engineering
Universiti Malaysia Pahang
Pekan, Malaysia

Ahmad Shahrizan Abdul Ghani
Faculty of Manufacturing and Mechatronic
Universiti Malaysia Pahang
Pekan, Malaysia

Muhammad Aizzat Zakaria
Faculty of Manufacturing and Mechatronic
Engineering Technology
Universiti Malaysia Pahang
Pekan, Malaysia

Wan Hasbullah Mohd. Isa
Universiti Malaysia Pahang
Pekan, Malaysia

Anwar P. P. Abdul Majeed
Faculty of Manufacturing and Mechatronic
Engineering Technology
Universiti Malaysia Pahang
Pekan, Malaysia

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-981-19-2094-3

ISBN 978-981-19-2095-0 (eBook)

<https://doi.org/10.1007/978-981-19-2095-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The second edition forum of The Innovative Manufacturing, Mechatronics & Materials Forum 2021 (iM3F 2021) organized by Universiti Malaysia Pahang through its Faculty of Manufacturing and Mechatronic Engineering Technology was held on September 20, 2021. The main field focuses on manufacturing, mechatronics as well as materials.

More than 132 submissions were received during iM3F 2021 and were reviewed in a single-blind manner, and 49 papers were advocated by the reviewers to be published in this lecture notes in electrical engineering. The editors would like to express their gratitude to all the authors who submitted their papers. The papers published in this proceeding have been thoroughly reviewed by the appointed technical review committee consisting of various experts in the field of mechatronics.

The conference had brought a new outlook on cutting-edge issues shared through keynote speeches by Prof. Ir. Dr. Jamaluddin Mahmud and Prof. Dr. Mohammad Osman Tokhi.

Finally, the editors hope that readers find this volume informative as we thank LNEE for undertaking this volume publication. We also would like to thank the conference organization staff and the international program committees' members for their hard work.

Pekan, Malaysia
November 2021

Ismail Mohd. Khairuddin
Muhammad Amirul Abdullah
Ahmad Fakhri Ab. Nasir
Jessnor Arif Mat Jizat
Mohd. Azraai Mohd. Razman
Ahmad Shahrizan Abdul Ghani
Muhammad Aizzat Zakaria
Wan Hasbullah Mohd. Isa
Anwar P. P. Abdul Majeed

Contents

Mapping and Navigation for Indoor Robot Using Multiple Sensor Under ROS Framework	1
Gigih Priyandoko and M. S. Hendriyawan Achmad	
Optimization of Waterjet Paint Removal Operation Using Artificial Neural Network	11
Abdullah Faisal Alzaghir, Mohd Nazir Mat Nawi, M. A. Gebremariam, and Azmir Azhari	
Light Path Simulation for Optical Switch Based on Digital Electromagnetic Actuators	21
Thi Thanh Vi Nguyen, Nguyen Duy Minh Phan, Ngo Quoc Huy Tran, Le Anh Doan, and Quang Truong Vo	
An Application of Charge-Coupled Device (CCD) Tomography System for Gemological Industry - A Review	31
Fatinah Mohd Rahalim, Juliza Jamaludin, Syarfa Najihah Raisin, Wan Zakiah Wan Ismail, Irneza Ismail, Ruzairi Abdul Rahim, and Yasmin Abdul Wahab	
Prediction of Abrasive Waterjet Machining of Sheet Metals Using Artificial Neural Network	43
Nur Khadijah Mazlan, Nazrin Mokhtar, M. A. Gebremariam, and Azmir Azhari	
You Are Too Loud! Classification of Psychological Conditions for Stress Detection System Using Galvanic Skin Response	51
Amirul Aidy Amiruldin and Aimi Shazwani Ghazali	
Universiti Malaysia Pahang Autonomous Shuttle Development: Lane Classification Analysis Using Convolutional Neural Network (CNN) . . .	61
Lee Yin Yee and Muhammad Aizzat Zakaria	

Eco-Design of Electric Vehicle Battery Pack for Ease of Disassembly	71
X. Q. Chew, W. J. Tan, N. Sakundarini, C. M. M. Chin, A. Garg, and S. Singh	
Electric Vehicle Drive Specification Modelling for Three Wheels Scooter Configuration	85
M. Faris, F. R. M. Romlay, A. R. Razali, M. R. Hanifah, and A. Ghazali	
Experimentation on Spectra Data Regression Using Dense Multilayer Neural Networks with Common Pre-processing	97
Kok Pin Chan, Mahmud Iwan Solihin, Chun Kit Ang, and Liew Phing Pui	
A Piecewise Affine System Modeling Approach of Thin McKibben Muscle Servo Actuator	113
Mohd Akmal Mhd Yusoff, Ahmad Athif Mohd Faudzi, Mohd Shukry Hassan Basri, and Mohd Fuaad Rahmat	
A Comparison of Type 1 and Type 2 Fuzzy Logic Controller for DC Motor System	125
Nafrizuan Mat Yahya, Nurainaa Elias, and Muhammad Hazim Mohd Nordin	
Modelling and PID Control of SEPIC-Boost Converter for 48 V Miniature Mild Hybrid System	135
M. N. Mamat, M. N. Abdullah, S. Kaharuddin, and D. Ishak	
Particle Swarm Optimization with Multi-Chaotic Scheme for Global Optimization	145
Wy-Liang Cheng, Koon Meng Ang, Cher En Chow, Wei Hong Lim, Sew Sun Tiang, Chun Kit Ang, Balaji Chandrasekar, and Eik Cheng	
Flexible Wearable Microstrip Antenna with DGS and Shorting Post for WBAN Application	157
Liu Yanan, Sew Sun Tiang, Wei Hong Lim, and Xiang Sheng Wong	
PID Controller Based on Bird Mating Optimizer for Vibration Cancellation of Horizontal Flexible Plate	169
Muhammad Farith Nazreen Abu Samah, Muhamad Sukri Hadi, Abdul Malek Abdul Wahab, Nadzirah Mohd Mokhtar, and Intan Zaurah Mat Darus	
Hub Angle Control of Flexible Manipulator Based on Bacterial Foraging Optimization	181
Muhammad Nazri Mohd Salme, Muhamad Sukri Hadi, Annisa Jamali, Hanim Mohd Yatim, Mat Hussin Ab. Talib, and Intan Zaurah Mat Darus	

Position and Attitude Control of Quadrotor MAV Using Sliding Mode Control with Tanh Function 193
 Aminurrashid Noordin, Mohd Ariffanan Mohd Basri,
 Zaharuddin Mohamed, and Izzuddin Mat Lazim

On the Partial Transmit Sequences (PTS) Optimization Using Firefly Algorithm for PAPR Reduction in OFDM Systems 205
 Hazmarini Husin, Rosmiwati Mohd-Mokhtar,
 and Aeizaal Azman A. Wahab

A Comparison of Machine Learning and Deep Learning in Hyperspectral Image Classification 221
 Fady Mohamed Sadek, Mahmud Iwan Solihin, Fahri Heltha,
 Lim Wei Hong, and M. Rizon

Identify Target Area of Panel for Spraying Using Convolutional Neural Network 237
 Y. K. Chin, S. S. N. Alhady, A. A. A. Wahab, W. A. F. W. Othman,
 E. A. Bakar, and M. N. Akhtar

Navigation of Cart Follower Using Artificial Intelligence 251
 C. Y. Ooi, S. S. N. Alhady, A. A. A. Wahab, W. A. F. W. Othman,
 E. A. Bakar, and M. N. Akhtar

Object Tracking for Autonomous Vehicle Using YOLO V3 265
 William Chin Wei Hung, Muhammad Aizzat Zakaria, M. I. Ishak,
 and P. M. Heerwan

A Brief Review on Motion Sickness for Autonomous Vehicle 275
 Sarah ‘Atifah Saruchi, Nor Aziyatul Izni, Mohd Hatta Mohammed Ariff,
 and Nurbaiti Wahid

Investigation of the Combination of Kinematic Path Planning and Artificial Potential Field Path Planning with PI Controller for Autonomous Emergency Braking Pedestrian (AEB-P) System 285
 Z. Abdullah, P. M. Heerwan, M. A. Zakaria, and M. I. Ishak

Simulation Studies on Bimorph and Unimorph PZT Piezoelectric Transducer for Energy Harvesting Application 299
 Sanjeev Singh A/L Sarbjeet Singh, Sew Sun Tiang, Wei Hong Lim,
 Kah Hou Teng, and Chin Hong Wong

Simulated Kalman Filter with Modified Measurement, Substitution Mutation and Hamming Distance Calculation for Solving Traveling Salesman Problem 309
 Suhazri Amrin Rahmad, Zuwairie Ibrahim, and Zulkifli Md Yusof

Development of Heart Rate Sensor Warning System to Estimate driver’s Cognitive Distraction Level 321
 Akmal Shahmi Bin Azhar and Ahmad Khushairy Bin Makhtar

Gain Scaling Tuning of Fuzzy Logic Sugeno Controller Type for Ride Comfort Suspension System Using Firefly Algorithm	335
Mat Hussin Ab Talib, Intan Zaurah Mat Darus, Hanim Mohd Yatim, Muhamad Sukri Hadi, Mohd Syahril Ramadhan Mohd Saufi, and Nor Hasrul Akhmal Ngadiman	
Evaluation of a DC Motor Temperature Response Characterization Method Under Different Sampling Interval	345
I. A. Kamaruddin, M. A. H. Rasid, N. F. Abdullah, and A. Abdul Wahab	
Intelligent Optimization of Novel Particle Swarm Optimization with Explorer (PSOE) for Identification of Flexible Manipulator System . . .	361
Hanim Mohd Yatim, Ahmad Nur Yussuf Zamri, Muhamad Sukri Hadi, Mat Hussin Ab. Talib, and Intan Zaurah Mat Darus	
SCADA System for Industrial Manipulator PLC Trainer	375
Nur Hanisah Azman, Muchamad Oktaviandri, and Muhammad Rizal Ramedan	
A Self-adaptive Hybridized Algorithm with Intelligent Selection Scheme for Global Optimization	387
Zhi Chuan Choi, Koon Meng Ang, Cher En Chow, Wei Hong Lim, Sew Sun Tiang, Chun Kit Ang, and Balaji Chandrasekar	
Development of Small Scale Machine for Non-Destructive Testing (NDT) to Assess Structural Integrity of Composite	399
Mohamad Akmal Ridzuan Bin Radzi, Mohammad Nishat Akhtar, Jazmina Binti Ja'afar, Elmi Abu Bakar, S. S. N. Alhady, and Mohamad Nazir Abdullah	
Hands and Fingers Tracking for Tactile Graphics Reading Assistive Device	413
Muhammad Ikmal Hakim Shamsul Bahrin, Hazlina Md Yusof, and Shahrul Na'im Sidek	
Assembly Sequence Optimization Using the Bees Algorithm	423
Shafie Kamaruddin, Nabilah Azmi, and Nor Aiman Sukindar	
The Application of Conjugate Gradient Method to Motion Control of Robotic Manipulators.	435
Ibrahim M. Sulaiman, Maulana Malik, Wed Giyarti, Mustafa Mamat, Mohd Asrul Hery Ibrahim, and Muhammad Zaini Ahmad	
Analysis of Multiple Prediction Techniques of Received Signal Strength to Reduce Surveying Effort in Indoor Positioning	447
Mohd Amiruddin Abd Rahman, Caceja Elyca Anak Bundak, and Muhammad Khalis Abdul Karim	

An Approach to Neutral Steering of a 4WIS Vehicle with Yaw Moment Control 459
 Li Maoqi, M. I. Ishak, P. M. Heerwan, and Muhammad Aizzat Zakaria

Investigation of Electroencephalogram (EEG) Sensor Position for Brain-Controlled Home Automation 471
 Norizam Sulaiman, Ailis Aimyilia Hasim, Md Nahidul Islam, Mahfuzah Mustafa, and Mohd Shawal Jadin

Modeling and Analysis of Omnidirectional Wheeled Vehicles Using Velocity-Based Impedance Control 485
 Norsharimie Mat Adam and Addie Irawan

An Observation of Different Clustering Algorithms and Clustering Evaluation Criteria for a Feature Selection Based on Linear Discriminant Analysis 497
 K. H. Tie, A. Senawi, and Z. L. Chuan

Recent Advances and Open Challenges in RFID Antenna Applications 507
 Nazmus Sakib Khair, Nurhafizah Abu Talip Yusof, Mohd Hisyam Mohd Ariff, Yasmin Abdul Wahab, and Bifta Sama Bari

Investigating Vehicle Characteristics Behaviour for Roundabout Cornering 519
 Ashaa Supramaniam, Muhammad Aizzat Zakaria, Mohamad Heerwan Peeie, and Gerogious Papaioannou

Rice Disease Identification Through Leaf Image and IoT Based Smart Rice Field Monitoring System 529
 Md Nahidul Islam, Fahim Ahmed, Md Tanvir Ahammed, Mamunur Rashid, and Bifta Sama Bari

Chili Plant Classification Using Transfer Learning Models Through Object Detection 541
 Muhammad Nur Aiman Shapiee, Amirul Asyraf Abdul Manan, Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin, and Anwar P. P. Abdul Majeed

The Classification of Impact Signal of 6 DOF Cobot by Means of Machine Learning Model 553
 Gavin Lim Jiann Kai, Ahmad Fakhri Ab. Nasir, Anwar P. P. Abdul Majeed, Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin, and Lim Thai Li

A Cluster Analysis of Identifying Team and Individual Sports Athlete Based on Anthropometric, Health and Skill Related Components 561
Noor Aishah Kamarudin, Mohamad Razali Abdullah,
Vijayamurugan Eswaramoorthi, Ahmad Bisyril Husin Musawi Maliki,
Aina Munirah Ab Rasid, Anwar P. P. Abdul Majeed,
Mohd Azraai Mohd Razman, and Rabiu Muazu Musa

Vision-Based Human Presence Detection by Means of Transfer Learning Approach 571
Jin Cheng Tang, Ahmad Fakhri Ab. Nasir, Anwar P. P. Abdul Majeed,
Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin, and Thai Li Lim

Mapping and Navigation for Indoor Robot Using Multiple Sensor Under ROS Framework



Gigih Priyandoko and M. S. Hendriyawan Achmad

Abstract A few criteria needed to be met in order to develop a mobile robot mapping system: wireless control of the mobile robot and obstacle detection. In this research, a Robot Operating System (ROS) is developed as the framework for the mobile robot control system. The laser range finder and Kinect camera serve as tools for detecting obstacles in the mobile robot's environment. The Hector mapping package is used to create a 2D/3D map based on the object detected by the laser range finder. The Hector package's parameters are investigated in order to determine the best map layout quality. Based on the experimental results, the first step is to develop a Wi-Fi network with the ROS framework for wireless sensor data collection and command execution. Second, the use of hector mapping allows for the creation of 2D/3D maps with minimal error. The development of a mobile robot control system allows for the execution of mapping missions and exploration of unknown environments without putting human lives at risk.

Keywords ROS · Mapping · Indoor · Robot

1 Introduction

The mobile robot has evolved in modern technology to perform a variety of tasks that benefit humanity. Exploration of dangerous areas, in particular, is one of the important roles that mobile robots can play in order to reduce risk to humans, reduce costs, and improve efficiency [1]. The generation of maps is a critical task for the mobile robot.

G. Priyandoko (✉)

Department of Electrical Engineering, Faculty of Electrical Engineering, University of Widyagama, 65125 Malang, Indonesia
e-mail: gigih@widyagama.ac.id

M. S. H. Achmad

Department of Electrical Engineering, Faculty of Science and Technology, Universitas Teknologi Yogyakarta, 55285 Yogyakarta, Indonesia
e-mail: hendriyawanachmad@uty.ac.id

Many successful mobile robot systems are able to execute their tasks because they use maps as guidance to achieve their goal [2]. In this paper, we only focus on the mapping system of the mobile robot.

Robot Operating System (ROS) the mobile robot's operating system for data collection, visualization, and control. A sensor is needed to collect data from the environment in order for a mobile robot to complete the mapping task. With the movement of the mobile robot along with the unknown environment, a safe and good quality collision-free path able to be obtained and a map is built along the process [3]. The map generated by the mobile robot is widely used in the automated mobile robot sector because it provides important information about the obstacle as well as a clear path for an automated robot to generate a path to the destination without colliding. The map can also help people visualize the layout of their surroundings in order to determine the goal location and detect hazards.

Beno et al. [4] had implemented a mobile robot 3D mapping system by using RGB-D Camera and Octree Maps. In his research, Visual Simultaneous Localization and Mapping (SLAM) had been achieved by using RGB-D Kinect. The depth-sensing camera in Kinect is used to estimate the movement between two frames, allowing the mobile robot to be localized. Instead of using wheel odometry, he used visual odometry to pinpoint the mobile robot's location on a map. The frame-matching approach was used for the mapping section, which involved identifying key points from both frames and paring them together for estimation transformation.

Yu et al. [5] had developed an omnidirectional automated guided vehicle with a MY3 wheel. The MY3 wheel used in his project is more resistant to dirt and fragment on the ground. As four spherical wheels in one set capable of mutual support, the MY3 wheel has a good load-carrying capability when compared to the omni-directional wheel. Optical color sensors and cameras are used to guide the project. The optical color sensor guides the AGV along a predetermined path, while the camera allows the operator to observe the working environment.

Poulaian [6] developed navigation of an Autonomous Robot Vehicle using optical rangefinder and odometry. The system necessitates the use of an off-line path planner for the vehicle. To move the vehicle from one location to another, a series of collision-free guiding tools with line and arc segments is required. A 2D representation was used to collect the line segment in the plane. The system is made up of two algorithms: one for matching the sensory data to the map and another for estimating the precision of the corresponding match or correction.

Li and Shavkin [7] had presented a mobile robot navigation system using a sensor network. This navigation system does not require a pre-decided environment map or access to GPS, IMU or a compass. The workspace is preinstalled with nodes, and navigation occurs through node wise motion, in which the robot moves from node to node, beginning with the starting node and ending with the goal node. When a destination is specified, the navigation field computation is triggered by the node that is closest to the goal node. During this computation, the optimal direction in which the robot should move is determined probabilistically. The navigation field indicates the "best possible" path for the robot to take in order to reach the goal.

Zhang et al. [8] had presented a motion planning of a mobile robot using the potential field method. The environment is a dynamic in which both the target and the obstacle move. They begin by defining a new potential function and the virtual force that corresponds to it. The use of attractive and repulsive potential for target and obstacle allows the robot to track the target in the desired direction. The motion planning is carried out with the assistance of an online measurement of the obstacle information.

Kim et al. [9] proposed a fully 3D thermal modelling of a building with an autonomous system. A mobile robot with a laser sable toner, thermal camera, and webcam attached to the top of its body performs a 3D mapping task. The robot can fit the plane model to the data points by using the ICP (Iterative Closest Point) algorithm. The calibration and matching of data collected from all sensors at various locations results in the creation of a common reference frame. As a result, the goals of this research are to create a wireless control system for smooth mobile robot operation as well as a 2D/3D mapping system using a laser sensor and a Kinect camera.

2 Robot Operating System

In recent years, a new software framework known as Robot Operating System (ROS) based on Linux has been developed for robot software development. ROS aims to be peer-to-peer, tool-based, multilingual, thin, and open-source. [10, 11]. Robots or systems can communicate with one another thanks to peer-to-peer topology. The ROS framework is built on the Real-Time Publisher-Subscriber Protocol (RTPSP), which allows different programs to be defined as nodes and interact with each other independently through message subscribing and publishing. This framework solved the previous complexity in robotics, in which every task was developed in a single program, causing the entire program to shut down if one tiny part went wrong. If one node fails in ROS, the other nodes continue to function normally, allowing the user to easily identify the problem and perform the debugging task. Figure 1 depicts the fundamental concept of ROS with distributed system capabilities.

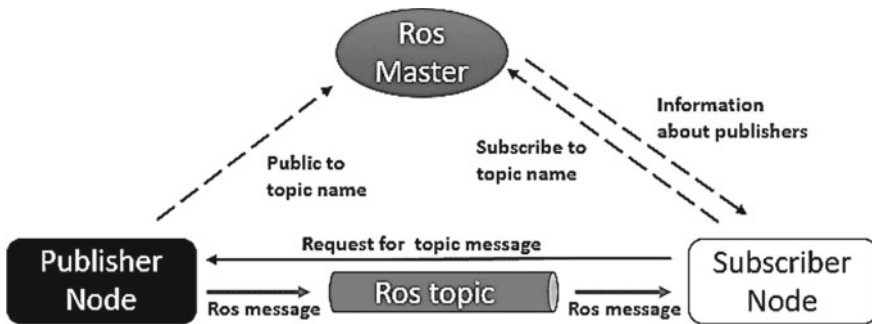


Fig. 1 ROS basic concept

ROS is an open-source Meta operating system that could abstract hardware, low-level device control, message passing, and package management. ROS differs from other robotic software platforms as it supports code reuse in robotic research and development, thus it is suitable for robotics research [10]. ROS is a language-neutral operating system that supports four different languages (C++, Python, Octave, and LISP). It consists of node and topic in ROS. The node includes the ability to process information, such as converting data collected from sensors into information that users can read. A topic serves as a link between topics for information transfer. Without a topic, nodes that have processed data are unable to send it to another node because they are unable to recognize each other for data transmission. As a result, data transmission is only possible when the next node is subscribed to the topic published by the previous topic. ROS platform has several advantages, including simple autonomy programming, modularity, flexibility, reusability, and a clear design. ROS is distinguished by its powerful and robust communication system, which is appropriate for users in a multimember team. Although ROS comes with various advantages, it consists of a barrier, which is, it only able to operate in the Linux platform [12–15].

3 Methodology

3.1 Mobile Robot Setup

Acrylic Glass is used to make the body of this mobile robot. Mobile robot's basic dimensions are 0.4 (width) \times 0.4 (length) \times 0.75 m (height). An Arduino mega serves as the base controller for this mobile robot, which also includes a Hokuyo laser sable toner and a Kinect camera. As shown in Fig. 2, a laptop will be placed in the center of the mobile robot to connect all of the mobile robot's sensors and base controller.

The rigid body transformation of a mobile robot used for exploration to generate the map is critical and required. The robot cannot generate a map without transformation information. The robot should be aware of the relative position of all movable part frames in the world frame so that the robot pose can be accurately estimated.

Figure 3 depicts the mobile robot's transformation tree during experiments. The odometry frame is a virtual frame that is always linked to the robot base frame and the IMU frame. This frame is an estimate of the robot's pose in relation to the world frame. The robot's transformation tree employs a parent and child diagram, with the child in a relative pose to the parent. The relative pose is made up of both linear and angular values. The linear values are relative translations (x , y , z) in meters, and the angular values are relative rotations (qx , qy , qz , qw) in quaternions.

Fig. 2 Mobile robot

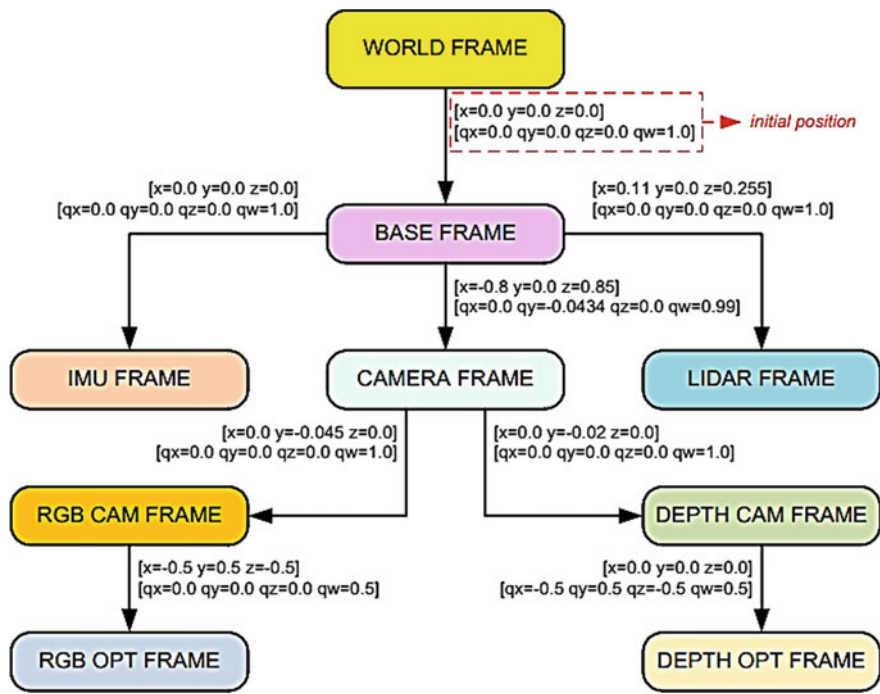
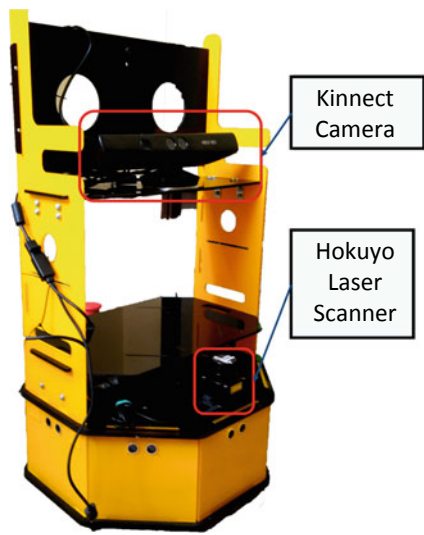


Fig. 3 Mobile robot body transformation

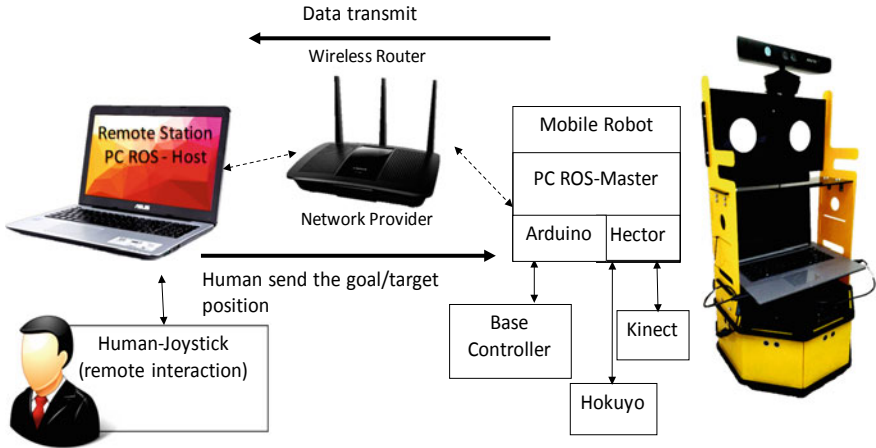


Fig. 4 Basic hardware setup

3.2 Initial Hardware Setup

Two laptops were installed with the Ubuntu operating system during the initial stage of apparatus setup. The reason for installing Ubuntu is that the Robot Operating System (ROS) can only run on Linux-based systems. After installing all necessary software on both laptops, one of them serves as the master, connecting to all of the hardware on the mobile robot, which includes the Kinect RGB-D camera, Hokuyo range sensor, and base controller. The prototype robot was equipped with a master computer, allowing it to move freely and without restriction. Another laptop acts as the Host where it act as the user control interface for the user to visualize the data collected from the mobile robot. The Host laptop will be used to control the prototype robot wirelessly by using the joystick as shown in Fig. 4. To establish a wireless connection between two computers, the wireless router had been used. By connecting both computers to the same Wi-Fi channel, the mobile robot able to transfer data from the sensor to the host computer and vice versa.

3.3 Initial Software Setup

A Wi-Fi network channel is used to connect two laptops. Using the ROS setting, the Host laptop can gain full access to the Master Laptop by using the Master Laptop's IP address. After both computers are connected to the same Wi-Fi network, data collection and command can be performed on the Master laptop by using the Host laptop. After connecting the Host and Master laptops, the next step is to connect all of the sensors and the base controller. Freenect node has been launched to connect the RGB-D camera (Kinect) with the Master laptop. For creating a connection between the

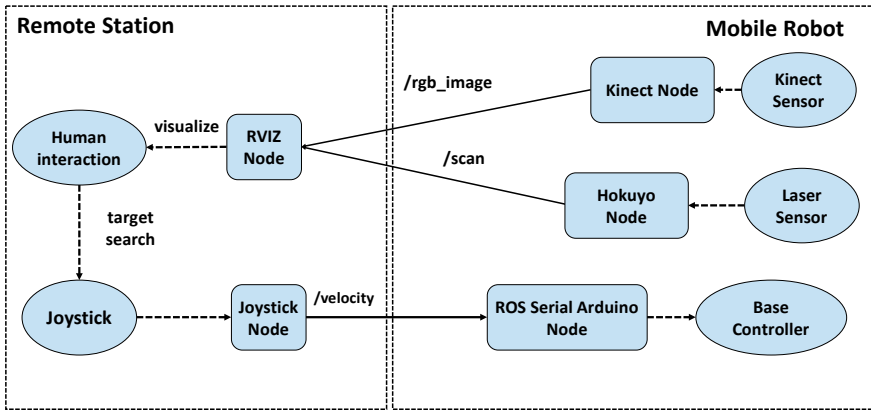


Fig. 5 Basic structure of the framework for mobile robot

Master laptop with the hokuyo sensor, hokuyo_node had been launch. The topic that has been published by freenect node are /rgb_image and /sable to from hokuyo_node. Both topic that been published (/rgb_image and /sable to) is then subscribed by rviz node which has the function of visualizing the entire signal that collected. Data published by the rviz node is capable of being received by the Host laptop via Wi-Fi data transmission. A few nodes had been launched in order to establish a connection for the user to gain control of the robot. The joystick node was used to convert joystick commands into a signal that the robot’s base controller could read. As shown in Fig. 5, the ROS serial Arduino node was launched for the base controller to subscribe to the signal published by the joystick node, allowing the user to gain control over the movement of the mobile robot.

4 Results and Discussion

Some preparations had been made prior to carrying out the mapping operation. When the mobile robot was running, the tf li-brary had to be run to create a tracking on multiple coordinate frames. The Tf function on the Hokuyo sensor and RGB-D Kinect had been established in relation to the base link (center base of mobile robot). A measuring tape was used to determine the distance between the Hokuyo sensor and the RGB-D Kinect and the base of the mobile robot. With the initial software setup complete, the user can visualize the robot’s surroundings by utilizing all of the mobile robot’s sensors. With the proper configuration, the user can use the joystick to give commands to the robot, and the mobile robot will move accordingly. The RVIZ (ROS Visualizer) window can display all updated results and conditions of the robot when the mobile executes the user’s command. The process is repeated until the user is satisfied with the result or achieves the goal.

A parameter of the Hector package must be changed in order to create a good map. To fit the mobile robot's movement, two parameters must be adjusted. These are the variables. The thresholds for performing map updates are map update distance thresh and map update angle thresh. map update distance thresh is the threshold for performing map updates; the map will be updated after the mobile robot has traveled a certain amount of distance. The map update angle thresh parameter controls the threshold for performing map updates based on the angle turned by the mobile robot. The goal of changing the parameter is to ensure that the map is nicely mashed and that no information is lost on the map.

A bag file is created to ensure that the mobile robot's path and movement are identical for each parameter set. A bag file is a file that records and saves all of the information that the laser sable toner generates. The ideal parameter for the Hector package can be identified as the true response of the laser sable toner by playing back the bag file.

Figures 6 and 7 depicts the occupied map from our experiment explorations and visualize it using RVIZ (ROS Visualizer) on a 2D/3D RGB image as a map of an unknown indoor area. Despite using the best parameter for the angular and distance update thresholds, there is some error in the map generated by the mobile robot. When the mobile robot explores an unknown area with a lack of features, the error occurs as a result of accumulating visual odometry error. The Hector package's mapping algorithm works by matching the previous frame of the map. As the map exploration process continues, the error will accumulate.



Fig. 6 Constructed 2D map

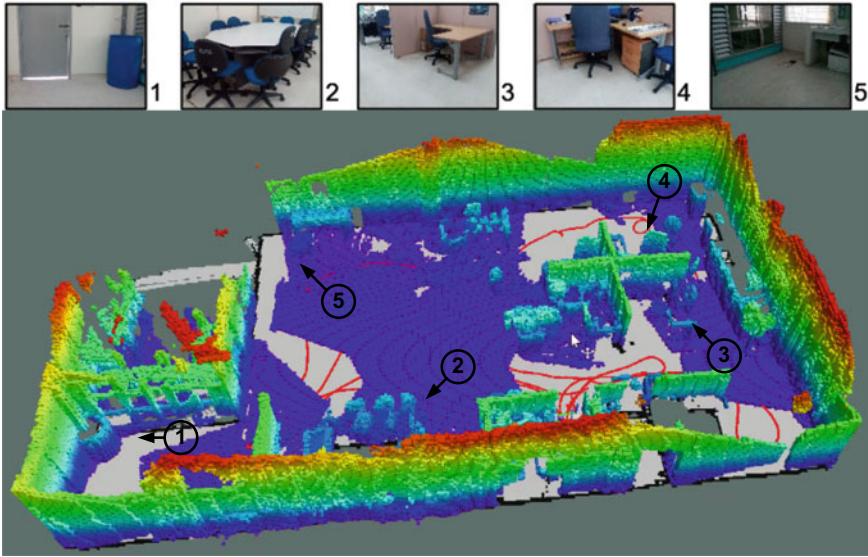


Fig. 7 Visualize 2D/3D occupied map

Table 1 Comparison the generated 2D map with the real layout

Real layout dimension (m)	Generated map dimension (m)	Percentage of error (%)
11.55	11.67	1.04
13.67	13.83	1.17

Figure 6 depicted the dimensions of the real layout and map generated by the mobile robot.

The percentage of error is calculated to compare the error between real and generated maps, as shown in Table 1. As a result, it is possible to conclude that the difference between the layout and the map is insignificant. Despite the fact that the generated map contains close loop error, the percentage of error is less than 2%. An increase in the range of laser scanner is proposed to solve the unexplored map error. One of the specs for each laser scanner is the scanning range. A higher scanning range laser scanner will be able to detect obstacles further away from the scanner, increasing mapping efficiency as the mobile robot's moving distance is reduced. Although laser scanner with a high scanning range is preferred for mapping operations, it is expensive.

5 Conclusion

Using the ROS framework, a wireless mobile robot control system was built. The user is able to execute a wireless mobile robot control by using joystick control by utilizing the Wi-Fi network. The use of a mapping package and laser scanner allows the user to visualize the environment on the robot's surroundings. The mobile robot can perform map exploration tasks to explore an unknown area by combining a 2D/3D mapping system with a wire-free control system. The accuracy and scale of the mobile robot control system are limited due to the limited budget. In the future, the system will be able to be improved by having better sensors and more accurate algorithms.

References

1. Benkrid A, Benallegue A, Achour N (2016) Robot's energy consumption based multi-robot exploration strategy, pp 1129–1134
2. Masehian E, Jannati M, Hekmatfar T (2017) Cooperative mapping of unknown environments by multiple heterogeneous mobile robots with limited sensing. *Robot Auton Syst* 87:188–218
3. Luo C, Jan GE, Zhang J, Shen F (2016) Boundary aware navigation and mapping for a mobile automaton, pp 561–566
4. Beno P, Pavelka V, Duchon F, Dekan M (2016) Using octree maps and RGBD cameras to perform mapping and navigation, pp 66–72
5. Yu S, Ye C, Liu H, Chen J (2016) Development of an omnidirectional automated guided vehicle with MY3 wheels. *Perspect Sci* 7:364–368
6. Poulaiin T (2019) Path generation and control of autonomous robot. *Int J Innov Technol Interdisc Sci* 2(3):200–211
7. Li H, Savkin AV (2018) An algorithm for safe navigation of mobile robots by a sensor network in dynamic cluttered industrial environments. *Rob Comput-Integr Manuf* 54:65–82
8. Zhang X, Wang J, Fang Y, Yuan J (2018) Multilevel humanlike motion planning for mobile robots in complex indoor environments. *IEEE Trans Autom Sci Eng* 16(3):1244–1258
9. Kim P, Chen J, Kim J, Cho YK (2018) SLAM-driven intelligent autonomous mobile robot navigation for construction applications, pp 254–269
10. Koubãa A (2017) *Robot Operating System (ROS)*, vol 1. Springer, Heidelberg
11. Shekhawat AS, Rohilla Y (2020) Design and control of two-wheeled self-balancing robot using arduino, pp 1025–1030
12. Joseph L (2018) *Robot Operating System (ROS) for Absolute Beginners*. Springer, Heidelberg
13. Dieber B, Breiling B, Taurer S, Kacianka S, Rass S, Schartner P (2017) Security for the robot operating system. *Robot Auton Syst* 98:192–203
14. Guan W, Chen S, Wen S, Tan Z, Song H, Hou W (2020) High-accuracy robot indoor localization scheme based on robot operating system using visible light positioning. *IEEE Photonics J* 12(2):1–16
15. Breiling B, Dieber B, Schartner P (2017) Secure communication for the robot operating system, pp 1–6

Optimization of Waterjet Paint Removal Operation Using Artificial Neural Network



Abdullah Faisal Alzaghbir, Mohd Nazir Mat Nawi, M. A. Gebremariam, and Azmir Azhari

Abstract Paint removal of automotive parts without environmental effects has become a critical issue around the world. The high pressure waterjet technology has received a wider acceptance for various applications involving machining, cleaning, surface treatment and material cutting. It offers an advantage to remove the automotive paint due to its superior environmental benefits over mechanical cleaning methods. Therefore, it is important to predict the waterjet cleaning process for a successful application for the paint removal in the automotive industry. In the present work, ANN model was used to predict the surface roughness after the paint removal process of automotive component using the waterjet cleaning operation. A response surface methodology approach was employed to develop the experimental design involving the first order model and the second order model of central composite design. Into training and testing, a back-propagation algorithm used in the ANN model has successfully predicted the surface roughness with an average of 80% accuracy and 3.02 mean square error. This summarizes that ANN model can sufficiently estimate surface roughness in waterjet paint removal process with a reasonable error range.

Keywords Waterjet cleaning · Paint removal · Multiple jet passes · Surface roughness · Artificial neural network

A. F. Alzaghbir (✉) · M. N. M. Nawi · A. Azhari
Faculty of Manufacturing Engineering and Mechatronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Malaysia
e-mail: abdullahalzaghbir@gmail.com

M. N. M. Nawi
Centre for Foundation Studies, International Islamic University Malaysia, 26300 Gambang, Malaysia

M. A. Gebremariam
College of Engineering, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia

1 Introduction

In the automotive component recycling industry, a paint removal procedure for automobile coatings is commonly employed. The need to reuse and recycle automotive parts while avoiding additional pollution from the removal process of paint is currently a major industrial problem across the world. The main goal of the paint removal procedure is to achieve high cleaning efficiency and erosion of paint with little damage to the substrate. Most of current research on cleaning processes focuses on the laser cleaning method [1]. However, considering laser cleaning is approximately 20 times more costly than traditional methods, its use is confined to historical and ornamental purposes [2]. Despite its fast growth, several studies on stone emission using laser ablation technology have shown that the laser cleaning process produces environmental health hazards and fails to meet health and safety standards [3].

High pressure waterjet technology is a relatively new process that can be used for machining, material removal, cleaning and surface treatment of various materials [4]. A cleaning operation using this technology has recently gained popularity to remove paint due to its superior environmental benefits over mechanical cleaning methods such as sandblasting, water brushing, hydropneumatics cleaning and controlled dry sanding [5]. Few studies have investigated the feasibility of using the waterjet cleaning in the paint removal process. However, very little attention has been given to the prediction of waterjet cleaning in removing the paint especially for automotive components. Therefore, it is important to predict the waterjet cleaning process for a successful application of the paint removal and to optimise the cleaning parameters in the automotive industry to get a desired result of roughness. Many studies have reported about the influence of WJ parameters, such as standoff distance, traverse rate and pressure, etc. on the cleaning output [6–9]. A series of experiments by previous studies were conducted to obtain the relationship between the cleaning parameters and the cleaning results. Most of these experiments were based on data fitting and analytical methods, however those methods were insufficient because of their possibility to lead to larger error between the real values and the predicted values [10].

Artificial neural networks (ANNs) are type of computer software that can learn from the results of previous experiments and predict the results of future ones. An artificial neural network consists of various artificial neurons that operate like synapses in the biological brain, transmitting messages to the other neurons to which they are linked [9]. The output from one node is used as an input for other node, thus data flows across the network of nodes and layers [9]. ANN was found to be a great tool to express the relationship between the cleaning parameters and the cleaning results and that is due to its ability to approximate such nonlinear and complex relationships [11]. Thus, the use of ANN in this study is a sufficient method to optimise the cleaning parameters to obtain and predict the surface roughness after the pain removal process. Many studies have shown that ANN can be used to forecast the surface finish in a variety of machining techniques, but there is little literature about their application

in the paint removal process using waterjet cleaning operation. Furthermore, as the number of machining factors increases, applicable experimental procedures require a large number of trials. Çaydaş & Hasçalık [12] employed a Taguchi approach with L_{27} orthogonal array to achieve a minimum number of experiments in developing an ANN model to predict the surface roughness during the abrasive waterjet (AWJ) machining process of aluminium alloy 7075-T6. Madara et al. [9] also developed an ANN model to predict the surface roughness in AWJ cutting of Kevlar 49 composite where L_9 orthogonal array was used in Taguchi’s design of experiments as the data for the ANN model [9]. The present work uses a response surface methodology approach to develop the experimental design involving the first order model and the second order model of central composite design. Compared to previous studies, ANN model is used to predict the surface roughness after the paint removal process of automotive component using the waterjet cleaning operation and that is to set the optimal set of parameters to obtain desired roughness.

2 Experimental Work

2.1 Material

The workpiece was prepared by cutting a commercial automotive part from standard plate. The sample was cut down to a small size from the entire component to create a flat workpiece with a surface area of nearly 100×100 mm. This project employed a commercial automobile part (spoiler) manufactured of polypropylene (PP) as the substrate. The plate was commercially painted with a conventional of coat layers. The average surface roughness (R_a) of the painted plate was about $0.112 \mu\text{m}$. The average thickness of the coat layers was determined to be $37.413 \mu\text{m}$. Table 1 shows the tensile mechanical characteristics of paint on the surface of automotive plastics components [13].

Table 1 Mechanical Properties of paint on the surface of automotive plastics component

Paint	Elastic modulus, E (GPa)	Yield strength, $\sigma_{0.2}$ (MPa)	Ultimate strength, σ_b (MPa)
Top coating	1.59–1.76	11.34–15.85	17.43–21.22
Intermediate coating	1.54–1.88	8.99–13.10	10.71–13.75
Primer	0.52–0.77	4.93–7.11	7.83–10.60

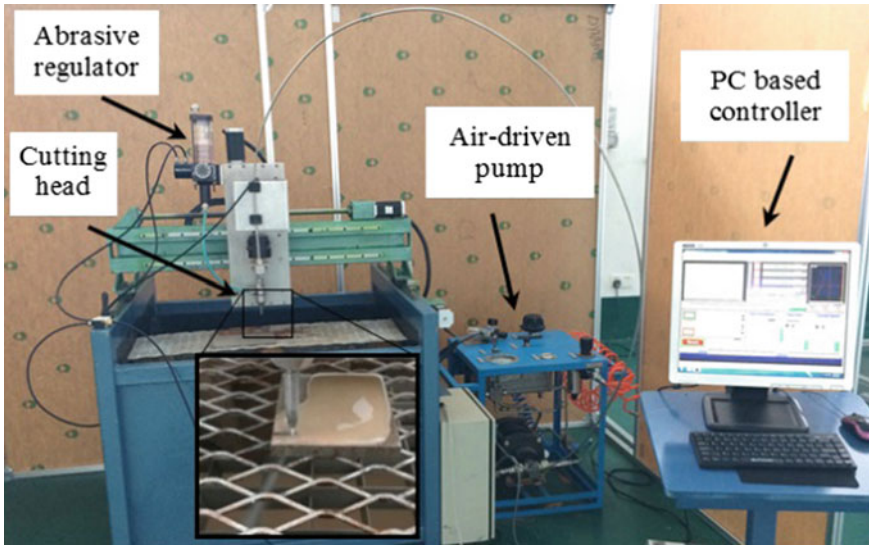


Fig. 1 The CNC waterjet machine

2.2 Equipment

A commercial waterjet cutting machine was used for the entire experiment. It uses an air-driven liquid pump with a capacity of water pressure of up to 200 MPa. A computer numerical control (CNC) system manages the movement of the nozzle in 3-dimensional directions. A ruby orifice with a diameter of 0.127 mm and a tungsten carbide focusing tube with diameters and lengths of 0.76 and 76.2 mm, respectively, were used for all experiments. Figure 1 shows the CNC waterjet machine.

2.3 Experimental Design

The experiment was conducted using a response surface methodology approach. Overall, five parameters were selected namely water pressure (p), traverse rate (u), standoff distance (s), number of passes (n) and lateral feed (f). Initially, a total of 16 experimental runs were conducted based on the first order regression model using a fractional two-level factorial design. Table 2 shows the setting of the input parameters and their resultant roughness from the first model.

Later, a second order model was created using Central Composite Design (CCD) once the substantial parameters were found. The CCD method is a two-level complete factorial (2^k) design that includes a center point and additional randomly chosen experiment trials. A total of additional 13 experiments were conducted as shown in Table 3.

Table 2 Experimental layout of two-level factorial

Run	p (MPa)	u (mm/min)	s (mm)	N (pass)	f (mm)	Ra (μm)
1	34	1000	10	1	0.2	7.647
2	34	1000	20	1	0.6	2.389
3	69	1000	20	3	0.6	6.63
4	34	500	10	3	0.2	7.824
5	34	500	20	3	0.6	7.824
6	69	500	10	3	0.6	6.221
7	69	500	10	1	0.2	5.683
8	69	1000	20	1	0.2	6.98
9	34	500	20	1	0.2	6.433
10	69	500	20	3	0.2	7.464
11	34	1000	10	3	0.6	5.742
12	34	500	10	1	0.6	2.48
13	34	1000	20	3	0.2	7.465
14	69	1000	10	1	0.6	4.476
15	69	1000	10	3	0.2	5.783
16	69	500	20	1	0.6	6.268

Table 3 Experimental layout of central composite design

Run	p (MPa)	u (mm/min)	s (mm)	n (pass)	f (mm)	Ra (μm)
1	35.3	500	10	1	0.7	4.754
2	33.6	500	10	1	0.5	5.504
3	34	500	10	1	0.6	5.983
4	32.8	500	10	1	0.6	2.655
5	34	500	10	1	0.6	5.983
6	34	500	10	1	0.6	5.983
7	36.2	500	10	1	0.6	6.404
8	34	500	10	1	0.4	7.729
9	33.6	500	10	1	0.7	3.259
10	35.3	500	10	1	0.5	8.234
11	34	500	10	1	0.6	5.983
12	34	500	10	1	0.6	5.983
13	34	500	10	1	0.8	2.104

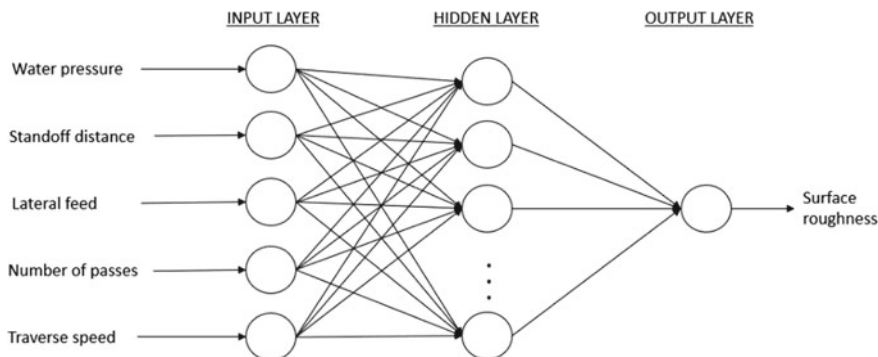


Fig. 2 ANN architecture for the surface roughness

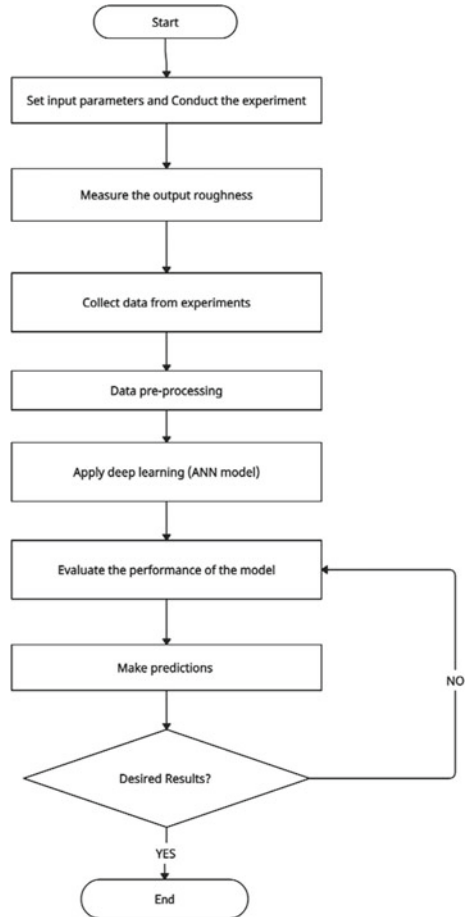
In all experiments, orifice diameter and impact angle were kept constant at 0.127 mm and 90° , respectively. The sample was clamped to the machine table. By applying repeated overlapping water-generated impacts, the nozzle moved in x and y directions to obtain coverage of the treated area of about 70×5 mm. The overlapping (lateral feed) rate was set between 0.2 and 0.6 mm to ensure a sufficiently uniform WJP treatment.

The surface quality of cleaned surface was assessed using a surface roughness instrument (Mahr Marsurf PS1). The average roughness profile of five successive sample lengths was assessed using the average arithmetic roughness parameters (R_a) according to EN ISO 4288 standard.

2.4 ANN Model

In the architecture of the ANN, there are three main layers in the model as shown in Fig. 2. The input layer, which is only one layer, where all input parameters of the system were fed into that layer, followed by the hidden layers which the number of hidden layers is optimized and set depending on the performance of the trained model. Hidden layers are not visible compared to the input and the output layers. Then, the output layer which consists of the response variable [14]. First, the input data were scaled and normalized and that is to avoid slow and unstable learning process of the ANN model. The input from the data was sent to the input layer, where each input neuron is connected to a new neuron in the next layer, and so on, until it reaches the output layer. In each neuron, initially a random weight is assigned. Each neuron has an activation function. Following feed propagation, backpropagation is essential in this study due to its ability to compute gradients in order to reduce the error by effectively updating the weights and bias after each cycle from the output layer and propagating it backwards for each layer to achieve a value closer to the desired output [9, 12]. The gradient of the error function with respect to the weights of the neural

Fig. 3 Flowchart of decision for experimental work and ANN model



network was calculated using chain rule. The Rectified Linear Unit (Relu) activation function is used to train the model, as it computes faster than other common activation function such as sigmoid etc. Where Relu only activates certain neurons at a time [15]. To evaluate the performance of the trained model, only three hidden layers were set to train the model after only 150 epochs and that is to avoid overfitting and high variance (Fig. 3).

3 Result and Discussion

The R_a for the two-level factorial design and central composite design were determined to be between 2.85 and 7.57 μm and between 2.104 and 8.234 μm , respectively as shown in Tables 2 and 3. Both experimental designs were combined with a

total of 29 experimental data to be fed to the ANN model where a 23 experimental data (80%) and 6 experimental data (20%) were used for the model training and testing, respectively. The ANN model used 5 input parameters namely water pressure, traverse speed, number of passes standoff distance and lateral feed whereas the surface roughness was taken as the targeted output.

After training the model, the learning curve from the neural networks is shown in Fig. 4. It can be observed that the learning curve becomes steady around 3.02 after 80 epochs or iterations. The trained ANN model was achieved after 150 epochs due to its repeated computations thus minimizing the computational cost of ANN. Table 4 shows the predicted R_a for the testing data involving 6 experimental runs. The predicted values of R_a are close to the actual values of R_a . Based on the testing data, the mean square error and the absolute error of the trained model were found to be 3.02 and 1.4, respectively. It can be concluded that that the developed ANN model

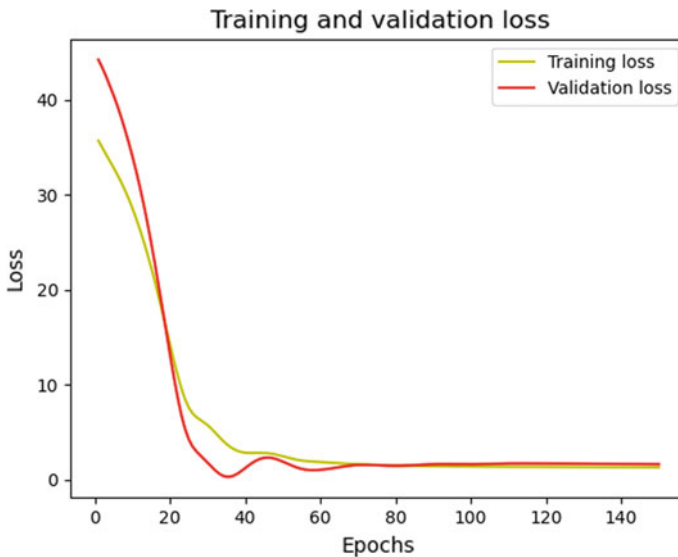


Fig. 4 The training and validation accuracy and loss at each epoch

Table 4 Actual Ra VS predicted Ra

p (MPa)	u (mm/min)	s (mm)	n (pass)	f (mm)	Actual: R_a (μm)	Predicted: R_a (μm)
69	1000	10	1	0.60000	4.476	3.6344
34	500	20	3	0.60000	7.824	5.37536
34	1000	20	1	0.60000	2.389	5.08775
34	1000	10	3	0.60000	5.742	5.03079
34	500	10	1	0.80000	2.104	4.00866
34	1000	20	3	0.20000	7.465	7.59611

is capable of estimating the surface roughness within a reasonable range of error for the waterjet paint removal process of automotive component.

4 Conclusion

The relationship between the cleaning input parameters and surface roughness in waterjet paint removal process was determined by developing an ANN model based on the experiment design according to the response surface methodology approach. A total of 29 experimental data were used from a combined of two-level factorial design (16 experiments) and central composite design (13 experiments) the ANN model. After testing the performance of the model on 6 randomly selected cases from both experimental designs, the ANN model was found to have an average of 80% accuracy and 3.02 mean square error. This summarizes that ANN model can sufficiently estimate surface roughness in waterjet paint removal process with a reasonable error range.

Acknowledgements Authors would like to gratefully acknowledge the financial support from Universiti Malaysia Pahang through RDU182203-2.

References

1. Li X, Wang H, Yu W, Wang L, Wang D, Cheng H, Wang L (2021) Laser paint stripping strategy in engineering application: a systematic review. *Optik* 241:167036. <https://doi.org/10.1016/j.ijleo.2021.167036>
2. Sanmartín P, Cappitelli F, Mitchell R (2014) Current methods of graffiti removal: a review. *Constr Build Mater* 71:363–374. <https://doi.org/10.1016/j.conbuildmat.2014.08.093>
3. Vergès-Belmin V, Wiedemann G, Weber L, Cooper M, Crump D, Gouverne R (2003) A review of health hazards linked to the use of lasers for stone cleaning. *J Cult Herit* 4:33–37. [https://doi.org/10.1016/S1296-2074\(02\)01224-4](https://doi.org/10.1016/S1296-2074(02)01224-4)
4. Folkes J (2009) Waterjet-an innovative tool for manufacturing. *J Mater Process Technol* 209:6181–6189. <https://doi.org/10.1016/j.jmatprotec.2009.05.025>
5. Carvalhão M, Dionísio A (2015) Evaluation of mechanical soft-abrasive blasting and chemical cleaning methods on alkyl-paint graffiti made on calcareous stones. *J Cult Herit* 16:579–590. <https://doi.org/10.1016/j.culher.2014.10.004>
6. Erzurumlu T, Oktem H (2007) Comparison of response surface model with neural network in determining the surface quality of moulded parts. *Mater Des* 28:459–465. <https://doi.org/10.1016/j.matdes.2005.09.004>
7. Risbood KA, Dixit US, Sahasrabudhe AD (2003) Prediction of surface roughness and dimensional deviation by measuring cutting forces and vibrations in turning process. *J Mater Process Technol* 132:203–214. [https://doi.org/10.1016/S0924-0136\(02\)00920-2](https://doi.org/10.1016/S0924-0136(02)00920-2)
8. Chien WT, Chou CY (2001) The predictive model for machinability of 304 stainless steel. *J Mater Process Technol* 118:442–447. [https://doi.org/10.1016/S0924-0136\(01\)00875-5](https://doi.org/10.1016/S0924-0136(01)00875-5)
9. Madara SR, Pillai SR, Pon Selvan MC, Van heirle J (2021) Modelling of surface roughness in abrasive waterjet cutting of Kevlar 49 composite using artificial neural network. *Mater Today Proc* 46:1–8. <https://doi.org/10.1016/j.matpr.2020.02.868>

10. Daoming G, Jie C (2006) ANFIS for high-pressure waterjet cleaning prediction. *Surf Coat Technol* 201(3–4):1629–1634
11. Zain AM, Haron H, Sharif S (2011) Estimation of the minimum machining performance in the abrasive waterjet machining using integrated ANN-SA. *Expert Syst Appl* 38:8316–8326. <https://doi.org/10.1016/j.eswa.2011.01.019>
12. Çaydaş U, Haşçalık A (2008) A study on surface roughness in abrasive waterjet machining process using artificial neural networks and regression analysis method. *J Mater Process Technol* 202:574–582. <https://doi.org/10.1016/j.jmatprotec.2007.10.024>
13. Zhang H, Chen M (2015) Theoretical analysis and experimental study on the coating removal from passenger-vehicle plastics for recycling by using water jet technology. *JOM* 67:2714–2726. <https://doi.org/10.1007/s11837-015-1424-6>
14. Singh B (2021) Predicting airline passengers ' loyalty using artificial neural network theory. *J Air Transp Manag* 94:102080. <https://doi.org/10.1016/j.jairtraman.2021.102080>
15. Sharma S, Sharma S, Anidhya A (2017) Understanding activation functions in neural networks. *Int J Eng Appl Sci Technol* 4:310–316

Light Path Simulation for Optical Switch Based on Digital Electromagnetic Actuators



Thi Thanh Vi Nguyen, Nguyen Duy Minh Phan, Ngo Quoc Huy Tran, Le Anh Doan, and Quang Trung Vo

Abstract The purpose of this paper is to model the light paths of optical switch construction based on four discrete positions of the actuator. Specifically, the electromagnetic digital actuators are used to create the switch because it is comprised of a mobile part that can switch along the two orthogonal directions and between 4 discrete positions. The principle of the digital actuators and their application in optical switch is first presented in the model. Then, an algorithm is proposed to validate the different output signals corresponding to the input signal once the actuator changes its own position. The simulation results of the 2×2 and 5×5 actuators switch with aiming to study and optimize the design before being fabricated, showed the effectiveness of the proposed method related to the configurations of digital actuators.

Keywords Optical switches · Fiber optical switches · Electromagnetic digital actuators · 4 discrete positions

1 Introduction

In order to increase the bit rate and implement new protocols in communication networks, the optical switch will gradually replace the core switch. Some of the typical researches on the manufacture of the optical switch have been published in recent years. Reyne et al. [1] have developed an assembly of several elementary optical switches to generate a network that permits reducing input signals but still guaranteeing the number of output signals. The elementary modules correspond to the optical switch developed by Maekoba et al., [2]. The input and output optical fibers are positioned around the assembly. When a moving mirror is in a high position, the optical signal from the input fiber is reflected back to the output fiber placed in the axis of the activated actuator. Hoffman et al., [3] have designed an optical switch network based on a vertical comb-drive actuator which allows a vertical movement

T. T. V. Nguyen (✉) · N. D. M. Phan · N. Q. H. Tran · L. A. Doan · Q. T. Vo
The University of Danang- University of Technology and Education, 48 Cao Thang,
550 000 Danang, Vietnam
e-mail: nttvi@ute.udn.vn

of up to 300 μm with an anisotropically etched comb-drive. With a high position, an optical signal can reflect through a mirror that is placed on a moving comb. With a bottom position, an optical signal passes through over the mirror. In [4], Jie et al., presented a 8×8 optical switches that uses the torsion beam electrostatic actuators. The 8×8 optical switch development is also proposed in [5]. This switch includes a MEMS-based silicon micro-mirror array and a solenoid-based bi-stable mini - actuator array. Usually, the literature listed above has two main directions: an optical signal can be reflected by a mirror, or an optical signal goes through the mirror. From the point of view of the optical switch, we focused on an optical switch using 4-positions digital actuator. This leads to obtaining a lot of input/output signals compared with other technologies. Furthermore, a low cost due to energy savings during the displacement of the actuator to different positions when it is compared with another approach, typically, the core switch. In literature, there are many studies proposing the physical designs for optical switch based on the digital actuators [6–12]. However, few researchers have addressed the problem of light path simulation in designing the optical switch bases on digital actuator. The contribution of this paper lies in modeling light paths of optical switch construction based on the digital actuator. This simulation allows to us determine the optimal parameters as wavelength and energy of input signal provided for optical switch. The simulation result shows the effectiveness of the proposed model.

This paper is organized as follows: Sect. 2 presents the operating principle of the actuator and the steps to generate a network of optical switches. Section 3 introduces the algorithm for light path simulation and its application for an 2×2 and 5×5 digital actuators optical switch. Section 4 shows the simulation results. Section 5 draws the conclusion and future work.

2 Background

2.1 Working Principle of the Actuator

In this work, a digital actuator with 4 discrete positions was studied [13]. As depicted in Fig. 1, the actuator is comprised of a Mobile Permanent Magnet (MPM) that is located in a square cavity, and 4 Fixed Permanent Magnets (FPMs). The holding effect is established magnetically with FPMs placed at each side of the MPM. Each MPM can move independently along the 2 orthogonal axes (x and y-axis). In fact, it can be at any of the 4 discrete positions thanks to the electromagnetic and the magnetic forces respectively. The Lorentz is used to switch the MPM between the 4 discrete positions. The Lorentz force is applied to the 2 axes due to the interplay between the wire's current and the magnetic flux density of the MPM. This Lorentz force is determined by the left hand rule. For example, in case the position of MPM is at $(-x, +y)$ and the current passed through $-y$ -axis direction of wire, the force acting applied on the wire will be in $-x$ direction because the flux density is in the

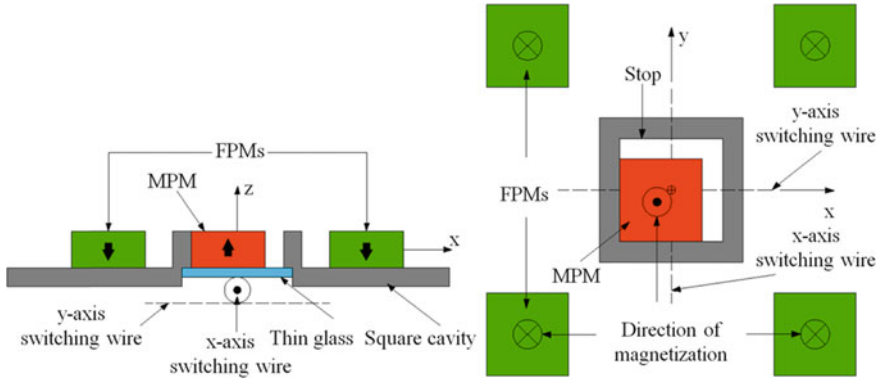


Fig. 1 4 discrete positions digital actuator adapted from [13]

$+z$ direction. When a second current passes through the $+x$ direction of the wire, a Lorentz force is exerted on the wire in the $-y$ direction which leads to an opposite force applied on the $+y$ direction of MPM. Thanks to this force, the holding effect will be improved in minimizing the straightness errors.

2.2 Optical Switch Design

On top of the MPM, a prismatic mirror with 2 reflecting surfaces is attached. One input fiber is located in front of the prismatic mirror in order that the reflection of the input optical signal is directed to one of the 4 output fibers depending on the MPM discrete position. While one collimator is used at the input signal stage in order to produce a collimated transmitted beam, the other 4 are employed at the 4 output fibers to couple the parallel-transmitted beam into the fiber core. Figure 2 shows the 4 states in which the input light signal is switched respectively to one of the 4 outputs [13, 14].

An elementary actuator can be employed to form an $N \times N$ optical switch. The Fig. 3 presents an application of a system including 4 actuators to develop an 2×2 actuators optical switch.

Fig. 2 Optical switch diagram

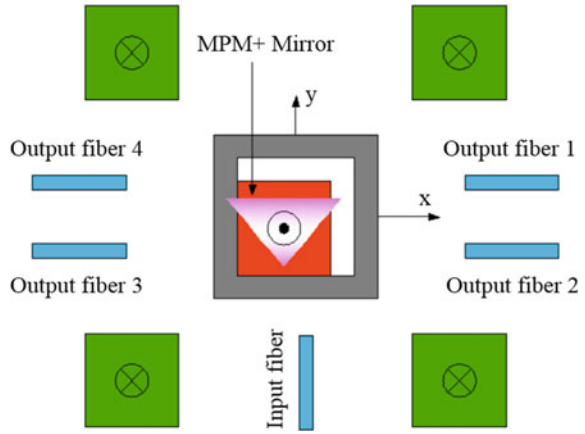
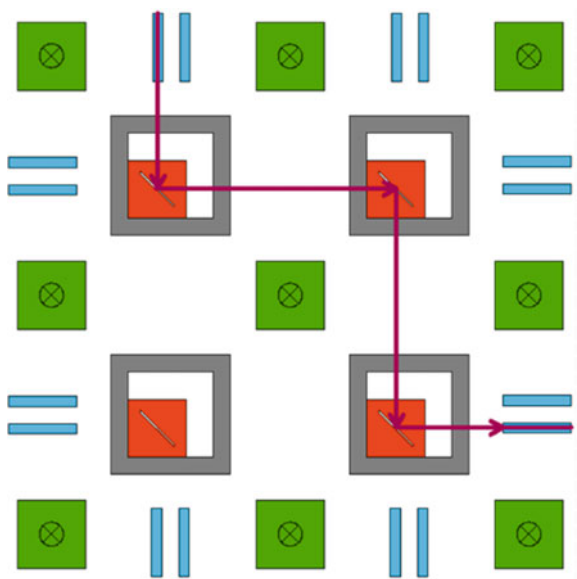


Fig. 3 2×2 digital actuators optical switch



3 Light Paths Simulation of Digital Actuators Optical Switch

In this section, the algorithm for light path simulation of the optical is proposed from the modelling of elementary actuator. This model will be applied for simulating the light paths of an 2×2 and 5×5 digital actuator optical switch. From given input values such as input optical signal, the position of actuator and the position of mirror, this model enables us to determine the output signal as well as the light paths through the optical switch constructed by 4 elementary actuators.

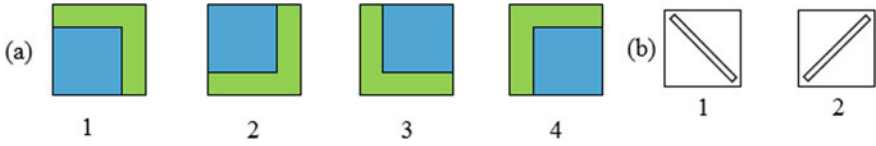


Fig. 4 a 4 discrete positions of actuator b 2 positions of mirror

3.1 Modelling of Elementary Actuator

The 4 discrete positions of the digital actuator are shown on Fig. 4. The digital actuator (in cyan) can be displaced at 4 different corners in the square cavity (in green), which corresponds to the 4 discrete positions of the actuator (Fig. 4a). To facilitate programming, each position of the actuator is labeled from 1 to 4. For example, if the actuator is at the bottom left of the square cavity, it corresponds to position number 1. A vertical mirror with 2 reflective faces is fixed at $+45^\circ$ or -45° corresponding to the position 1 or 2 (Fig. 4b).

It is worth noting that the input/output optical signals related to a reflection of mirror are identified depending on the positions of actuator (1, 2, 3, 4) and the positions of the mirror (1, 2). The two 4×4 matrices $A = (a_{i,j})$ and $B = (b_{i,j})$ are the matrices of input/output optical signals corresponding the position of mirror 1, 2 respectively. Each matrix row is corresponding to each position of actuator. For example, in case of position of actuator 1 and position of mirror 1, the input/output optical signals are at 1, 3, 6, 8. The algorithm shown in Fig. 6 is used to determine the output optical fiber from the set of input data. The input data is represented by $F = [f_{11}, f_{12}, f_{13}]$ with f_{11} , the input optical signal; f_{12} , position of actuator and f_{13} , position of mirror.

3.2 Application for 2×2 and 5×5 Digital Actuators Optical Switch

The algorithm proposed (as depicted in Fig. 6) for an elementary actuator will be applied for a 2×2 and 5×5 digital actuators optical switch. The 2×2 digital actuators optical switch is created by integrating 4 elementary actuators (Fig. 5b). The label of the input/output optical signals is defined by 2 digits, the first digit presents the actuator order, the second presents an input–output optical fibers order. The modelling of light paths for 2×2 optical switch array is developed easily from the algorithm for the elementary actuator modelling when the output of the previous actuator is also the input of the current actuator. The 5×5 digital actuators optical switch is created by integrating 25 elementary actuators. The label of the input/output optical signals for 5×5 digital actuators array is numbered in the same way as the 2×2 one. The numbering for each input/output is given in Fig. 7.

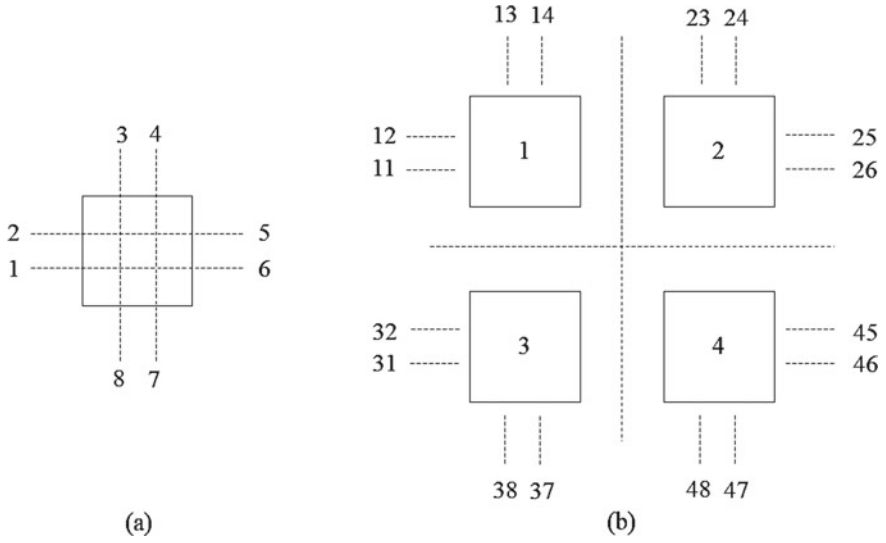


Fig. 5 a Input/output of elementary actuator b Input/output of 2 × 2 actuators

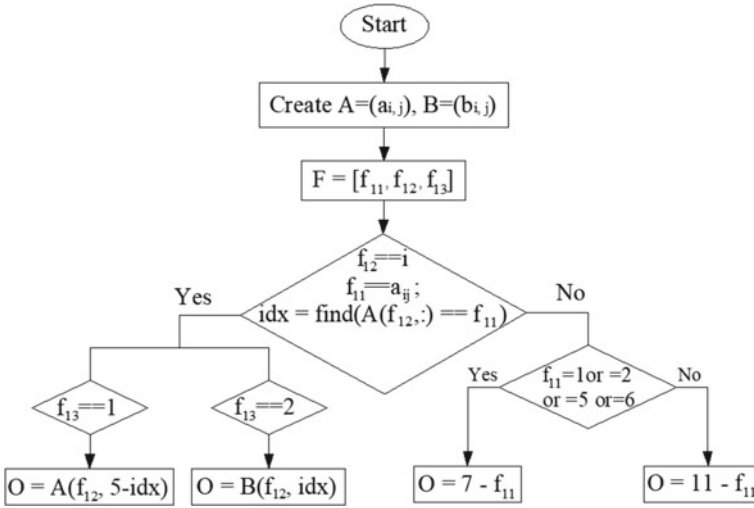
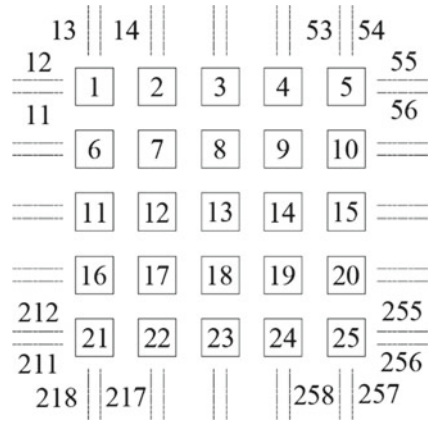


Fig. 6 Algorithm for light path modelling of an elementary actuator

Fig. 7 Input/output of 5×5 actuators



4 Results and Discussion

The light path modelling is shown on the graphic model programmed in Matlab (Fig. 8). The square with a green color is a square cavity that contains a MPM. The square with a blue color is a MPM of the actuator. A red color line and inclined at an angle of $\pm 45^\circ$, it is a mirror. The lines with a black color are optical fibers while the red line are optical signals and optical paths.

For each input signal, the path of the reflected light (red line) through the mirror reaches an appropriate location. In Fig. 8a, the input signal from the optical fiber 3 will be sent to the mirror of digital actuator 2, the mirror of actuator 1, the mirror of actuator 3, the mirror of actuator 4 and finally the output signal is terminated at

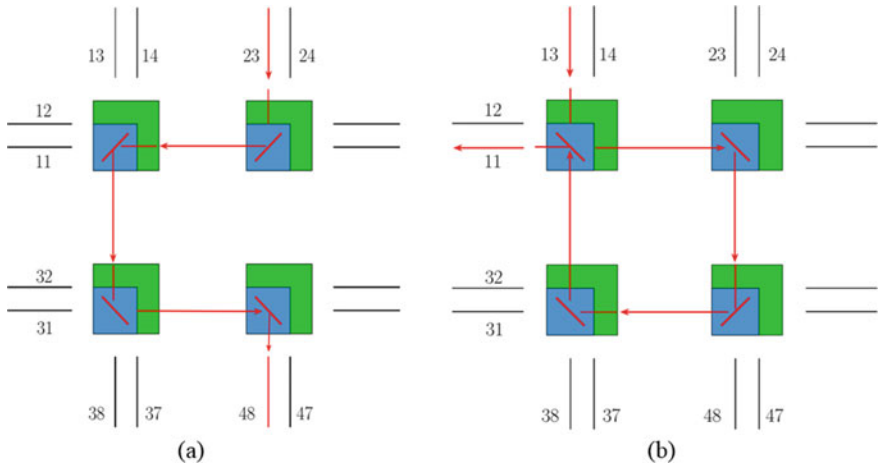
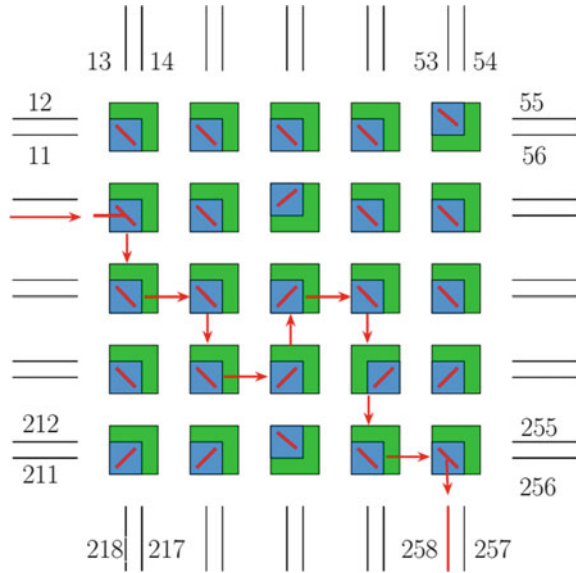


Fig. 8 Modelling of optical path of an 2×2 actuators switch

Fig. 9 Modelling of optical path of an 5×5 actuators switch



the optical fiber 8 of actuator 4. The Fig. 9 shows the light path through the mirror of actuator in the 5×5 actuators. The signal from the optical fiber 6 will be sent to the mirrors of actuator 11, 12, 17, 18, 13, 14, 19, 24 and the output signal is at the optical fiber 8 of actuator 25.

By validating visually the light paths inside the optical switch via the simulation results of two types of 2×2 and 5×5 actuators, all the output signals are predicted correctly by the proposed algorithm when there are input signals going through the actuator switch. Thanks to this model, we can determine the number of mirrors that the light has reflected through and therefore estimate the required wavelength of input signal should be applied to pass through the desired mirrors as well as the energy needed to provide for the optical switch.

5 Conclusion

In this paper, the light path simulation of an optical switch based on the electromagnetic digital actuators has been presented. An algorithm for modelling light path is firstly developed for the elementary actuator and then applied for the 2×2 and 5×5 optical actuators by integrating 4 or 25 elementary actuators respectively. The obtain results have confirmed our algorithm. The simulation of light paths so far have been very promising and allows for studying and optimizing the optical switch design. In the future, this model will be upgraded for modelling an $N \times N$ actuators optical switch and optimizing the switch design to shorten the switching time as well as the energy consumption.

References

1. Reyne G (2002) Electromagnetic actuation for MOEMS, examples, advantages and drawbacks of MAGMAS. *J Magn Magn Mater* 242:1119–1125
2. Maekoba H, Helin P, Reyne G, Bourouina T, Fujita H (2001) Self-aligned vertical mirror and V-grooves applied to an optical-switch: modeling and optimization of bi-stable operation by electromagnetic actuation. *Sens Actu A Phys* 87(3):172–178
3. Hoffmann M, Voges E (2002) Bulk silicon micromachining for MEMS in optical communication systems. *J Micromech Micro Eng* 12:349–360
4. Jia C, Zhou J, Dong W, Chen W (2009) Design and fabrication of silicon-based 8×8 MEMS optical switch array. *Microelectron J* 40(1):83–86
5. Yang YJ, Liao BT, Shih SC, Fan KC (2010) The development of an 8×8 optical switch. *J Chin Inst Eng* 33(1):37–43
6. Cochran KR, Fan L, DeVoe DL (2005) High-power optical microswitch based on direct fiber actuation. *Sens Actu A* 119(2):512–519
7. Chen WC, Lee C, Wu CY, Fang W (2005) A new latched 2×2 optical switch using bi-directional movable electrothermal H-beam actuators. *Sens Act A* 123:563–569
8. Pieri F, Piotto M (2000) A micromachined bistable 1×2 switch for optical fibers. *Microelectron Eng* 53(1–4):561–564
9. Fu S, Ding G, Wang H, Yang Z, Feng J (2007) Design and fabrication of a magnetic bi-stable electromagnetic MEMS relay. *Microelectron J* 38(45):556–563
10. Hajjar HA, Petit L, Lamarque F, Fracasso B (2015) 1×4 optical switch based on digital actuator. *Int J Optomechatron* 9(2):141–150
11. Petit L, Dupont E, Doré E, Lamarque F, Prella C (2015) Design and characterization of a high-precision digital electromagnetic actuator with four discrete positions. *Actu Multidisc Dig Publ Inst* 4(4):217–236
12. Deshmukh A, Petit L, Khan MU, Lamarque F, Prella C (2017) Development of a six positions digital electromagnetic actuator. In: 2017 IEEE international conference on advanced intelligent mechatronics (AIM), pp 975–980
13. Petit L, Al Hajjar H, Prella C, Lamarque F (2016) Design, modeling, and characterization of an optical switch based on four positions digital actuator. *IEEE/ASME Trans Mechatron* 21(3):1518–1527
14. Petit L (2009) Contribution aux techniques d'actionnement numérique: cas d'un système électromagnétique 2D (Doctoral dissertation, Compiègne)

An Application of Charge-Coupled Device (CCD) Tomography System for Gemological Industry - A Review



Fatinah Mohd Rahalim, Juliza Jamaludin, Syarfa Najihah Raisin, Wan Zakiah Wan Ismail, Irneza Ismail, Ruzairi Abdul Rahim, and Yasmin Abdul Wahab

Abstract Charge-Coupled Device (CCD) is a semiconductor chip with a light-sensitive sensor. The CCD has been used in many fields of engineering, including astronomy, medical sciences and processing. CCD is capable to detect light sources and convert this analogue signal into electrical signal. CCD is an integrated circuit that contains a large number of small photo elements with high sensitivity to light energy. The main focus of this research paper is on the review of CCD basic operating principle and construction, CCD characteristic, and the application of CCD in tomography system. The potential use of CCD in the gemological industry is also highlighted in this paper. Gemology is one of the important industries that considered profitable and crucial that deals with precious stones. This industry is in need of standardized grading valuation of gemstones as the current technique is prone to errors. An approach to the standardized grading technique is proposed where CCD tomography is used to detect and analyze the light distribution characteristic in ruby stones.

Keywords Charge-Coupled Device (CCD) · Gemology · Light distribution · Ruby · Tomography

1 Introduction

Since the 1980s, Charge-Coupled Devices (CCDs) have been the most widely used high-performance imaging detector in almost all scientific and industrial imaging

F. M. Rahalim · J. Jamaludin (✉) · S. N. Raisin · W. Z. W. Ismail · I. Ismail
Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia (USIM),
71800 Nilai, Negeri Sembilan, Malaysia
e-mail: juliza@usim.edu.my

R. A. Rahim
School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, UTM
Johor Bahru, 81310 Johor, Malaysia

Y. A. Wahab
Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan
Campus, 26600 Pekan, Pahang, Malaysia

applications [1]. In today's industries, the CCD sensor is a multipurpose electronic component that is in high demand. This type of sensor is unique in its architecture design, since it is made up of thousands of very small sensors that are very responsive to light sources [2]. It is extensively used in optical imaging, target tracking, and other products due to its wide spectral response, large dynamic range, high sensitivity, low power consumption, shock resistance and anti-electromagnetic interference capabilities. With this special characteristic, CCD is used as core imager components in resource satellite, ocean satellite and also meteorological satellite [3]. CCDs are also utilized in medical applications, particularly in X-ray tomography systems. X-ray tomography systems provide extremely high-resolution image reconstructions, which are critical in the medical industry since a clear and correct image output is important. Nowadays, CCD tomography systems are widely used in process industries for plant monitoring purposes because this system provide a non-intrusive and non-invasive inspection technique. CCD tomography system is also known as a hard-field sensor system because CCD sensor only depends on the change of the light attenuation or absorption. Aside from that, CCD sensors are resistant to electrical noise and interference, have a high resolution, and operate at a high speed. Many studies have been conducted on the analysis and monitoring of multiphase flow, solid contamination, sewerage blockage, and object measurement using CCD tomography systems.

The primary goal of this research paper is to provide an overview of CCD basic operating principles and construction, CCD performance criteria, and CCD applications in tomography systems. This paper also discusses the potential applications of CCD tomography in the gemological industry. The scientific study of gems is known as gemology [4]. Gemstones are natural inorganic minerals which have been mined and valued since prehistoric times as precious stones in jewelry or adornment [5]. The gemology sector is tremendously profitable, but it is far more complex than many other industries, such as iron, silver, or gold, because the quality of the gems, particularly the high-demand gems such as ruby, diamond, and sapphire, do not depend solely on carat weight or purity. With many artistic, subjective, and cultural factors in grading techniques, the relevance is dependent on a much more complicated quality [6].

2 Basic Operating System of CCD

A CCD is a type of image detector that comprises of an array of pixels that generate potential wells from applied clock signals in order to store and transfer charge packets. The charge packets for most CCDs consist of electrons created by the photoelectric incident photons or the internal dark signal. These pixels are defined by gate structures on the silicon surface. A time variable voltage sequence is supplied to these gates in a certain pattern, which moves the charge physically to a charge-to-voltage converter output amplifier. The output sequence of voltages is converted into a two-dimensional (2D) digital image by external electronics (typically a computer) [1].

Although the CCD's early promise as a memory element has faded, its superior ability to detect light has made it the industry's standard image sensor technology. Its light sensitivity was immediately exploited for imaging applications, resulting in a tremendous revolution, particularly in the field of astronomy [7]. Linear CCDs, in which one row of pixels is transferred to the output, and matrix CCDs, in which the bottom row is always entirely pushed out when all columns are shifted down by one pixel, are the two types of CCD sensors [8].

The underlying principle is based on the movement of electrical charge. A Metal Oxide Semiconductor (MOS) capacitor serves as the sensor for this component. When light strikes the CCD's surface, photon charges attract an electron charge, causing it to depart from its covalence band. The number of electrons created is relative to the quantity of photons reaching the detector's glass. The electrons will then be moved to the next sensor until the last sensor is reached. The data from the last sensor will be sent to the computer, which will process it and create an image reconstruction [2]. The number of electrons produced was proportional to the number of incident photons [2].

There are many types of CCDs invented where every type of CCD has their own function that can suit in certain application. For process industries application, it is usually low cost and comprises two types of sensor: color CCD and monochromatic CCD. Color sensors are typically found in video and photographic cameras, whereas monochromatic sensors are found in facsimile machines and scanners. In 1991, researchers discovered that CCD Linear sensor with color scheme had drawbacks. Because the CCD needs to scan three times for three different colors [2], this color CCD Linear Sensor requires a large memory for output signals.

There are several types of monochromatic CCD sensors, each with its own function that can be used in a specific function. For example, CCD ILX551 and CCD ILX551A are monochromatic CCD sensors. The main features of the CCD sensors ILX551A and ILX511, as shown in Fig. 1, differ slightly. CCD ILX551A has a smaller pixel than CCD ILX511 with $14 \times 14 \mu\text{m}$ area, although having the same number of effective pixels. Moreover, both CCDs have a high efficiency, but the CCD ILX551A has a higher maximum frequency of 5 MHz than the CCD ILX511, which has a maximum frequency of just 2 MHz. As a result, the CCD ILX551A can send and receive more data in less time. Above all, the CCD ILX511 is designed to be used for bar



Fig. 1 CCD sensor ILX551A (left) [11] and ILX511 (right) [12]

code point-of-sale (POS) hand scanner and optical measuring equipment where these devices do not require higher accuracy [9]. While for CCD ILX551A is designed to be used for facsimile, image scanner and optical character recognition (OCR) which obviously need greater accuracy and efficiency [10].

Researchers might explore a variety of criteria in order to improve CCD performance. In general, when it comes to CCD performance, there are seven criteria to consider. Quantum Efficiency (QE), signal noise per ratio, spectrum sensitivity, transfer efficiency, spatial resolution, blooming, and the dark current are the criteria. CCD is considered to have a high fraction of Quantum Efficiency (QE), ranging from 40 to 90% [13], when compared to other optoelectronic sensors. CCD has a typical wavelength range of 400 to 1000 nm. This range encompasses the visible spectrum as well as a major part of the infrared spectrum [14]. In today's market, CCD sensors come in a variety of sizes, each with a different level of spatial resolution. Blooming can occur when CCD is exposed to high-intensity light sources or intense illumination. Another factor that influences CCD performance is dark current. Even if the sensor is positioned in a dark area, dark current will occur when the signal is recognized [13].

Based on previous literature review, this proposed fundamental research is capable to provide a standardized quantitative grading valuation of gemstones via CCD tomography approach. Light reflection from the gemstone will strikes onto the CCD surfaces. This light source will be converted into an electrical signal which is in voltage output value. The CCD voltage output represents the light intensity value of ruby stone. CCD pixel will collect photon charges received from the light. The more photon charges received by CCD, the higher the transparency of gemstone [15].

3 Applications of CCD

Smith and Boyle, who were the inventor of this sensor, used CCD in solid-state cameras at an early point in its development from 1970 to 1975 [2, 16, 17]. These CCDs were also used in broadcast television [2]. This sensor was first used in the field of astronomy in 1983. Then, several large corporations began to manufacture this type of sensor to meet the needs of astronomy technology [3]. CCD cameras used in astronomy are expensive. To obtain an image of outer space, astronomical engineers typically employed area arrays of CCDs. The greatest CCD size recently recorded is installed on a 6" wafer size [2]. The application determines the essential principle of detecting an object using a CCD. In the area of astronomy, for example, the area array CCD is used to identify stars, planets, or meteors from millions of miles away via light reflection. So, the actual principle is to detect a bright light spot while the background of the unspotted area is dark. This is how it uses CCD to detect the presence of stars [2]. Then, in 1991, CCD became popular in the photographic world. Many digital cameras using CCD technology that create high-resolution photos have been released. A CCD used in a digital camera is less expensive than one used in a telescope. Furthermore, its pixel properties and sizes differ since they are determined

by consumer preferences [2]. CCD guarantees a more compact, high-quality image, and perhaps even more efficient camera [16, 17].

Modern space X-ray astronomy detection techniques have advanced significantly in the direction of low noise, low power, and huge scale. Because it outperforms conventional high-energy particle detectors in terms of both energy and spatial resolutions, the X-ray charge-coupled device (CCD) has a wide range of applications in the X-ray astronomy field [18]. CCD technology is also employed in X-ray Computed Tomography (CT). Since it can provide morphological and physical information about the interior structure of the investigated sample, X-ray CT is one of the most powerful non-destructive testing techniques for full-volume inspection of an object. CCDs are combined with other components to achieve digital radiography for x-ray, such as a CCD camera coupled with a taper for micro-CT, a CCD camera coupled with a fan of coherent optical fiber ribbons (multi-slice linear detector) and a scintillating screen seen by a CCD camera (cone-beam CT) [19]. The recent advancement of CCD based optical Computed Tomography (optical CT) scanners has enabled rapid and low-cost three-dimensional (3D) gel dosimetry for current radiation applications. CCD based optical CT can acquire the entire plane of data at each step and quickly provide a complete 3D dosage distribution [20]. One of the scanners developed is the CCD lasers range scanner (CCD-LRS) method, which is capable of recording both geometric and color information, thus improves scanning and tracking accuracy [21]. CCD scanners are widely used in medical field such as the MRI scanner [22].

The most common application of CCD is as an optical displacement sensor [17, 23], surface detection sensor [25], thickness detection sensor [26] and object detection system [14]. Because CCD cameras have excellent precision for 2D image measurement and lasers have tremendous precision in the axial direction, Z. Fei et al. [13] concluded that the best combination is laser and CCD cameras. According to Yang Ni et al. [13], the combination of CCD and laser has facilitated the development of thickness detecting tools with high sensitivity, exact accuracy, and reading distance stability. These circumstances demonstrated that the best transceivers for opaque object measurement technologies are CCD and laser diode. In comparison to solid items, light may penetrate more through transparent objects. Because light intensities obtained by CCD may not differ considerably, it may be difficult for CCD to detect transparent objects. J. Jamaludin et al. [27, 28] developed the OPT system, which detects moving air bubbles in crystal-clear water. In crystal-clear water, this CCD tomography system can reconstruct a cross-section image of the moving item with multiple low opacities.

4 Applications of CCD in Gemological Field

The science of gems is known as gemology. People involved in gem analysis, from amateurs to professional gemmologists, such as archaeologists, art historians, conservators, mineralogists, and gem merchants, have a difficult challenge. They must

not only comprehend the fundamentals of numerous sciences (mineralogy, crystallography, geology, chemistry, physics, and sometimes biology), but they must also consider economic factors [29]. The study of gems starts from the analysis of their chemical composition, determining whether they are natural or “artificial” (i.e., imitation, or synthetic), screening for enhancing treatments, grading and sometimes determining their geographic origin. In addition, all the information should be gathered by utilizing non-destructive and non-invasive procedures. When such diamonds are placed or inserted in jewels or artworks, their analysis becomes more complicated. Where in collections and museums, the gems and jewels to be tested are conserved, due to their great value, it is frequently not possible to remove them and to prevent any damage, such gems will therefore require in situ identification techniques. As a result, it is critical that the gemmologist handling the stones be professionally trained in the various ways of identification while considering all of the aforementioned scenarios and duties [29].

For an example, ruby depends on the quality and value of the combination; carat, cut, color and clarity. Excluding carat, the other parameters still require the presence of experienced gemmologists [30] in deciding on the value and quality of the ruby stone. This is because these three criteria are incredibly difficult to distinguish by the human vision [31]. Many gemological tools are used to distinguish or measure these characteristics [32]. Haüy (1817) and his colleagues began the development of gemology as a contemporary science. Many gemological instruments like the refractometer and polarizing filters were invented throughout the nineteenth century (at the time, made of gem tourmaline). These two equipment examples remain critical for gemological identification [33].

The International Gem Society claimed that the necessary instruments in the laboratory for gemology include magnifying the gem using a loupe and a microscope, a refractometer for measuring the refraction index, birefringence and the optical sign for gems, a dichroscope, a polariscope and a spectroscope. These instruments are commonly used in conjunction with other tools, such as the eyes, the diamond detector, and hardness sets [32]. Many researchers came out with more advance tools in detecting the grading value of gemstones.

In previous studies, the relevance of ruby stone being graded based on its clarity was emphasized. By capturing them in a fixed setting, the clarity of the various ruby stones was assessed and then each one of them was analyzed based on the acquired image. The authors of in other publications investigated the color of several rubies and analyzed the composition and the components of stones [34]. As the earlier approaches are prone to human error, further advancing techniques such as Raman spectroscopy and photoluminescence (PL) are introduced to identify the color variation of gemstones with better and more dependable findings [35]. The approach of grading is also proposed by Tariwong et. al using the X-ray micro-CT scan in which the interior part of the rubies is analyzed. Synthetic and treated gemstones are often seen in the trading markets as imitating high-quality gemstones, which cause scams and big losses to consumers. However, identifying instruments are generally huge, heavy and costly, while detecting tools require experience and operational expertise in order to prevent misleading results [36]. These tools depend

mostly on human vision and are prone to inaccuracy [37]. Moreover, these established measuring instruments cannot perform standardize ruby stone measurements, since the instruments cannot distinguish between different degrees of optical properties (crystal clearness and color).

4.1 Research Background

In this paper, an idea of designing the standardized system for grading the gems specifically on the ruby stone is proposed. The study on the light characteristic of ruby stone is conducted and the concept of the CCD is applied in this research. The light of ruby stone will hit onto the CCD sensor when the laser is transmitted and passes through that ruby stone. The CCD sensor analyses the received light and converts it into to voltage. The system will be designed with the Arduino Nano Microcontroller. The final voltage value will be shown on the computer. The whole process is illustrated in the block diagram in Fig. 2.

The light is moving across a transparent particle with three consequences: absorption, reflectance and dispersion (neglected due to its complex mathematical model and the fact that the particle size of interest is much greater than the wavelength of the incident light) [38]. Energy loss occurs when light passes across an interface of light reflection. The incidence angle of the incident ray increases, reflecting a bigger proportion of light. This reflection reduces the amount of light that is emitted by the particle.

Light attenuation is a process in which light is absorbed and subsequently transformed into energy when travelling through a substance. The output light intensity

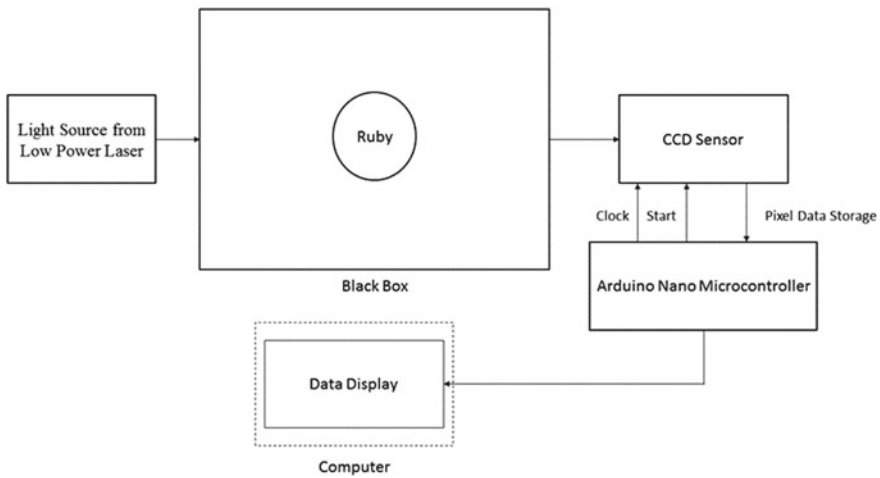


Fig. 2 The overall block diagram of the CCD system

Table 1 CCD voltage output and laser light intensity in off and on mode

Condition of laser	CCD voltage output (V)	Light intensity
Off	4.7419	0
On	1.8142	1

Table 2 CCD voltage output and laser light intensity in different situation of light propagation

Situation of LIGHT PROPAGATION	Light intensity	CCD voltage Output (V)
Air/Ruby	0.8072	2.3787

is attenuated exponentially by the object density along the optical path according to the law of Beer Lambert.

A linear CCD image sensor senses the shadow that falls onto the sensor if a collimated light is sent straight to a CCD linear image sensor. In line with the amount of light in the CCD sensor, the light intensity is transformed into voltage. When there is no object in the system, the saturation voltage is achieved (V_{sat}). In any situation, the reflection on the front of the picture sensor is overlooked [8]. Table 1 shows the theoretical value of CCD voltage output in the laser condition off and on and their respective light intensity. The CCD voltage output is the reference value from the previous research, whereby a clear water pipeline is used to transmit the light [27].

The laser intensity is proportional directly to the output of the CCD voltage. The following equation interprets the relationship between the light intensity and the CCD voltage output.

$$V = -4.5497I + 4.7419 \quad (1)$$

The following table illustrates the theoretical value of the light intensity when the light enters the ruby by the air and the theoretical value of CCD voltage generated. According to Table 2, when a system comprises light intensity and a CCD voltage without a single object, it is demonstrated that the voltage output is higher. The repeated absorption and reflectance process of the laser light results in an increased voltage output [27].

5 Conclusion

CCD is proven as one of the best imaging detectors that have been used widely in various industries with its ability to detect light and convert it into useful data for image reconstruction. With this important qualification, CCD can detect the internal activity of a system non-intrusively and non-inclusively that will not disturb the internal environment of a system [39]. Gemology industry is a crucial industry that deals with high price and valuable gems and needs a standardized grading technique to determine the gem grades without any error. A research on the grading of ruby

gemstone using CCD is proposed based on the characteristic of light distribution in the ruby stone. Based on the theoretical value obtained from the calculation, the voltage value is greater when there is an object in the system compared to when object is absence. Hence, the CCD can possibly be a good choice in building a standardized grading system for gemstones, especially for ruby, based on the light distribution characteristic.

Acknowledgements The authors would like to thank to Universiti Sains Islam Malaysia and ADS research group for their cooperation in this research paper. The research is supported by the Malaysian Ministry of Higher Education Fundamental Research Grant Nos. (FRGS/1/2020/WAB07/USIM/02/1) & (USIM/FRGS/FKAB/KPT/53020).

References

1. Lesser M (2020) Charge-Coupled Device (CCD) image sensors. In: High performance silicon imaging: fundamentals and applications of CMOS and CCD sensors, pp 75–93
2. Jamaludin J, Rahim RA, Rahim HA, Rahiman HF, Muji SZM, Fadzil NSM, Ling LP, Jumaah J, Ahmad A, Ayob NMN, Hong LE (2015) Introducing an application of a charged coupled device (CCD) in an optical tomography system. *Jurnal Teknologi* 73(3):97–102
3. Yang Z, Peng B, Gao L (2016) Research on the design of mechanical shock test fixture of ceramic package for packaging CCD. In: 2016 17th international conference on electronic packaging technology (ICEPT)
4. Clark D (2021) An introduction to gemology. International Gem Society. <https://www.gemsociety.org/article/an-introduction-to-gemology/>, Accessed 24 June 2021
5. Cartier LE (2019) Gemstones and sustainable development: perspectives and trends in mining, processing and trade of precious stones. *Extract Ind Soc* 6(4):1013–1016
6. Moses TM, Johnson ML, Green B, Blodgett T, Cino K, Geurts RH, Gilbertson AM, Hemphill TS, King JM, Kornylak L, Reinitz IM, Shigley JE (2004) A foundation for grading the overall cut quality of round brilliant cut diamonds. *Gems Gemol* 40(3):202–228
7. Ramli N, Ibrahim MN, Idroas M, Harun FC, Mahmood NH (2011) LabVIEW based driver for charge-coupled device linear image sensor. *Jurnal Teknologi*, 129–140
8. Dominec F (2010) Design and construction of a digital CCD spectrometer. Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Physical Electronic
9. ILX511. Sony Corporation
10. ILX551A. Sony Corporation
11. Manufacturer: SONY MPN: ILX551A Encapsulation: CDIP-22,2048-pixel CCD Linear Sensor. AFULLSCO. https://www.afullscout.com/index.php?main_page=product_info&products_id=277896, Accessed 25 June 2021
12. SONY ILX511 - Enhanced Scientifics. ENSCIX. www.enscix.com, Accessed 25 June 2021
13. Buil C (1991) Principle and Performance of CCD. In *CCD Astronomy: Construction and Use of an Astronomical CCD Camera*. Willmann Bell Inc, Virginia, pp 26–28
14. Jay M (2011) What Wavelength Goes with Colour? National Aeronautic and Space Administration. http://scienceedu.larc.nasa.gov/EDDOCS/Wavelengths_for_Colors.html, Accessed 24 Mar 2014
15. Gao J, Zhang Z, Yao R, Sun J, Zhang Y (2011) A robust smear removal method for interframe charged coupled device star images. *Nat Comput* 3:1805–1808
16. Zhou A, Guo J, Shao W (2011) Automated detection of surface defects on sphere parts using laser and CCD measurement. In: IECON 2011–37th annual conference on IEEE industrial electronics society. 2666–2671

17. Ni Y, Tian WY, Tao LJ, Huan LH (2009) Research on thickness measurement of transparent object based on CCD
18. Mehta S, Patel A, Mehta J (2015) CCD or CMOS Image sensor for photography. In: 2015 international conference on communications and signal processing (ICCSPP)
19. Lu B, Huo J, Chen Y, Li B, Liu H, Luo J, Zhou Y (2018) XCR4C: a rad-hard full-function CDS ASIC for X-ray CCD applications. In: 2018 IEEE nuclear science symposium and medical imaging conference proceedings (NSS/MIC)
20. Morigi MP, Casali F, Bettuzzi M, Brancaccio R, D'Errico V (2010) Application of X-ray computed tomography to cultural heritage diagnostics. *Appl Phys A* 100(3):653–661
21. Chang Y-J (2015) Use of a speckle reduction technique to improve the reconstruction image quality of CCD-based optical computed tomography scanner. *Nucl Instrum Methods Phys Res, Sect A* 784:585–589
22. Pheiffer TS, Simpson AL, Lennon B, Thompson RC, Miga MI (2012) Design and evaluation of an optically-tracked single-CCD laser range scanner. *Med Phys* 39(2):636–642
23. Zhou A, Guo J, Shao W (2011) Automated detection of surface defects on sphere parts using laser and CCD measurement. In: IECON 2011–37th annual conference on IEEE industrial electronics society, 2666–2671
24. Jamaludin J, Rahim RA, Rahim HA, Rahiman HF, Muji SZM, Jumaah MF, Fadzil NSM, Ahmad N, Sahlan S, Ahmad A, Yunus Y, Abas KH (2015) Analysis on the performance of led and laser diode with charge coupled device (CCD) linear sensor measuring diameter of object. *Jurnal Teknologi* 77(17):1–7
25. Fei Z, Guo J, Wang J, Luo G (2010) The application of laser and CCD compound measuring method on 3D object detection. In: Proceeding of the international conference on mechatronic and automation, Xian, China, pp 1199–1202
26. Li FM, Nixon O, Nathan A (2004) Degradation behaviour and damage mechanism of CCD image sensor with deep UV laser radiation. *IEEE Trans Electron Dev* 51(12):2229–2236
27. Jamaludin J, Rahim RA, Rahim HA, Fazalul Rahiman MH, Mohd Muji SZ, Rohani JM (2016) Charge coupled device based on optical tomography system in detecting air bubbles in crystal clear water. *Flow Meas Instrum* 50:13–25
28. Jamaludin J, Abdul Rahim R, Abdul Rahim H, Fazalul Rahiman MH, Mohd Rohani J (2017) Charge-coupled device based on optical tomography system for monitoring two-phase flow. *Electron Lett* 53(5):331–333. <https://doi.org/10.1049/el.2016.3084>
29. Chen S, Guan M, Lian H-J, Ma L-J, Shang J-K, He S, Zhang J-W (2014) Crossed cerebellar diaschisis detected by arterial spin-labeled perfusion magnetic resonance imaging in subacute ischemic stroke. *J Stroke Cerebrovasc Dis* 23(9):2378–2383
30. Karampelas S, Kiefert L, Bersani D, Vandenabeele P (2020) *Gems and Gemmology*. Springer Nature Switzerland AG
31. Tariwong Y, Chantima N, Rajaramakrishna R, Kim HJ, Kaewkhao J (2020) X-ray induced luminescence, optical, compositional and structural investigations of natural and imitation rubies: Identification technique. *Radiat Phys Chem* 177:109089
32. Intarasiri S, Bootkul D, Tippawan U, Songsiriritthigul P (2016) Color improvement of rubies by ion beam technique. *Surf Coat Technol* 306:205–210
33. Recommended Gemology Tools and Instruments - Gem Society. <https://www.gemsociety.org/article/tools-for-gemology/>, Accessed 4 Jan 2021
34. Fritsch E, Rondeau B (2009) Gemology: the developing science of gems. *Elements* 5(3):147–152
35. Aung HL, Zin TT (2017) Nanyaseik ruby of phakant township, Kachin state, northern Myanmar. In: 2017 IEEE 6th global conference on consumer electronics (GCCE)
36. Laurs BM, Renfro ND, Williams C, Williams B, Falster AU, William 'Skip' B S, Hart A (2017) Gem notes. *J Gemmol* 35(6):468
37. Liao KW, Lu KJ, Luo RC, Yeh JA (2016) Portable dielectric tunable forensic lens design for jadeite analysis. In: 2016 international conference on optical MEMS and nanophotonics (OMN)

38. Colored Stone Grading • A Heretic's Guide. <http://www.lotusgemology.com/index.php/library/articles/343-colored-stone-grading-a-heretic-s-guide>, Accessed 25 June 2021
39. Jamaludin J, Rahim RA, Rahiman MHF, Rohani JM (2017) Analysis on the effect of sensor views in image reconstruction produced by optical tomography system using charge-coupled device. *IEEE Trans Image Process* 27(4):1689–1696

Prediction of Abrasive Waterjet Machining of Sheet Metals Using Artificial Neural Network



Nur Khadijah Mazlan, Nazrin Mokhtar, M. A. Gebremariam,
and Azmir Azhari

Abstract High pressure waterjet technology has received a wider acceptance for various applications involving machining, cleaning, surface treatment and material cutting. Machining of soft and thin materials with acceptable cutting quality requires a relatively low waterjet pump capacity typically below 150 MPa. The present study attempts to predict the surface roughness during the waterjet machining process for a successful cutting of sheet metals using low pressure. Artificial neural network model was used as the method for prediction. Taguchi method with L_{36} orthogonal array was employed to develop the experimental design. A back-propagation algorithm used in the ANN model has successfully predicted the surface roughness with the mean squared error to be below 10%. This summarizes that ANN model can sufficiently estimate surface roughness in the abrasive waterjet machining of sheet metals with a reasonable error range.

Keywords Waterjet cutting · Abrasive waterjet · Surface roughness · Artificial neural network · Sheet metal

1 Introduction

High pressure waterjet technology is a relatively new process that can be used for machining, material removal, cleaning and surface treatment of various materials [1]. Many studies have been reported to explore the influence of the major abrasive waterjet (AWJ) machining process parameters on the cutting quality. In the study, AWJ is clearly versatile as it can cut through a wide range of material, environmentally friendly as there are no harmful fumes are produced, capable of thin material cutting and lastly no thermal distortion on the workpiece [1]. Generally, the Abrasive

N. K. Mazlan (✉) · N. Mokhtar · A. Azhari
Faculty of Manufacturing Engineering and Mechatronics Engineering Technology, Universiti
Malaysia Pahang, 26600 Pekan, Malaysia
e-mail: nurkhadijah1997@gmail.com

M. A. Gebremariam
College of Engineering, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia

waterjet machining technique can produce complicated parts with excellent precision and accuracy [2, 3]. Machining of sheet metals requires a relatively low waterjet pump capacity usually less than 150 MPa [4, 5]. This pressure range is sufficient to machine soft and thin materials with acceptable cutting quality [5]. In previous studies involving low waterjet pressure application, it has shown the feasibility to drill dolomite rock up to a depth of 70 cm at a pressure of 35 MPa [6]. Furthermore, it was found that low-pressure pre-mixed abrasive waterjet machining process can also successfully machine composite materials [7]. Therefore, the application of AWJ machining at low-pressure is worth to explore especially for thin sheet metals due to its wide industrial usage. Furthermore, it is also crucial to predict the waterjet machining process for a successful cutting of sheet metals using low pressure in industry.

Artificial neural network (ANN) is a type of computing system by learning from previous experimental results to predict future results. It consists of various artificial neurons that operate like synapses in the biological brain, transmitting messages to the other linked neurons [8]. The input of one node is derived from the output of another node thus allowing data flows across the network of nodes and layers [8]. Because of its ability to simulate such nonlinear and complex relationships, ANN was found to be an excellent tool for expressing the link between the parameters and the surface roughness. As a result, using ANN in this study is a sufficient strategy for predicting surface roughness after cutting. Many studies have shown that ANN can be used to forecast the surface finish in a variety of machining techniques [9–12], but there is little literature about their application in the abrasive waterjet machining of sheet metals. Furthermore, as the number of machining factors increases, applicable experimental procedures require many trials. Çaydaş & Haşçalık [13] employed an ANN approach to develop a prediction model during AWJ machining process of aluminium alloy 7075-T6 utilizing a Taguchi design of experiment with L_{27} orthogonal array. Madara et al. [8] also predicted the surface roughness during AWJ cutting of Kevlar 49 composite using an ANN model with experimental data from Taguchi's L_9 orthogonal array. Furthermore, Pappas et al. [14] performed a prediction model of surface quality features in AWJ machining of steel sheets using a hybrid modelling technique based on the Taguchi method and ANN. In all previous studies, it was found that the final prediction models using the ANN approach were able to verify the experimental results with acceptable accuracy of an error below 10%. Hence, the ANN approach can be conveniently applied to make a prediction in AWJ machining process of other materials. The present work uses an ANN model to predict the surface roughness during AWJ machining of sheet metals at low cutting pressure by utilizing experimental data from Taguchi's L_{36} orthogonal array.

2 Experimental Work

2.1 Equipment

A commercial waterjet cutting machine was used for the entire experiment. It uses an air-driven liquid pump with a capacity of water pressure of up to 200 MPa. A computer numerical control (CNC) system manages the movement of the nozzle in 3-dimensional directions. A ruby orifice with a diameter of 0.127 mm and a tungsten carbide focusing tube with diameters and lengths of 0.76 and 76.2 mm, respectively, were used for all experiments.

2.2 Experimental Design

The experiment was conducted using Taguchi method. Overall, four machining parameters were selected namely water pressure (p), traverse rate (u), standoff distance (s) and abrasive flowrate (g). Additionally, the type of material is also selected as another control factor in the Taguchi's design of experiments. With a total of five control factors, consequently, a L_{36} orthogonal array with 36 rows which corresponds to the number of experiments was chosen. Table 1 shows the experimental layout using the L_{36} orthogonal array and their measured surface roughness (R_a) values.

The workpiece was prepared by cutting commercially available sheet metals of stainless steel 304 and T2 pure copper as received. Both sheet metals have a thickness of 1 mm. In all experiments, orifice diameter and impact angle were kept constant at 0.127 mm and 90° , respectively. The sample was clamped to the machine's cutting table. The machining parameters were set to the intended levels according to the L_{36} orthogonal array for each experimental run. Once all the parameters were set to their respective levels, then the test samples were through cut with a slot length of 30 mm.

The surface quality of cleaned surface was assessed using a surface roughness instrument (Mahr Marsurf PS1). The average roughness profile of five successive sample lengths was assessed using the average arithmetic roughness parameters (R_a) according to EN ISO 4288 standard.

2.3 ANN Model

In the architecture of the ANN, there are three main layers in the model as shown in Fig. 1. The input layer where all input parameters of the system were fed into that layer, followed by the hidden layers which were not visible compared to the input and the hidden layers. Then, it went to the output layer which consists of the response variable [15]. The input from the data was sent to the input layer, where each input

Table 1 Experimental design using L₃₆ orthogonal array

Experiment no.	Material	p (MPa)	u (mm/min)	s (mm)	g (g/min)	R_a (μm)
1	Stainless Steel	69	5	3	9	3.087
2	Stainless Steel	103	10	6	36	3.676
3	Stainless Steel	138	15	9	45	4.665
4	Stainless Steel	69	5	3	9	3.056
5	Stainless Steel	103	10	6	36	3.593
6	Stainless Steel	138	15	9	45	4.745
7	Stainless Steel	69	5	6	45	2.614
8	Stainless Steel	103	10	9	9	3.123
9	Stainless Steel	138	15	3	36	4.419
10	Stainless Steel	69	5	9	36	2.754
11	Stainless Steel	103	10	3	45	3.826
12	Stainless Steel	138	15	6	9	4.820
13	Stainless Steel	69	10	9	9	3.462
14	Stainless Steel	103	15	3	36	3.189
15	Stainless Steel	138	5	6	45	4.843
16	Stainless Steel	69	10	9	36	2.506
17	Stainless Steel	103	15	3	45	3.519
18	Stainless Steel	138	5	6	9	4.918
19	Copper	69	10	3	45	3.469
20	Copper	103	15	6	9	3.564
21	Copper	138	5	9	36	4.630
22	Copper	69	10	6	45	3.613
23	Copper	103	15	9	9	3.157
24	Copper	138	5	3	36	4.158
25	Copper	69	15	6	9	3.193
26	Copper	103	5	9	36	4.123
27	Copper	138	10	3	45	4.101
28	Copper	69	15	6	36	2.870
29	Copper	103	5	9	45	4.090
30	Copper	138	10	3	9	4.928
31	Copper	69	15	9	45	3.212
32	Copper	103	5	3	9	3.936
33	Copper	138	10	6	36	4.476
34	Copper	69	15	3	36	2.944
35	Copper	103	5	6	45	4.387
36	Copper	138	10	9	9	6.389

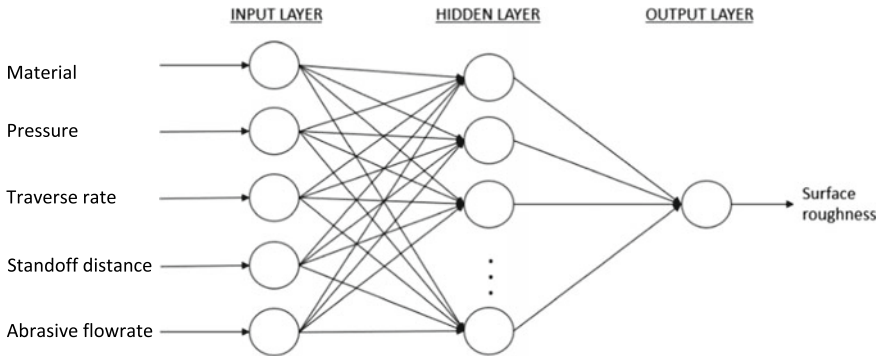


Fig. 1 ANN architecture for the surface roughness

neuron was connected to a new neuron in the next layer, and so on, until it reached the output layer. In each neuron, initially a random weight was assigned. Each neuron has an activation function. Following feed propagation, backpropagation was used to compute gradients in order to reduce the error by effectively updating the weights and bias from the output layer and propagating it backwards for each layer to achieve a value closer to the desired output. The gradient of the error function with respect to the weights of the neural network was calculated using chain rule.

3 Result and Discussion

The R_a for the experimental layout with the L36 orthogonal array were found to be between 2.506 and 6.389 μm as shown in Table 1. The ANN model used five input parameters namely material, water pressure, traverse speed, standoff distance and abrasive flowrate whereas the surface roughness was taken as the targeted output. All networks were feed-forward backpropagation models trained with the Levenberg Marquardt algorithm. All 36 experimental data to be fed to the ANN model where 26 experimental data (70%) were used for training, and 5 experimental data (15%) each were used for validation and testing, respectively.

After training the model, the learning curve (i.e., iteration number versus mean squared error) from the neural networks is shown in Fig. 2(a). It can be observed that the learning curve becomes steady after 18 epochs or iterations. The trained ANN model was achieved quickly thus minimizing the computational cost of ANN. Furthermore, the mean squared error for R_a is observed to be below 10% thus showing the well-trained network model can give a high level of accuracy in predicting the R_a [13]. Figure 2(b) shows 5 experimental data for testing with coefficient of determination (R^2) 0.99915 and Fig. 2(c) shows the comparison of R_a between the experimental and predicted data. Based on the model fitness, the coefficient of determination (R^2)

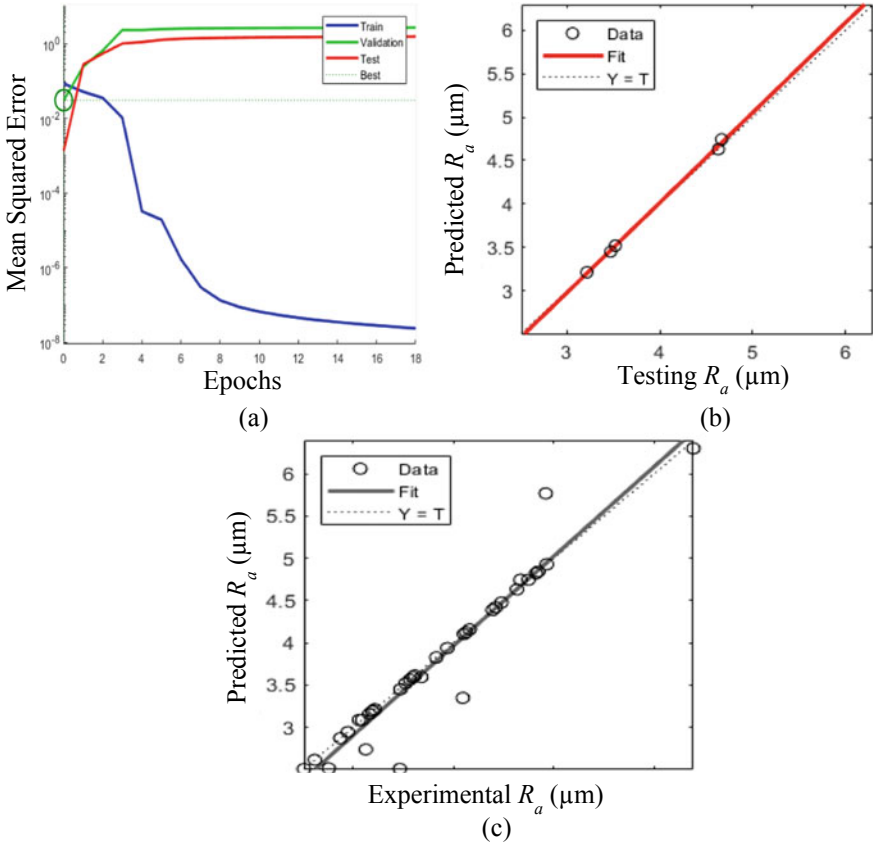


Fig. 2 a iteration number versus mean squared error b R_a for testing data c comparison of R_a between experimental and predicted data

was found to be 0.961. This shows that there is a good correlation between the experimental and predicted data for the R_a . It can be concluded that that the developed ANN model is capable of estimating the surface roughness within a reasonable range of error for the abrasive waterjet machining of sheet metals (Table 2).

Table 2 Actual Ra VS Predicted Ra

Material	p (MPa)	u (mm/min)	s (mm)	g (g/min)	Actual R_a (μm)	Predicted R_a (μm)
Stainless Steel	69	5	3	9	3.087	3.087
Stainless Steel	103	10	6	36	3.676	3.593
Stainless Steel	138	15	9	45	4.665	4.745
Copper	69	10	3	45	3.469	3.450
Copper	103	15	6	9	3.564	3.564
Copper	138	5	9	36	4.630	4.630

4 Conclusion

The relationship between the input parameters and the surface roughness during the abrasive waterjet machining of sheet metals was determined by developing an ANN model based on the Taguchi experiment design. A total of 36 experimental data were used from L_{36} orthogonal array. The ANN model was found to produce a high level of prediction accuracy with the mean squared error for the surface roughness to be below 10%. This summarizes that ANN model can sufficiently estimate surface roughness in the abrasive waterjet machining of sheet metals with a reasonable error range.

Acknowledgements Authors would like to gratefully acknowledge the financial support from Universiti Malaysia Pahang through RDU182203-2.

References

1. Folkes J (2009) Waterjet—an innovative tool for manufacturing. *J Mater Process Technol* 209(20):6181–6189. <https://doi.org/10.1016/j.jmatprotec.2009.05.025>
2. Begic-Hajdarevic D, Cekic A, Mehmedovic M, Djelmic A (2015) Experimental study on surface roughness in abrasive water jet cutting. *Procedia Eng* 100:394–399. <https://doi.org/10.1016/j.proeng.2015.01.383>
3. Khan AA, Haque MM (2007) Performance of different abrasive materials during abrasive water jet machining of glass. *J Mater Process Technol* 191:404–407. <https://doi.org/10.1016/j.jmatprotec.2007.03.071>
4. Murugan M, Gebremariam MA, Hamedon Z, Azhari A (2018) Performance analysis of abrasive waterjet machining process at low pressure. *IOP Conf Ser Mater Sci Eng* 319:012051
5. Ahmed TM, El Mesalamy AS, Youssef A, El Midany TT (2018) Improving surface roughness of abrasive waterjet cutting process by using statistical modeling. *CIRP J Manuf Sci Technol* 22:30–36. <https://doi.org/10.1016/j.cirpj.2018.03.004>
6. Summers DA, Yao J, Blaine JG, Fossey RD, Tyler LJ (1992) Low pressure abrasive waterjet use for precision drilling and cutting of rock. In: *Jet cutting technology*. Springer, Heidelberg, pp 233–251
7. Zou ZL (2012) Study of cutting composite materials with low pressure abrasive-water jet. *Appl Mech Mater* 130–134:1480–1483

8. Madara SR, Pillai SR, Pon Selvan MC, Van heirle J (2021) Modelling of surface roughness in abrasive waterjet cutting of Kevlar 49 composite using artificial neural network. *Mater Today Proc* 46:1–8. <https://doi.org/10.1016/j.matpr.2020.02.868>
9. Azmir MA, Ahsan AK (2008) Investigation on glass/epoxy composite surfaces machined by abrasive water jet machining. *J Mater Process Technol* 198:122–128. <https://doi.org/10.1016/j.jmatprotec.2007.07.014>
10. Selvan CP, Midhunchakkaravarthy D, Pillai SR, Madara SR (2019) Investigation on abrasive waterjet machining conditions of mild steel using artificial neural network. *Mater. Today Proc* 19:233–239. <https://doi.org/10.1016/j.matpr.2019.06.757>
11. Markopoulos AP, Manolakos DE, Vaxevanidis NM (2008) Artificial neural network models for the prediction of surface roughness in electrical discharge machining. *J Intell Manuf* 19:283–292. <https://doi.org/10.1007/s10845-008-0081-9>
12. Sanjay C, Jyothi C (2006) A study of surface roughness in drilling using mathematical analysis and neural networks. *Int J Adv Manuf Technol* 29:846–852. <https://doi.org/10.1007/s00170-005-2538-8>
13. Çaydaş U, Haşçalık A (2008) A study on surface roughness in abrasive waterjet machining process using artificial neural networks and regression analysis method. *J Mater Process Technol* 202:574–582. <https://doi.org/10.1016/j.jmatprotec.2007.10.024>
14. Pappas M, Ntziantzias I, Kechagias J, Vaxevanidis N (2011) Modeling of abrasive water jet machining using Taguchi method and artificial neural networks. In: *NCTA 2011 – proceedings of international conference neural computing theory applications*, pp 377–380. <https://doi.org/10.5220/0003681103770380>.
15. Singh B (2021) Predicting airline passengers ' loyalty using artificial neural network theory. *J Air Transp Manag* 94:102080. <https://doi.org/10.1016/j.jairtraman.2021.102080>

You Are Too Loud! Classification of Psychological Conditions for Stress Detection System Using Galvanic Skin Response



Amirul Aidy Amiruldin and Aimi Shazwani Ghazali 

Abstract Stress is the most common mental health problem nowadays and the number of people who suffer due to this condition rises drastically. In identifying human psychological states, the project is dedicated to develop a minimally invasive stress detection system using physiological signals. The physiological sensor used is Galvanic Skin Response (GSR), which is carefully chosen based on the best-fitted sensor to match the study's objective. Stroop color test, digit span test, and noisy music are used to elicit stress conditions while calming music is employed to invoke relax conditions. Using within-subject design, results for classifying these psychological conditions (stress vs relax) show that Lazy IBk and Random Forest correctly classify the GSR readings with 97.8% accuracy. For extending and quantifying the outcomes from the sensor, a set of questionnaires is also utilized for psychometric measures. Outcomes from the questionnaires demonstrate that the participants experience higher anxiety and higher reactance in stress condition than in relax condition. A future study might include other physiological signals for attaining robust classifications in developing a better stress detection system.

Keywords Stress detection system · Galvanic Skin Response (GSR) · Classification

1 Introduction

According to World Health Organization [1], over 260 million people in the world suffer because of mental health problems such as depression, and this number increased significantly. Known as a common mental disorder, depression is caused by long-term stress. This condition can trigger people to think irrationally which later may lead to suicide. Suicide tendencies due to stress and depression have been a huge problem around the globe, including Malaysia [2] due to several causes such as financial pressure and time management issues.

A. A. Amiruldin · A. S. Ghazali (✉)

Department of Mechatronics Engineering, International Islamic University Malaysia,
53100 Kuala Lumpur, Gombak, Malaysia
e-mail: aimighazali@iium.edu.my

Stress is defined as an emotional situation experienced when someone perceives expectations beyond that individual's personal and social resources that he or she can manage [3]. Ramachandiran and Dhanapal (2018) [4] explained that stress can be triggered when "the internal and/or environmental requirements are greater than the adaptive capacity of an individual or social structure" [5].

An earlier study showed that stress and human physiology can be correlated to one another [6]. Anatomically, physiological reactions that is related to stress are constrained by Autonomic Nervous System (ANS). The system controls significant functions of the body including digestion, temperature, circulatory pressure, and numerous parts that relate to the behavior of the emotions. ANS can be split into two parts which are Parasympathetic Nervous System (PNS) and Sympathetic Nervous System (SNS). While PNS regulates the essential functions of energy to restore rest, maintenance, and reconstruction, SNS monitors the bodily response that is deployed during emergencies by characterizing the response to flight or fight in several situations.

By taking the SNS part as an advantage, on one hand, earlier research used physiological signals such as Skin Temperature (ST), Electromyography (EMG), and Heart-Rate Variability (HRV) to detect stress [7–9]. However, these types of signals are hassle to be measured due to noise or contamination from several factors like mechanical and stimulus artefacts [10] besides the uniqueness of human bodily conditions (such as heart oscillations) from one another [11]. On the other hand, some studies [12, 13] used questionnaires for example Adolescent Stress Questionnaire [14] and State-Trait Anxiety Inventory (STAI) [15] to identify stress. However, findings of these studies are limited [7–9, 12] as none of them has properly related the physiological measures with psychometric measures. Importantly, the development of a stress detection system would far more convincing and rigid if the system considers differentiating the stress and relax conditions using classifiers.

Thereby, the main objective of the study is to develop a minimally invasive stress detection system using Galvanic Skin Response (GSR) by classifying the signal (stress and relax conditions) accordingly. In achieving the main objective, several sub-objectives are identified which are (1) To design a structured experimental procedure in measuring stress and relax conditions using physiological signals (2) To establish standard stressors (stimuli) that trigger stress condition during the experiment (3) To analyze and classify the readings from the GSR sensor (4) To establish the relationship between physiological signal (from the GSR) and psychological data collected from the questionnaire.

GSR or known as electrodermal activity (EDA) is a marker of stimulation for human nervous system. GSR is considered one of the most responsive and true emotional excitation markers [7]. Under stress conditions, skin resistance will decrease as the secretion in the sweat glands increased. This situation causing the GSR readings to decrease. As GSR sensor can measure the intensity of the sweat secretion precisely and easily on both feet and hands [16], GSR is selected to be used in this study. Additionally in predicting the stress vs relax condition, the GSR readings will be classed using supervised learning technique. The readings will be segregated into two types of datasets which are testing and validating datasets. Using

k-fold cross-validation technique, the readings will be fed to selected classifiers such as Naïve Bayes [17]. The accuracy of classifications will be compared in selecting the most suitable classifier to be used with our GSR readings.

Additionally, stressors or stimulus are introduced to the participants in triggering significant changes in their physiological signal level - from relax to stress condition vice versa. Various types of stimuli and stressors have been tested and used in the earlier research for eliciting stress namely Stroop Color and Word Test [18], Digit Span [16], as well as Trier Social Stress Test (TSST) [19]. For instance, Stroop Color and Word Test focused on mind tricks in increasing stress levels. That is, the subjects are given an orange-colored word written as “BLUE”. There are six color boxes displayed on left- and right-hand sides of the word, for example red, green, yellow, pink, orange, and blue boxes. For a right answer, the subjects must click or select an orange color box, instead of a blue color box in a very short duration of time. However, as people recognize words much quicker than they could even recognize the color of the words, most people fail to complete this test. As such, stress level is increased. Differently for Digit Span [16], the participants are asked to read a list of numbers and reproduce the same series in the forward order or the backward order. Forward order catches productivity and focuses on the power of the person while backward order is an executive function that depends especially on working memory. Specifically for this study, Stroop Colour and Word Test besides Digit Span (we name these tests as IQ test) were chosen to invoke stress based on the earlier research’s findings [16, 18]. Differently for eliciting relax conditions, music titled “Weightless” by the Macaroni union was chosen for the first relax phase [20] as the music able to reduce anxiety levels and bring calmness. The subjects are asked to just sit and doing nothing.

2 Study Design

2.1 Subjects

Twenty-eight subjects (22 males and 6 females) participated in this study with ages ranging between 17 and 56 ($M = 26.36$, $SD = 8.90$). The experiment lasted for twenty-five minutes in which all subjects were given some sweets and beverages after the experimental session was completed. Subjects were randomly selected with no restriction of gender and age.

2.2 Measures

Physiological Measure: GSR. Grove GSR sensor was used as an extension for Arduino UNO to measure the level of skin conductance of the subjects. Coding

was done in Arduino IDE to read and print the voltage value of the measured GSR. Figure 1 illustrates the set ups for physiological measures utilized in this study. It can be seen that the electrodes were placed on the distal phalanges of the subject's middle and index fingers on one non-dominant hand as the dominant hand will be used for playing the dedicated "game" in stress conditions later.

Psychometric Measure: Questionnaire. A set of questionnaires was given to the subjects to evaluate their psychological condition consist of four questions from the psychological reactance scale [21] and six questions from State-Trait Anxiety Inventory (STAI) [15] as used in the earlier study [16]. The Likert scale of 4 ranging from "Not at all" to "Very much" toward the following statements:

Reactance Scale [17]. I feel *angry, annoyed, irritated, aggravated*.

STAI or Anxiety Scale [15]. I feel *upset, uncomfortable, indecisive, jittery, confused, worried*.

Also, three hypotheses for psychometric analysis were investigated:

Hypothesis 1. There is a significant correlation between anxiety and reactance scores for each relax and stress condition.

Hypothesis 2. There is a significant difference of phases on (a) anxiety and (b) reactance perceived in all relax condition.

Hypothesis 3. There is a significant difference of phases on (a) anxiety and (b) reactance perceived in all stress condition.

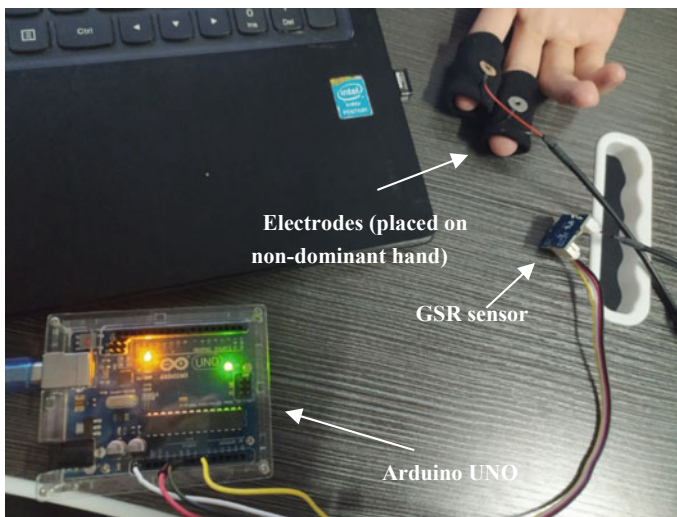


Fig. 1 Hardware set ups for physiological measure using GSR

2.3 Experimental Set Ups

The subjects were asked to take a sit in front of a dedicated laptop prior to the experiment. The experimental set ups were designed to position the subjects in a state of emotional and cognitive discomfort repetitively. This was done to analyze the differences of GSR readings when the subjects in their stress-induced physiological parameters or known as “stress” condition and in the rest or known as “relax” condition. There are five phases in this experiment namely “first relax phase (R1)”, “first stress phase (S1)”, “second relax phase (R2)”, “second stress phase (S2)”, and “final relax phase (R3)” as summarized in Fig. 2. Each phase lasts five minutes.

Prior to the experiment, the subjects were asked to read and sign a dedicated consent form. For R1, the subjects were asked to sit in a quiet environment without any distractions including from the cell phone. They were asked to listen to a piece of calming music using headphones. This process was crucial for acquiring each subject’s baseline as physiological metrics between all subjects are varies. Then, a series of psychological questions (consist of reactance and STAI questions) were given to the subjects before S1 was pursued.

After that, the subjects carried out the first stress induction phase (S1), in which they were required to complete the first set of IQ tests (Stroop Color and Word test) provided by the experimenter to elicit stress conditions. In parallel, the subjects were asked to listen to a piece of noisy background music to create tension during their tests; replicating the procedure by Asif et al. (2019) [22]. The subjects were equipped with headphones so that they can fully immerse themselves with the provided music background during the test.

After S1 was completed, R2 was subsequently conducted in the same conditions as in R1 process. Then, S2 was carried out in the same condition as in S1 but with a different set of IQ tests (Digit Span Test). Lastly, the experiment ended with R3 to reduce the subject’s stress level back to normal condition. This phase duplicated the same condition as R1. The subjects were asked to wear the GSR sensors during



Fig. 2 Summary: flow of the experiment

the entire experimental process (R1, S1, R2, S3 and R3) to continuously monitor the changes in their skin conductance.

3 Results and Discussion

3.1 Physiological Analysis

The raw voltage value of GSR has similar patterns of voltage regardless of the individual's GSR readings. That is, low voltages were recorded in stress conditions while high readings in relax conditions. However, as the data was too large to be analyzed, the sample size data was further reduced by averaging the voltage readings for every 30 s. The newly obtained dataset was then being normalized as a whole, regardless of the phases. After normalization, each subject only has a total of 46 data to be analyzed as the data during the transition of phase were omitted. Figure 3 illustrates the graphical representation of the normalized GSR data. The X-axis denotes the phase while the y-axis shows the GSR readings.

The dataset obtained was then being further evaluated to assess the relationship between the voltages generated by GSR sensor (as the input for the classifiers) and the emotional phases involved (as the predicted class for the classifiers). Using Waikato Environment for Knowledge Analysis (WEKA) software¹, two classifiers were selected to be used in this study namely Lazy IBk and Random Forest. As the test option, the classification process was set to a cross-validation fold of ten (by default). Lazy IBk is a k-nearest-neighbour classifier that uses a distance measure to classify k "close" instances for each test instance in the training data and uses those selected instances to make a prediction [23]. Differently, Random Forest consists of a large number of individual decision trees which act as an ensemble. A class

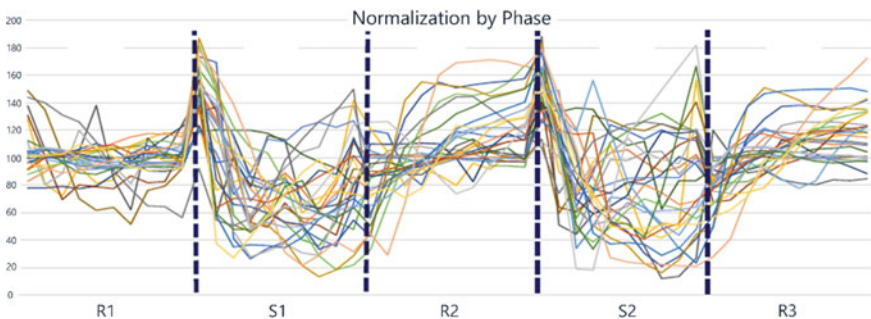


Fig. 3 GSR patterns after being normalized

¹ <https://www.cs.waikato.ac.nz/ml/weka/>.

Table 1 Confusion matrix for Lazy IBk and random forest classifications in two classification tests

Actual psychological condition	Predicted psychological condition		Total	Accuracy (%)
	Relax	Stress		
Relax	28	0	28	100.0
Stress	1	17	18	94.4

prediction is generated by each tree in the Random Forest and the class with the most votes become the prediction of the model [24].

As a result of two separate classification tests, Lazy IBk and Random Forest successfully classified the data with an accuracy of 97.8%. That is, both classifiers correctly classified all relax conditions datasets (100.0%) and only 1 set of data (5.6%) was incorrectly classified for stress conditions. As classification outputs for both classifiers are exactly the same, details for Lazy IBk and Random Forest classification are elaborated using one confusion matrix shown in Table 1.

In terms of classifier selection, both Lazy IBk and Random Forest produced the highest classification accuracy compared to other classifiers such as Input Mapped Classifier (with an accuracy of 60.9%), Lazy LWL (with an accuracy of 93.5%), and Naïve Bayes Multinomial (with an accuracy of 95.7%).

In summary, the classification results proved the effectiveness of using GSR in determining human psychological condition subjected to stress and relax conditions.

3.2 Psychometric Analysis: Hypotheses Testing

A set of questionnaire consists of ten questions (six questions from STAI for anxiety scale [15] and four questions from reactance scale [17]) was utilized in the study. These questions were given to the subjects at the end of each phase. For analysis, mean and standard deviation values for each question in each phase were then calculated as presented in Table 2.

The data accumulated from the questionnaires were then analyzed using the Statistical Package for Social Science (SPSS)².

Hypothesis 1. Using correlation test, results showed that hypothesis 1 was partially accepted. That is, anxiety for R1 and R3 were found to be moderately correlated, $r(28) = 0.40, p = 0.02$, anxiety and reactance for S1 to be weakly correlated, $r(28) = -0.26, p = 0.88$, reactance for S1 and anxiety for S2 to be weakly correlated, $r(28) = 0.26, p = 0.89$, anxiety for R2 and reactance for R3 to be moderately correlated, $r(28) = -0.44, p = 0.01$, reactance for R2 and R3 to be weakly correlated, $r(28) = -0.30, p = 0.06$ besides anxiety and reactance for S2 to be strongly correlated, $r(28)$

² <https://www.ibm.com/my-en/products/spss-statistics>.

Table 2 Means and standard deviations of questions based on phase

	Question(s)	Phase: Mean (Standard Deviation)				
		R1	S1	R2	S2	R3
Anxiety	Upset	1.07 (0.26)	3.36 (0.73)	1.50 (0.51)	3.43 (0.57)	1.54 (0.51)
	Uncomfortable	1.07 (0.26)	3.21 (0.79)	1.75 (0.44)	3.50 (0.51)	1.32 (0.48)
	Indecisive	1.21 (0.42)	3.14 (0.80)	1.50 (0.51)	3.64 (0.49)	1.46 (0.51)
	Jittery	1.00 (0.00)	3.25 (0.75)	1.71 (0.53)	3.21 (0.88)	1.68 (0.48)
	Confused	1.18 (0.39)	3.04 (0.84)	1.50 (0.51)	3.32 (0.77)	1.64 (0.56)
	Worried	1.14 (0.45)	3.14 (0.85)	1.46 (0.58)	3.25 (0.59)	1.54 (0.51)
Reactance	Angry	1.00 (0.00)	3.18 (0.77)	1.61 (0.50)	3.14 (0.71)	1.46 (0.51)
	Annoyed	1.00 (0.00)	3.32 (0.77)	1.50 (0.51)	3.39 (0.57)	1.50 (0.51)
	Irritated	1.00 (0.00)	3.43 (0.69)	1.46 (0.51)	3.46 (0.58)	1.50 (0.51)
	Aggravated	1.00 (0.00)	3.36 (0.68)	1.57 (0.57)	3.29 (0.66)	1.43 (0.50)
Overall		1.07 (0.18)	3.24 (0.77)	1.56 (0.52)	3.36 (0.63)	1.51 (0.51)

= 0.91, $p < 0.01$. Other correlations were not significant, $p > 0.10$. Other correlation outputs were not reported due to insignificant results.

Hypothesis 2. A repeated one-way measure of ANOVA test for STAI (anxiety questions) was run by comparing the three relax phases as independent variables and the anxiety scores as dependent variables. Results showed that there was a significant difference of the relax phases (R1, R2 and R3) on anxiety perceived by subjects, $F(2, 26) = 119.74$, $p < 0.001$, partial $\eta^2 = 0.90$. Based on descriptive analysis, it can be observed that the subjects experienced the lowest anxiety scores on the R1 ($M = 1.11$, $SD = 0.12$) in which they were not exposed to any stress environments yet, followed by R3 ($M = 1.53$, $SD = 0.18$). The highest anxiety score was recorded during R2 ($M = 1.57$, $SD = 0.19$), that is after the subjects were exposed to S1.

Analysis for reactance used the same test as in anxiety. Results recorded that there was a significant difference of relax phases (R1, R2 and R3) on reactance, $F(2, 26) = 149.30$, $p < 0.001$, partial $\eta^2 = 0.08$. Interestingly, based on the analysis of mean value, subjects showed the lowest reactance scores on R1 ($M = 1.00$, $SD = 0.00$), followed by R3 ($M = 1.47$, $SD = 0.28$), and the highest reactance recorded in R2 ($M = 1.54$, $SD = 0.25$).

To conclude hypothesis 2, both anxiety and reactance scores were significantly depending on relax conditions (the lowest score during R1, subsequently R3 and R2) which indicated that hypothesis 2 was accepted.

Hypothesis 3. A repeated measure one-way ANOVA was run to evaluate the main effect of stress phase (as independent variables) on anxiety scores (dependent variables). Results demonstrated that there was no significant difference of S1 and S2 on anxiety perceived by subjects, $F(1, 27) = 1.48$, $p = 0.235$, partial $\eta^2 = 0.052$.

For reactance, results showed that there was a significant difference of reactance measured in S1 and S2 with $F(1, 27) = 4.55$, $p = 0.04$, partial $\eta^2 = 0.14$. Subjects

recorded the highest reactance scores on the S1 ($M = 3.32$, $SD = 0.41$) while the lowest reactance scores recorded on S2 ($M = 3.00$, $SD = 0.80$) in which the subjects were more familiar with the stress-inducing phase.

Hypothesis 3 was partially accepted; that is, only reactance score was significantly depending on stress conditions (S1 and S2).

4 Conclusions

The study indicates the potential in measuring human psychological conditions using GSR for developing stress detection system. That is, we successfully design a structured experimental procedure to invoke stress and relax by establishing standard stressors (stimuli) to trigger such conditions. Also, this study has successfully classified the psychological conditions of human mental health using collected data from GSR using Lazy IBk and Random Forest with 97.8% accuracy. The psychometric analysis strengthens the results found in the physiological analysis. That is, high anxiety and reactance scores are recorded in the stress conditions, while low anxiety and reactance scores are found in the relax conditions. As GSR proposed in this study may not be the best suggestions for long term usage due to comfortability issue, future work might venture in using other physiological sensors that are smaller in size and more ergonomic than GSR in attaining higher accuracy results in detecting stress.

References

1. Depression (2020). <https://www.who.int/news-room/fact-sheets/detail/depression>
2. Subramaniam A (2021) Addressing joblessness, suicide. <https://www.nst.com.my/opinion/letters/2021/03/678017/addressing-joblessness-suicide>
3. Bhargava D, Trivedi H (2018) A study of causes of stress and stress management among youth. *IRA Int J Manage Soc Sci* 11(03):108–117
4. Ramachandiran M, Dhanapal S (2018) Academic stress among university students: a quantitative study of generation Y and Z's perception. *Pertanika J Soc Sci Humanit* 26(3):2115–2128
5. Lazarus RS, Cohen JB (1977) Environmental stress. In: Altman I, Wohlwill JF (eds) *Human behavior and environment*, pp 89–127. Springer, Boston. https://doi.org/10.1007/978-1-4684-0808-9_3
6. Thawabieh AM, Qaisy LM (2012) Assessing stress among university students. *Am Int J Contemp Res* 2(2):110–116
7. Kyriakou K et al (2019) Detecting moments of stress from measurements of wearable physiological sensors. *Sensors* 19(17):3805
8. Michael S, Mwape F, Jeneffa L (2019) An assessment of inoculation of academic stress in education sector. *DMI-St. Eugene Univ Mult Disciplinary J Sci Manage Technol* 1(1)
9. Salai M, Vassányi I, Kósa I (2016) Stress detection using low cost heart rate sensors. *J Healthcare Eng* 2016:1–13. <https://doi.org/10.1155/2016/5136705>
10. Yang D et al (2017) Accurate EMG onset detection in pathological, weak and noisy myoelectric signals. *Biomed Signal Process Control* 33:306–315

11. Shaffer F, Ginsberg J (2017) An overview of heart rate variability metrics and norms. *Front Public Health* 5:258
12. Blanca MJ et al (2020) Psychometric properties of a short form of the adolescent stress questionnaire (ASQ-14). *Psicothema* 32(2):261–267
13. Hung H-Y, Hung S-P, Chang Y-J (2020) Development and validation of the prenatal activity restriction stress questionnaire: a Rasch rating scale analysis. *BMC Pregnancy Childbirth* 20(1):1–13
14. Byrne DG, Davenport S, Mazanov J (2007) Profiles of adolescent stress: the development of the adolescent stress questionnaire (ASQ). *J Adolesc* 30(3):393–416
15. Spielberger CD et al (1999) STAI: Cuestionario de ansiedad estado-rasgo. TEA ediciones Madrid
16. Acerbi G et al (2017) A wearable system for stress detection through physiological data analysis. In: Cavallo F, Marletta V, Monteriù A, Siciliano P (eds) *Ambient assisted living. ForItAAL 2016*. LNEE, vol 426. Springer, Cham. https://doi.org/10.1007/978-3-319-54283-6_3
17. Satti FA et al (2021) User stress modeling through galvanic skin response. In: 2021 15th international conference on ubiquitous information management and communication (IMCOM). IEEE
18. Scarpina F, Tagini S (2017) The stroop color and word test. *Front Psychol* 8:557
19. Mozos OM et al (2017) Stress detection using wearable physiological and sociometric sensors. *Int J Neural Syst* 27(02):1650041
20. Rather JA, Shrivastava Y (2019) Effect of music therapy on pre-competition anxiety in college level soccer players of Kashmir
21. Dillard JP, Shen L (2005) On the nature of reactance and its role in persuasive health communication. *Commun Monogr* 72(2):144–168
22. Asif A, Majid M, Anwar SM (2019) Human stress classification using EEG signals in response to music tracks. *Comput Biol Med* 107:182–196
23. Alshammari M, Mezher M (2020) A comparative analysis of data mining techniques on breast cancer diagnosis data using WEKA toolbox. (IJACSA) *Int J Adv Comput Sci Appl* 8:224–229
24. Phan TN, Kuch V, Lehnert LW (2020) Land cover classification using google earth engine and random forest classifier—the role of image composition. *Remote Sens* 12(15):2411

Universiti Malaysia Pahang Autonomous Shuttle Development: Lane Classification Analysis Using Convolutional Neural Network (CNN)



Lee Yin Yee and Muhammad Aizzat Zakaria

Abstract In recent years, the widespread adoption of autonomous vehicle, advanced driver assistant systems (ADAS) have acquired great interests as it provides safe and better driving by automating, adapting, and enhancing the driving experience. Road accidents can be avoided with the identification of various road infrastructures such as merging or splitting lanes as well as ending lanes must be well detected, providing a driver with a more convenient and safe intelligent function. However, the image for lane detection failed to be detected due to the visibility of image is affected because it may consist of noise, occlusion, undesired background blur and the image pixels. To predict the lane markers on road pavement correctly, a robust lane classification system using deep learning approach requires guidance so that it can detect significantly. Four significant operations involve in developing the system which are data acquisition, data pre-processing, data training and data testing. In this study, an improved classification algorithm using deep learning specifically convolutional neural network is used to detect the lane markers. The big dataset consists of 5000 images. It is distributed into are 4000 images as training data, 700 images as validation data and 300 images as testing data respectively. For the evaluation of lane detection system, the evaluation metrics are in terms of accuracy, false positive (FP) and false negative (FN). The accuracy of the lane classification system network is 91.97%.

Keywords Machine learning · Convolutional Neural Network · Lane classification

L. Y. Yee · M. A. Zakaria (✉)

Intelligent Robotic and Vehicle (IROV), IMAMS Laboratory, Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

e-mail: maizzat@ump.edu.my

Autonomous Vehicle Laboratory, Centre for Automotive Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

1 Introduction

Autonomous vehicles (AVs) have the potential to improve logistics and passenger transfer methods [1]. Land transport such as road transport is the main mode of transport in Malaysia. Road accidents are commonly caused by reckless drivers such as misconduct, carelessness and diversion [2]. According to World Health Organization (WHO), every year, approximately 1.35 million of lives are taken because of land transport crashes. Land transport injuries are the leading cause of death for youngsters aged between five to twenty-nine years and there are between twenty to fifty millions of lives suffer from non-fatal injuries but resulting a disability from the injury [3]. However, road accidents can be lessened by improving driving assistances with modern technologies. Lane classification on autonomous vehicle has been famous in the field of (AI) artificial intelligence used in transportation. Lane detection is a crucial part of autonomous vehicles. At present, several algorithms have been proposed to overcome and improve the issue as it is one of the most extensive research topics for driverless cars.

To accelerate the development of autonomous vehicles on public roads, several studies on autonomous vehicles are being conducted in specific areas such as university campus. In campus regions, autonomous cars can provide transportation services while minimizing the number of vehicles on campus and this improves the environment by reducing congestion and pollution. The Autonomous Vehicle Laboratory team from Universiti Malaysia Pahang is also conducting research on the development of autonomous vehicles [4]. Figure 1 illustrates the autonomous prototype of the compact shuttle. The camera embedded on the autonomous shuttle will be used to collecting data for research like lane classification and object tracking for autonomous vehicle.

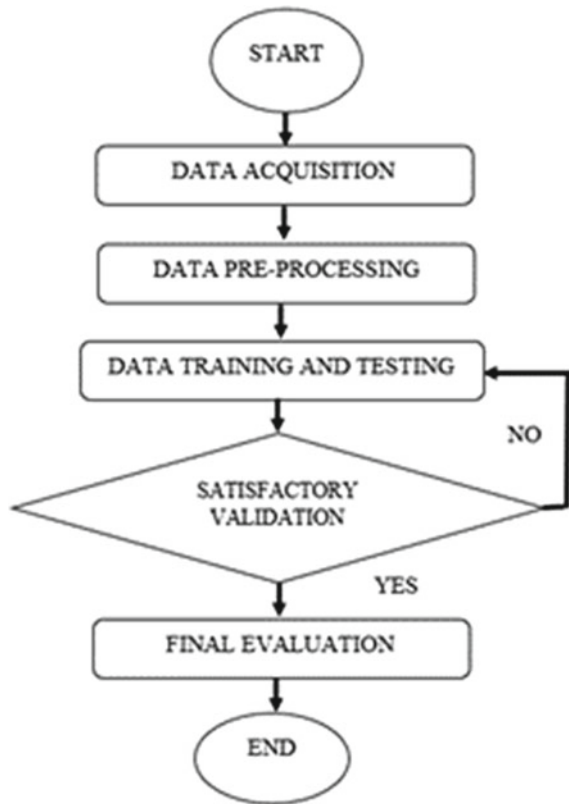
Recent studies on deep learning enables the unification of classification issue into a semantic segmentation task [5]. Convolutional Neural Networks are widely used in classification tasks and recently adapted to segmentation task manage to score accuracy that exceed any traditional approaches. China, Korea, Japan, and Singapore

Fig. 1 Autonomous prototype of compact shuttle with camera and lidar sensor mounted on top



are among the Asian countries who are contributing significantly in the field of self-driving vehicle [6]. More effort is needed to develop self-driving vehicles before these technologies can be used be deployed on a large scale in a reliable manner. Several algorithms are suggested such as [7] proposed a deep neural network-based method, in two stages: lane edge design and lane line localization. The lane line localization network is responsible for determining the location of each lane in the image based on the lane edge map. Authors [8] proposed a hybrid neural network combining CNN and RNN and structural analysis method able to achieve outstanding results for the curvature lane detection and it yields higher accuracy and gain a better fitting results for lane detection [9]. In this research, a lane classification system was developed based on Convolutional Neural Network (CNN) based on the encoder-decoder network which has a good robustness in deep learning approach.

Fig. 2 Research methodology



2 Methodology

The study conducted aims to evaluate, fine-tune and develop a suitable deep learning model to conduct lane classification for autonomous vehicle for industrial application. The custom dataset will undergo data sampling to obtain the training set and testing set. The training set will undergo several data augmentation procedures to improve the generalization of the model. The resulting data set will be fed into several deep learning models for feature extraction. Four major operations involve in the development of the lane classification system are as shown in the chart below (Fig. 2).

2.1 Dataset

As for the image dataset, the proposed method for lane classification system is tested on a new data set of 300 images that is obtained for different places at Universiti Malaysia Pahang, Pekan Campus under sunny weather and illumination conditions only. The dataset includes 5000 images of highway sceneries with dotted lane markers and solid white lane markers only. The collection of dataset used for evaluation purpose will be images taken using camera embedded at autonomous shuttle UMP at various place whereas Tusimple dataset is captured under stable lighting conditions [10] (Table 1).

2.2 Data Acquisition

The UMP dataset used in this research is captured with a camera with horizontal mode at 1280×720 resolution. Additional dataset of UMP road images consists of two lanes and various curvature lanes. This will cover the real driving scenarios in UMP Pekan. Besides, different illumination images are included in the dataset to produce better generalization even though the scope of project is narrowed to sunny

Table 1 Dataset description using TUSimple and UMP Road

Options	Descriptions
Dataset	TuSimple and UMP road
Frame	5000
Train	4000
Test	300
Validation	700
Resolution	1280×720
Road type	Highway and UMP road

day only. For road curvature dataset, curvy lane, mixture of straight and curvy and straight lane consists of 40, 32, and 28% respectively. However, for the illumination and weather dataset, it consists of 65% sunny afternoon, 15% morning and 20% cloudy evening.

2.3 Training and Testing Network

For training, 4000 images from the total frames are used as training dataset. Network 1 is trained with images obtained from online source while Network 2 is trained with the combination dataset of online dataset and UMP road dataset. Both networks are trained under CNN architecture based on the encoder-decoder network. Encoder is used to extract and encode the information contained in the initial image into lower resolution feature map while decoder is to preserve the context information from the lower resolution map while recovering fine-grained details through up-sampling layers. ResNet-34 is used fine-tuned and trained to apply on lane extractor for lane classification system. The selected backbone used in extracting preliminary feature at this stage is VGG-16 which could perform classification on raw image without causing overfitting occurs. In the evaluation stage, the evaluation metrics which will be taken into consideration is accuracy, false positive and false negative (Table 2).

2.4 Data Testing

In practical engineering, pixel accuracy (PA) and mean intersection over union (mIoU) are applied on evaluation of the performance of the semantic segmentation method. Pixel Accuracy is used to measure the object contour segmentation. Intersection over Union is a metric for determining how accurate an object detector is on a given dataset and Mean of Intersection Over Union is a ratio between the expected segmentation and the ground truth segmentation which reflects the overall improvement in sematic segmentation accuracy.

Table 2 Parameter configuration for training parameters

Options	Descriptions
Image height	368
Image width	640
Weight	0.4
Learning rate	0.001
Batch size	4
Threshold value	0.60
Epoch	100

$$Pixel\ Accuracy = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij'}} \quad (1)$$

$$Intersection\ over\ Union = \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (2)$$

P_{ii} : the representation of true positives;

P_{ij} : the amount of pixels of class I inferred to class j

P_{ij} : False positive;

P_{ji} : False positive;

For the evaluation metrics used to make a representative comparison for lane classification system is calculated as:

$$Accuracy = \frac{N_{predicted}}{N_{groundtruth}} \quad (3)$$

where $N_{predicted}$ is the number of lane markers that are correctly predicted while $N_{groundtruth}$ is the number of ground truth

$$FP = \frac{F_{predicted}}{N_{predicted}} \quad (4)$$

$F_{predicted}$ is the number of lane markers wrongly predicted.

$$FN = \frac{M_{predicted}}{N_{groundtruth}} \quad (5)$$

$M_{predicted}$ is the number of missed lane based on ground truth.

FP is the number of background pixels that are incorrectly predicted as lanes whereas FN is the number of lane pixels that are incorrectly predicted as background.

3 Results and Discussion

3.1 Image Segmentation

In the experiment conducted, the pixel accuracy (PA) running at 50 epochs is 98.7% with 0.325 mIoU. The PA and mIoU increase as it raises to 80 epochs, 110 epochs, and 150 epochs which are 98.9% with 0.396, 99% with 0.445 and 99.1% with 0.478 respectively. For dataset running at 160 epochs, the pixel accuracy (PA) achieved the same accuracy as in dataset running at epoch 150 which is at 99.1% but the mIoU for 160 epochs is lower than that of 150 epochs at 0.449. All in all, the PA ratings display the proposed algorithm has a better object contour detection while for

Table 3 Image segmentation performance

Trial	Epoch	Pixel Accuracy (PA)	Mean Intersection over Union (mIoU)	Response Time (s)
1	150	0.991	0.478	0.066
2	160	0.991	0.449	0.064
3	110	0.99	0.445	0.064
4	80	0.989	0.396	0.066
5	50	0.987	0.325	0.095

the score of mIoU, training dataset running at epoch 150 has the best classification accuracy when compared to the other four different epochs. The best results acquired is the model running at epoch 150 for two experiments with two different amounts of dataset. We can conclude that the best epoch for running the model to obtain most favorable results is at epoch 150 with a greater number of images input during training process. As the pixel accuracy is high in this network, this can be summarized as the network has excellent detection which is suitable to be applied on self-driving cars as it provides accurate prediction which helps in navigation in lane detection system (Table 3).

3.2 Evaluation on Deep Learning Method

This section displays the prediction of lane classification system trained by Network 1 for 10 times. Network 1 is trained with the combination of online dataset and UMP road dataset which increase the accuracy of the system. Ten times of trial, most of the trial able to achieve higher FN and lower FP except for the last trial where FP score is higher than FN. Theoretically, bigger dataset with higher resolution able to generate a better detection system as what the proposed model shows. A classification-based network could easily over-fit the training set and perform poorly on the validation set because to the underlying structure of lanes. However, the proposed model did not relate with it. Therefore, we can conclude that, the model does not overfitting during training where its good performance in training process able to generalize well on new or unseen data. In other words, the model manages to learn particular pattern of the training data which is relevant in other data. To avoid this problem and gain generalization ability, an augmentation method consisting of rotation, vertical, and horizontal shift is applied. A large number of epochs is chosen because our structure-preserving data augmentation takes a long time to learn. This adds to more accurate and precise lane prediction in practical systems. The experimental results also proven that lane classification performance is significantly improved from traditional approach to deep learning method. Table 5 illustrates the accuracy obtained at average after 10 trials (Table 4).

Table 4 Results obtained from training process of lane classification system

Trial	Accuracy	False Positive (FP)	False Negative (FN)
1	0.7436	0.5331	0.5732
2	0.8825	0.2199	0.2366
3	0.9306	0.0800	0.0965
4	0.9349	0.0769	0.0917
5	0.9478	0.0468	0.0638
6	0.9462	0.0547	0.0634
7	0.9530	0.0518	0.0523
8	0.9461	0.0633	0.0646
9	0.9554	0.0442	0.0505
10	0.9572	0.0463	0.0431

Table 5 Table of mean accuracy of lane classification system

Trial	Accuracy
1	0.7436
2	0.8825
3	0.9306
4	0.9349
5	0.9478
6	0.9462
7	0.9530
8	0.9461
9	0.9554
10	0.9572
Average	0.9197

3.3 Evaluation on Deep Learning Method on Sunny Day

As explained, the training model has performed in various places in UMP Pekan with different scene variations under sunny day condition. In the presented output images, the position of lanes in the outcomes of the machine learning framework is strictly aligned with the ground truth. The lane markers are highlighted in different colours and these coloured pixels are the result of the processing phase of the algorithm (Figs. 3, 4, 5 and 6).

Fig. 3 Faculty of Computing, UMP



Fig. 4 Road to main entrance of UMP



Fig. 5 Road to Canseleri, UMP

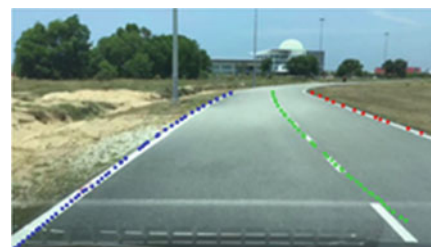


Fig. 6 Dewan Serbaguna, UMP



4 Conclusion

In this paper, the development of lane classification system to detect road lane markers at Universiti Malaysia Pahang Pekan Campus is formulated. To develop a lane classification system, the main part is not only limited to computer vision but also the ability to articulate the road lane markers. The weakness of using traditional image

processing method is that the parallel road lanes might face some issue like white arrow on the road which will affect the accuracy of lane detection. Therefore, a deep learning approach is used. A set of training data are loaded into a network after data acquisition and data pre-processing to extract the features from the images. A total of 5000 frames is separated into 4000 as training data, 700 as validation data and 300 as the testing data respectively. From the trained network, the testing images are detected. Then accuracy is calculated based on the prediction and evaluation metrics used from the trained network. The evaluation metrics used to validate the robustness of the system are accuracy, FP and FN. In the end of the research, the lane classification system is able to achieve accuracy of 91.97%. It would be interesting if LiDAR and radar sensors are used for future work.

Acknowledgements The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) NO. FRGS/1/2021/TK08/UMP/02/2 (University reference RDU190104) and Universiti Malaysia Pahang for laboratory facilities as well as additional financial support under Internal Research grant RDU1903139.

References

1. Sell R, Rassölkin R, Wang R, Otto R (2019) Integration of autonomous vehicles and Industry 4.0. *Proc Est Acad Sci* 68(4):389. <https://doi.org/10.3176/proc.2019.4.07>
2. Bellis E, Page J (2008) National motor vehicle crash causation survey (NMVCCS) SAS analytical users manual
3. Peden M et al (2004) World report on road traffic injury prevention. World Health Organization
4. Zakaria MA, Kunjuni B, Peeie MHB, Papaioannou G (2021) Autonomous shuttle development at Universiti Malaysia Pahang: LiDAR point cloud data stitching and mapping using iterative closest point cloud algorithm. In: Hamid UZA, Al-Turjman F (eds) *Towards connected and autonomous vehicle highways*. EAISICC, pp 281–292. Springer, Cham. https://doi.org/10.1007/978-3-030-66042-0_11
5. Romera E, Alvarez JM, Bergasa LM, Arroyo R (2017) ERFNet: efficient residual factorized convnet for real-time semantic segmentation. *IEEE Trans Intell Transp Syst* 19(1):263–272
6. Daily M, Medasani S, Behringer R, Trivedi M (2017) Self-driving cars. *Comput (Long Beach Calif)* 50(12):18–23. <https://doi.org/10.1109/MC.2017.4451204>
7. Wang Z, Ren W, Qiu Q (2018) LaneNet: Real-time lane detection networks for autonomous driving. arXiv preprint [arXiv:1807.01726](https://arxiv.org/abs/1807.01726)
8. Zou Q, Jiang H, Dai Q, Yue Y, Chen L, Wang Q (2019) Robust lane detection from continuous driving scenes using deep neural networks. *IEEE Trans Veh Technol* 69(1):41–54
9. Ye YY, Hao XL, Chen HJ (2018) Lane detection method based on lane structural analysis and CNNs. *IET Intell Transp Syst* 12(6):513–520
10. Zheng T et al (2020) RESA: recurrent feature-shift aggregator for lane detection. arXiv preprint [arXiv:2008.13719](https://arxiv.org/abs/2008.13719)

Eco-Design of Electric Vehicle Battery Pack for Ease of Disassembly



X. Q. Chew, W. J. Tan, N. Sakundarini, C. M. M. Chin, A. Garg,
and S. Singh

Abstract Electric vehicle being one of the leading green technologies nowadays, is leaving a humongous amount of spent lithium-ion batteries untreated. Current research on lithium-ion battery waste management is at its minimal because the huge power range of the battery is much attractive than the battery waste dismantling process. Treating these battery wastes are crucial for rare metal recovery due to its limited resources on land. Thus, this study aims to propose an eco-design battery pack to ease the recycling process in a more economical and sustainable manner. SolidWorks is used to generate the 3D modelling and ANSYS is utilized to carry out the simulation of the product's mechanical performance in a drop and impact tests. Results shows that the proposed design of EV battery pack has a design efficiency of one with Easy Fixings indicator of 28%. In the drop test of 0.3 m height, it yields a maximum deformation of $1.015e^{-3}$ m and a generated Von-Mises stress of $4.827e^8$ N/m². Other than that, $2.5227e^6$ N/m² of Von-Mises stress is obtained in the impact frontal test. With a great impact of cruising at a speed of 15.6464 m/s, $5.6053e^{-8}$ m deformation is obtained in the same test. As a result, the proposed EV battery pack design has showed the potential to improve the sustainability, performance, and ease of disassembly.

Keywords Electric vehicle · Battery pack · Design of disassembly · Lithium-ion battery · Eco design

X. Q. Chew · W. J. Tan (✉) · N. Sakundarini · C. M. M. Chin
Department of Mechanical, Material and Manufacturing Engineering, University of Nottingham
Malaysia, Selangor, Malaysia
e-mail: kedy6twj@nottingham.edu.my

N. Sakundarini
e-mail: novita.sakundarini@nottingham.edu.my

A. Garg
State Key Lab of Digital Manufacturing Equipment and Technology, School of Mechanical
Science and Engineering, Huazhong University of Science and Technology, Wuhan, China

S. Singh
Department of Mining, Metallurgy and Materials Engineering, University of Laval, Quebec City,
QC, Canada

1 Introduction

Climate change has been around for the past decades. Greenhouse gases are reported as the main root cause of global warming which is a consequence from large consumption of fossil fuel for transportation [1]. Unfortunately, there is inadequate recycling technology and methodology to handle this amount of batteries due to difficulties in disassembling the battery packs [3]. Without proper management, massive EV battery waste will be produced, creating future environmental problems.

There are many types of EV batteries, but lithium-ion battery (LIB) dominates EV battery industry due to its relative low weight, high standard potential, and triple energy density compared to traditional batteries [4]. Therefore, LIBs are selected and examined in this study. In an EV, battery cells are stacked to create a module while modules are combined to make up battery pack. A typical rechargeable LIB system comprises of cells, busbars, wiring harness, battery management system (BMS), traction cable, vehicle interface, current measuring device, isolation monitoring device, main contactor relays, cooling system, covers for certain parts and some fastening components [2, 4].

Challenges in disassembling the battery pack include the heterogeneous nature of its components. Having different materials, with joining technologies and connection makes the process time-consuming and economically difficult [3]. With that being said, what makes it even complicated is the diversity of screw types and flexible components used on different connecting application. It directly reflects on the number of different types of screwdrivers or tools needed in the dismantling process [2]. Furthermore, battery pack are not accessible from the same direction as joining comes from different orientations [3]. Other than that, safety risk has always been one of the challenges due the short circuit, discharge failure, high voltage and chemicals contained in battery module [2, 3]. High product variance in the EV battery pack market and the non-existent standard of battery pack have further increased the difficulties of dismantling and recycling since most of the car manufacturers implement battery system according to their existing car design [5].

The cost of dismantling is even higher as detailed designs of each products are unavailable for the recycler causing lower effectiveness in disassembling [6]. Designing a standardized modular battery pack can be justified as it might be one of the ways to speed up the disassembling and recycling line and even implement fully automated disassembly. Even, research and public concerns have been raised over the years, disassembly and recycling path remains at laboratory level due to the complexity of LIBs [7]. Hence, it is essential to establish an eco-design of EV battery to ease the recycle process of spent LIBs.

2 Design of Lithium-Ion EV Battery

Generally, LIB packs used in mainstream EVs have a common set of internal components. Essentially, a battery pack consists of a battery case, which houses the battery modules connected to the battery management system (BMS), heating and cooling systems, and power electronics. Additionally, the battery case also holds various electrical wiring, fuses, and plugs. Meanwhile, the battery case itself consists of an insulated upper and lower housing, which maintains the structural integrity of the battery pack [8].

2.1 Structure of Lithium-Ion Battery (LIB)

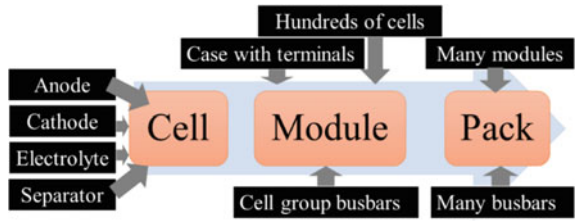
EV battery packs are generally designed and manufactured in a pack–module–cell structure [9]. The cell itself can be in the form of cylindrical or prismatic constructions depending on the design. Cell is the basic unit of LIB that converts the stored chemical energy into electrical energy by charging and discharging [10]. The set of single cells or groups forms battery modules, finite, and discrete sub-components, then aggregated through series connections into the final constructions called pack. With this structure, LIB can be disassembled into two stages, one from pack to modules level and one from modules to cells level.

2.2 Disassembly Issues in Battery Systems

There is a common agreement that the design of battery packs is not standard and varies depending on the car model [11, 12]. The variation of battery design is largely due to the requirement of car manufacturers to develop new designs that balance several aspects such as crash safety, center of gravity, space efficiency, and serviceability. Due to this concern, battery packs from different manufacturers are differ in size and structure for electronics arrangement. Therefore, the disassembly of battery packs will be inefficient and affecting the recycling of spent batteries [13]. This also results in a requirement for a wide range of tools for the removal of the fasteners and screws within the batteries. Figure 1 depicts the disassembly of each level and with elements obtained.

According to Gentilini [14], generic process of EV battery disassembly are removal of battery cover, service plug or safety fuse removal, coolant removal, junction block removal, Battery Management System (BMS) removal and lastly battery modules removal. Components in modules are detached to go for downstream process.

Fig. 1 Disassembling procedure of EV batteries in three possible levels as per the elements of each level [5]



3 Methodology

The methodology comprises of four phases which include case study, eco-design of EV battery pack, simulation and testing, and validation (Fig. 2).

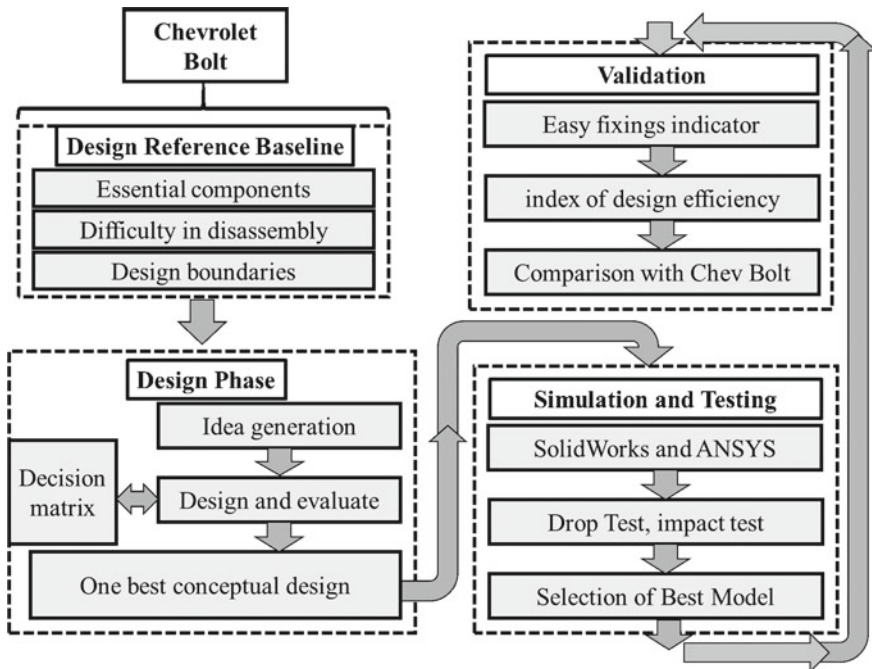


Fig. 2 Methodology comprises of four phases such as design reference baseline, design phase, simulation and testing, and validation adopted to design an eco-friendly battery pack

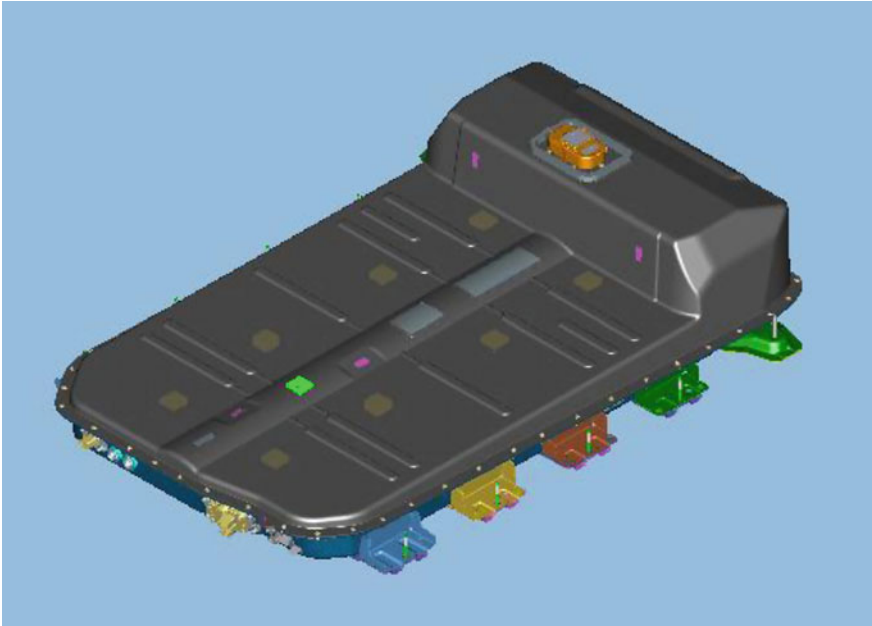


Fig. 3 The Chevrolet Bolt EV battery [15]

3.1 Chevrolet Bolt EV (CBEV) Battery as a Design Reference

In this study, the battery found in the Chevrolet Bolt EV (CBEV) is chosen as the case study for design reference baseline since this design has been declared as an excellent sample in battery development history [15]. The CBEV battery pack design is observed to understand the essential components, structure, and disassembly difficulties. In this paper, design boundaries are set where the study is focused on the battery pack level in which electronic connections and cooling system have been neglected. The battery pack basic requirements is evaluated for comparison with the proposed design on its ease the disassembly.

The battery pack of the CBEV is shown in Fig. 3 and Table 1 shows the battery performance of CBEV.

Table 1 CBEV battery performance [15]

CBEV Performance	
Range	>321,869 m
Battery Chemistry	Lithium Ion
Battery Nominal Energy	60 kWh
Battery Mass	436 kg
Max Battery Power	150 kW
Battery Nominal Voltage	350 V
Battery Position	Underfloor
Battery Cooling System	Liquid Active Thermal Control
Manual Service Disconnect Access Position	Interior

3.2 Design Phase of EV Battery Pack for Ease of Disassembly

3.2.1 Idea Generation

Idea generation is the first essential step to establish creative ideas. It emphasizes the quantities of ideas rather than quality to start the design process. After all the evaluation made on the reference battery pack from CBEV, the main notion is used to build up the initial battery pack design. In this stage, quantity of ideas is prioritized over the quality of it to boost as many ideas as possible. Then, the ideas consolidated later for further development.

3.2.2 Conceptual Design

Ideas generated in brainstorming stage are then compiled into individual designs and further assessed in terms of practicality, competence, strengths, and weaknesses. There are four possible design concepts which were generated. Subsequently, the designs were evaluated using a decision matrix which resulted in design concept 3 being chosen to be built in SolidWorks for further development. The final design has been established and brought forward to the next step which is simulation and testing to validate its functionality. The lithium-ion pouch cell has been constructed based on the Chevrolet Bolt EV which is a nickel-rich LIB cell.

3.3 Simulation and Testing

Simulation and testing of the design is done using SolidWorks and ANSYS while drop test has been done in SolidWorks software. In these tests, mechanical abuse is

prioritized in the battery abuse testing context, to validate the joining of the product and its performance. Drop test has been conducted as the battery abuse test. The battery pack has been dropped from a height of 0.3 m with 9.81 m/s^2 gravitational pull acceleration [9]. Furthermore, impact test has been carried out to test the crash worthiness of the battery pack using ANSYS software. A 15.6464 m/s front impact has been simulated on the battery pack to imitate the crash test on a vehicle. The battery pack will be simulated to run straight at a speed 15.6464 m/s into a solid concrete wall which duplicates a vehicle crash with another vehicle at the same speed and comparable weight [16].

3.4 Validation

Finally, the product has been validated with the Chevrolet Bolt. The main features such as “index of design efficiency” and “easy fixings indicator” have been compared and validated through Eq. 1 and Eq. 2 respectively. For calculating design efficiency, time taken to assemble a “theoretical” part has been assumed as 3 s. The Easy Fixings indicator examines the disassembly complexity in many ways to evaluate the percentage of difficulty in disassembly procedure for a product [11].

$$\text{Design efficiency} = \frac{3 \times \text{min no.of parts}}{\text{total assembly time}} \quad (1)$$

$$I_{f_f} = \frac{\sum_{i=1}^n A_{NDD}CNDD_i + A_{OD}COD_i + A_{TC}CTC_i + A_{TCO}CTCO_i + A_{QO}CQO_i + A_{ER}CER_i}{n} \quad (2)$$

where:

CNDD: complexity of disassembly directions,

COD: complexity of disassembly tool,

CTC: complexity of the type of contact,

CQO: complexity of the operators' qualification,

CER: complexity of require equipment,

n: number of links involved,

If: Easy fixings indicator (In%).

Furthermore, the proposed EV battery pack has been examined and compared with the baseline design, the Chevy Bolt battery pack (in Fig. 4) in terms of number of bolts and nuts, design efficiency, easy fixings indicators, disassembly time, number of casing for three cell group, and number of components for three cell group.

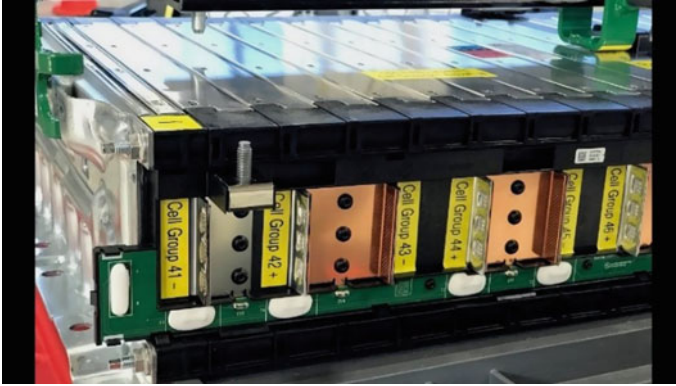


Fig. 4 Modular battery pack of the Chevy Bolt EV [17]

4 Results and Discussion

4.1 Design and Analysis

The exploded view of the final design of the battery pack in 3D model with its component has been represented in Fig. 5 with five main elements. Single cells of LIBs have been slotted into the battery pack case with separator alternatively slotted together. A snap-fit terminal connector with terminals attached has been fitted into the battery pack case to connect the electricity of each single cells. Three cells have

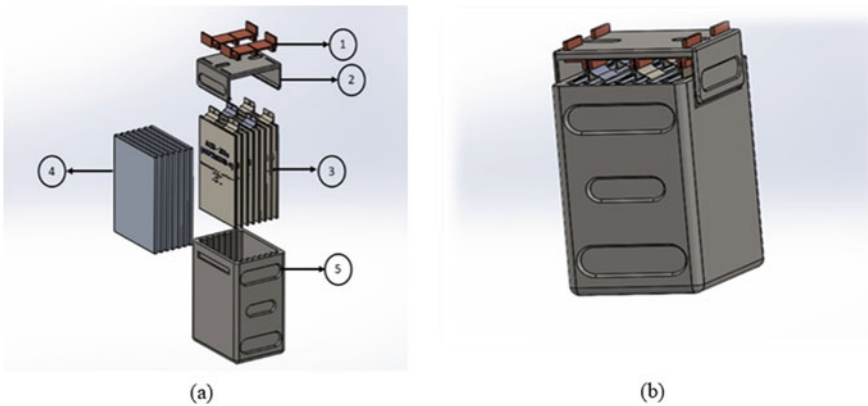


Fig. 5 **a** The exploded view of the proposed design of the battery modules in 3D model with its component. 1) terminal made of copper, 2) terminal connectors made of aluminum alloy, 3) lithium-ion single cell made of nickel and lithium, 4) separator made of aluminum, and 5) battery pack case made of aluminum alloy. **b** The assembly of the proposed battery pack design with zero nuts and bolts

been welded together as a single group cell. The proposed EV modular is composed of 9-cell battery pack having 44 kg weight. The entire battery pack comprises of 288 cells, which is equivalent to 96 cell groups that would give out a power range of 66 kWh.

4.2 Simulation and Testing

4.2.1 Drop Test

A drop test has been designed to investigate the structural integrity of the product. In this context, the battery pack has been dropped from a predefined height of 0.3 m to a surface with a gravitational force of 9.81 m/s^2 . The simulation result is the displacement yield obtained from the drop test which is represented in Fig. 6. The drop test follows the International Safe Transit Association (ISTA)'s A1 procedure where 0.3 m height is applicable to the product of weighing 44 kg [9]. It has been observed from Fig. 7, that the main displacement yield from the terminal connector where it yields to maximum is $1.015 \times 10^{-3} \text{ m}$. This indicates that the yielding is still within the safe zone where deformation is not too broad, and fracture does not occur during a drop. Besides that, it shows that the maximum von Mises stress generated in the element is $4.827 \times 10^8 \text{ N/m}^2$. This is relatively a great stress obtained however it is still under the red zone which shows the critical stage of stress. Thus, the stress generated from the drop test is acceptable.

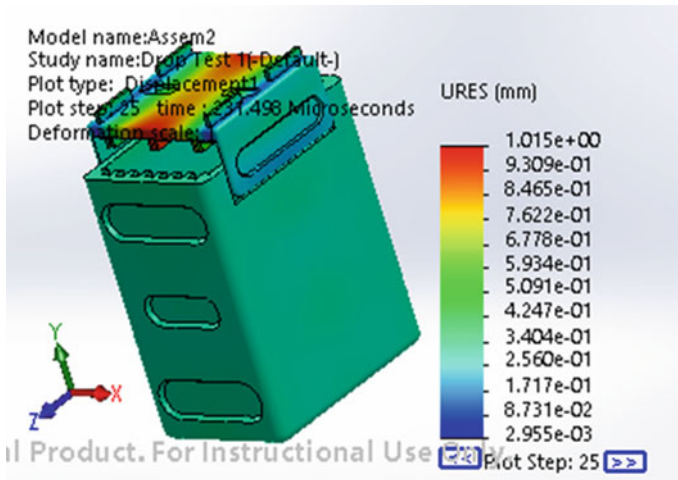


Fig. 6 ANSYS displacement yield simulation results via the drop test of designed battery pack

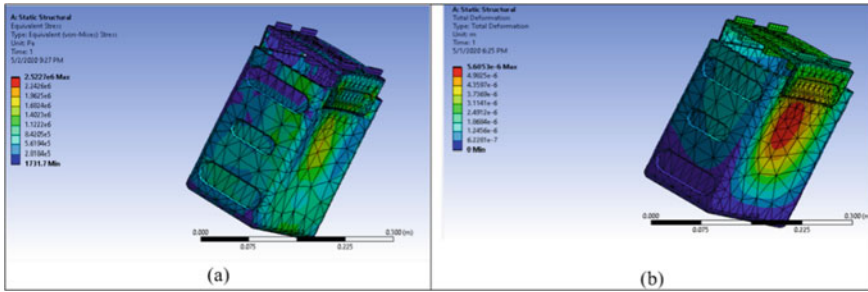


Fig. 7 **a** von Mises stress yield simulation results via impact test of designed battery pack using ANSYS software, **b** Displacement yield simulation results via impact test of designed battery pack using ANSYS software

4.2.2 Impact Test

ANSYS von Mises stress yield and displacement yield simulation results obtained through impact test of designed battery pack have been shown in Fig. 7(a) and (b) respectively. The impact test examines the amount of energy absorbed by a product during deformation. The absorbed energy is a measure of the product robustness. It also determines whether the product is brittle or ductile in nature.

A 15.6464 m/s front impact is performed on the battery pack as to imitate the crash test on a vehicle. However, the maximum von Mises stress generated on the battery pack is obtained as $2.5227e^6$ N/m². With great impact of cruising at a speed of 15.6464 m/s, $5.6053e^{-8}$ m deformation is obtained in Fig. 7(a). As seen in Fig. 7(b), deformation is caused however the level of robustness of the battery case is high as no obvious deformation is formed. This indicates that the battery pack case can protect the battery cells securely despite the huge frontal impact.

4.3 Comparison of Proposed Design with CBEV

Table 2 shows the evaluation and comparison between the Chevy Bolt EV and the

Table 2 Evaluation and comparison between the CBEV and the proposed design

CBEV	Comparison	Proposed design
12	No. of Bolts and Nuts	0
0.95	Design efficiency	1
35%	Easy Fixing Indicator	28%
110 s	Disassembly time	78 s
3	No. of casing for 3 cell group	1
35	No. of component for 3 cell group	26

final design. The number of components in Chevy Bolt EV is greater than the final design as it includes the number of bolts and nuts, casing, and total components. This also indicates that more time and technical effort is required when it comes to the disassembly procedure. Other than that, the design efficiency of the final design is 1. It is calculated from Eq. (1) where time of assembly and minimum number of parts are involved.

The assembly of the final modular battery pack design has been shown in Fig. 5(b). It shows that the proposed design performed minimal assembly time while also having minimal number of elements in the system. This is a significant difference as it reduced the number of bolts and nuts to zero while the Chevy Bolt EV has 12 of it [17]. Another reason the proposed design has an improved disassembly is that it greatly reduced the time taken to unscrew the bolts and nuts unlike in the case of Chevy Bolt EV.

Furthermore, the Easy Fixings Indicator (If) is used to examine the complexity of the product design. It evaluates the complexity of disassembly direction, disassembly tool, type of contact, operators' qualification, required equipment and the number of links involved. Therefore, the easy fixings indicator formulated for the final design is lower than the Chevy Bolt EV which shows that it is much less complicated in terms of design and many other aspects. This can greatly simplify the disassembly line of the EV battery waste. This is largely due to the parallel and identical disassembly direction of components designed in the proposed battery pack. Less disassembly tools are required due to the proposed design contains less components. The type of contact in the Chevy Bolt is considered as "many points contact" which has a scale of 5 compared to the proposed design has linear and surface contact of scale of 3 to 4. This may increase the complexity of Chevy Bolt EV battery pack due to the bolts and nuts contact type while the proposed design has a snap-fit contact between the terminal connector and the battery pack case. In the case of operators' qualification and required equipment, both designs have the same scale as no fire protection and air filtration are needed in the case. It is assumed that both have the same level of power range and discharge expertise needed in the disassembly procedure.

However, there are some drawbacks in the proposed design. In the final battery pack design, the terminal connector uses the snap-fit connection to link the battery cell connections. Lower pressure will be formed as no bolts and nuts exerted onto the tabs of the battery cells. This may lead to unstable connections between the cells. Other than that, by only using one casing to group three cell groups together, there are more separators used in the proposed design. Hence, this may increase the weight of the modular battery pack and act as a burden to the system. Thus, further study is needed to improvise this study's output.

5 Conclusion

The proposed EV battery pack has a design efficiency of 1 with Easy Fixings indicator of 28%, demonstrating its greatly improved aspects compared to the battery pack reference, the Chevy Bolt EV battery pack. In a 30 cm drop height test, it yields a maximum deformation of $1.015e^{-3}$ m and a generated von Mises stress of $4.827e^8$ N/m². Additionally, $2.5227e^6$ N/m² of von Mises stress is obtained in the impact frontal test. With great impact of cruising at a speed of 15.6464 m/s, $5.6053e^{-8}$ m deformation is obtained in the same test. The results show that the proposed design have the potential to improve the ease of disassembly and can thus improve the recyclability of batteries.

For the future development of the proposed design, modal vibration analysis can be done in ANSYS to evaluate its capability in absorbing vibration caused in the EV. Additionally, proposed materials to be tested in laboratory level to examine its suitability in the battery pack application and to obtain validation from battery pack manufacturers for its practicality of design and implementation.

References

1. Lallanilla M (2019) Greenhouse gases: causes, sources and environmental effects. Live science. <https://www.livescience.com/37821-greenhouse-gases.html>
2. Wegener K, Andrew S, Raatz A, Dröder K, Herrmann C (2014) Disassembly of electric vehicle batteries using the example of the Audi Q5 hybrid system. *Procedia CIRP* 23(C):155–160
3. Zhang J, Li B, Garg A, Liu Y (2018) A generic framework for recycling of battery module for electric vehicle by combining the mechanical and chemical procedures. *Int J Energy Res* 42(10):3390–3399
4. Johnson Matthey Battery Systems (2015) Guide to Batteries 3rd edition. Johnson Matthey Battery Systems
5. Coffin D, Horowitz J (2018) The supply chain for electric vehicle batteries, December 2018
6. Wegener K, Chen WH, Dietrich F, Dröder K, Kara S (2015) Robot assisted disassembly for the recycling of electric vehicle batteries. *Procedia CIRP* 29:716–721
7. Zheng X et al (2018) A mini-review on metal recycling from spent lithium ion batteries. *Engineering* 4(3):361–370
8. Wallentowitz H (2006) Unkonventionelle Fahrzeugantriebe, Institut für Kraftfahrzeuge, vol 142
9. Zwicker MFR, Moghadam M, Zhang W, Nielsen CV (2020) Automotive battery pack manufacturing - a review of battery to tab joining. *J Adv Joining Process* 1:100017
10. Bernardes AM, Espinosa DCR, Tenorio JAS (2004) Recycling batteries: a review of current processes and technologies. *J Power Sources* 130:291–298
11. Pehlken A, Albach S, Vogt T (2017) Is there a resource constraint related to lithium ion batteries in cars? *Int J Life Cycle Assess* 22(1):40–53
12. Wegener K, Andrew S, Raatz A, Dröder K, Herrmann C (2014) Disassembly of electric vehicle batteries using the example of the Audi Q5 hybrid system. *Procedia CIRP* 23:155–160
13. Elwert T, Römer F, Schneider K, Hua Q, Buchert M (2018). Behaviour of Lithium-Ion batteries in electric vehicles. Springer, Cham, pp 289–321. <https://doi.org/10.1007/978-3-319-69950-9>
14. Gentilini L, Mosalli E, Angilus A, Colledani M (2020) A safety oriented decision support tool for the manufacturing and recycling of post use H&EVs Lithium-ion batteries. *Procedia CIRP* 90:73–78

15. Liu J, Anwar M, Chiang P, Hawkins S et al (2016) Design of the Chevrolet Bolt EV propulsion system. *SAE Int J Altern Powertrains* 5(1):79–86. <https://doi.org/10.4271/2016-01-1153>
16. Steps to the Carton Drop Test. <https://www.intouch-quality.com/blog/5-steps-to-the-carton-drop-test>. Accessed 28 June 2021
17. Chevrolet Bolt EV Battery Reassembly (2018) WeberAuto. <https://www.youtube.com/watch?v=ZBzRKglr95U>

Electric Vehicle Drive Specification Modelling for Three Wheels Scooter Configuration



M. Faris, F. R. M. Romlay, A. R. Razali, M. R. Hanifah, and A. Ghazali

Abstract An electric vehicle is a vehicle that uses an electric motor for propulsion. Electric vehicles use fewer components that make electric vehicles low cost, easier for maintaining and very environmentally friendly as electric vehicles use no fuel. In east Asia, three wheels electric scooters have always become their choice because of their low cost and ease to perform maintenance. A major problem for three wheels electric scooter is the vehicle cannot reach the desired distance travelled because there is a problem when selecting motor and battery for three wheels electric scooters. This research study aims to construct the electric vehicle specifications design tools for motor and battery sizing, simulate and analyze the electric motor performance, and simulate and analyze drive system performance for three wheels electric scooter. For the electric vehicle specification design tool, DC motor sizing calculation is used to obtain the motor's power, speed, and torque based on some parameters required by user limitation. Simulation is performed using MATLAB R2020b Simulink to model and analyze the three wheels electric scooter's performance. BLDC motor with 800 W power, 1000 rpm speed and 12 Nm torque is chosen to simulate three wheels electric scooters. A lead-acid battery with 60 V, 12 Ah is chosen to simulate three wheels electric scooters. The effect of parameters such as motor's rated torque and power vehicle performance has been analyzed.

Keywords Electric vehicle · BLDC motor specification and motor torque

M. Faris · F. R. M. Romlay (✉) · A. R. Razali
Machine Manufacturing Union in Mechatronics Laboratory, Manufacturing Focus Group,
Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: fadhlur@ump.edu.my

M. R. Hanifah
Automotive Engineering Centre, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

A. Ghazali
Sapura Technical Centre, No 11, Jalan P/1, Section 13, Kawasan Perindustrian Bangi, 43650
Selangor, Bandar Baru Bangi, Malaysia

1 Introduction on Electric Vehicle

To this day, the automobile has become one of the biggest successes in the development of the internal combustion engine. This success leads to rapid growth in automotive industries through high demand in society [1]. However, the achievement will cause some poor results, especially to the environment. 24% of the world's total energy resource consumption is vehicle energy resource consumption [2].

The conventional petroleum-consuming automotive industry has already caused the economy to become deeply dependent on fossil oil supplies and has amplified the dispute between energy production and consumption. The issue of energy resources is becoming increasingly evident and dangerous as the keeping of vehicles continues to increase. High use of the conventional vehicle also leads to global warming and air pollution as internal combustion engines emitted CO₂ to the surrounding.

It is very difficult to solve this issue only by enhancing the performance of the engine. Developing new energy vehicles will become the driving force behind the growth of the car industry in the future. A solution has been made to solve these problems which are to develop electric vehicles and hybrid vehicles. This is a comparison between internal combustion engines, hybrid vehicles and electric vehicles [3].

2 Electric Vehicle Technology Review

In east Asia, people started to choose an electric vehicle as their public transports to go anywhere such as work, market and others. In Thailand, most of the public transports is rickshaws or 3-wheel electric scooters that can carry four people to their desired destination. 3-wheel electric scooters always become their choice because of their low cost and ease to perform maintenance.

Electric vehicle (EV) has a wide potential in maintenance and average saving cost. EV is powered by using an electric motor. Direct current and alternating current can be the electric motor. A different motor will be created a different result. Electric vehicles do not pollute the environment because EVs do not produce carbon dioxide (see Table 1).

However, the major problem of three wheels electric scooters was not fully efficient due to the various component's integration. The components selected are not suitable for three wheels electric scooter. This may cause the three wheels electric scooters cannot achieve the performance that is desired by the users such as in terms of the travel range and speed [5, 6].

Thus, the objectives of the study are established to construct the EV specifications design tools modeling for motor and battery sizing. Then, the tasks are preceded to simulate and analyze the electric motor drive system performance for three wheels electric scooter.

Table 1 Comparison between ICE vehicle, hybrid vehicle and electric vehicle

Parameters	ICE vehicles	Hybrid vehicles	Electric vehicle
Efficiency	Converts 20% of the energy stored in gasoline to the power the vehicle	Converts 40% of the energy stored in gasoline to the power the vehicle	Converts 75% of the chemical energy from batteries to power the vehicle
Speed (Average Top Speed)	199.5 km per hour (kmph)	177 km per hour (kmph)	48–153 km per hour (kmph)
Acceleration (average)	0–96.5 kmph in 8.4 s	0–96.5 kmph in 6–7 s	0–96.5 kmph in 4–6 s
Maintenance	High maintenance owing to more number of moving parts	Same as an ICE vehicle	Maintenance is minimal due to lesser number of moving parts
Mileage (average)	Can go over 480–500 kms before refueling. Typically achieves 10–12 kmpl	Typically achieves 20–25 kmpl	Can travel 120–200 kms before recharging
Cost (average)	INR 0.7–1.1 million	INR 1.2–2 million	INR 0.9–6 million

Source [4]

3 Electrical Vehicle Drive Specification

3.1 Electric Motor Specification Calculation

When choosing an electric vehicle motor factors must be taken into consideration to decide maximum power is needed. The elements are rolling, gradient, and aerodynamic drag force [7–9].

$$F_{total} = F_{rolling} + F_{gradient} + F_{aerodynamic\ drag} \tag{1}$$

where, F_{total} = Total force

$F_{rolling}$ = Rolling force

$F_{gradient}$ = Gradient force

$F_{aerodynamic\ drag}$ = Aerodynamic Drag force

The total force is a total tractive force that the output of the motor must overcome, to move the vehicle.

Rolling resistance is the resistance offered to the vehicle due to the contact of tires with the road as referring to Table 2 coefficient (Fig. 1).

$$F_{rolling} = C_{rr} \cdot M \cdot g \tag{2}$$

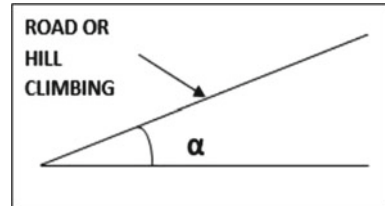
where, C_{rr} = coefficient of rolling resistance.

M = Mass in kg.

Table 2 Values of C_{rr} coefficient

Contact surface	C_{rr}
Concrete (good/fair/poor)	0.010/0.015/0.020
Asphalt (good/fair/poor)	0.012/0.017/0.022
Macadam (good/fair/poor)	0.015/0.022/0.037
Snow (2/4 in.)	0.025/0.037
Dirt (smooth/sandy)	0.025/0.037
Mud (firm/medium/soft)	0.037/0.090/0.150
Grass (firm/soft)	0.055/0.075
Sand (firm/soft/dune)	0.060/0.150/0.300

Fig. 1 The angle between the ground and the slope of a path



g = acceleration due to gravity.

$$F_{gradient} = \pm M \cdot g \cdot \sin \alpha \tag{3}$$

where, M = Mass in kg.

g = acceleration due to gravity.

α = angle between the ground and slope of the path.

$$F_{aerodynamic\ drag} = 0.5 \cdot C_d \cdot A \cdot \rho \cdot v^2 \tag{4}$$

where, C_d = Drag Coefficient.

A = Front Area of electric vehicle.

ρ = density of air.

v = velocity of electric vehicle.

After calculated the total traction force, the total force must be changed to power in watt by using this Eq. (5).

$$Power = F_{total} \cdot v \tag{5}$$

where, $Power$ = Power in watt.

F_{total} = Force total.

v = velocity of an electric vehicle.

From the value of power, many specifications of the motor can be determined such as RPM motor and torque motor. The equation below is used to determine the

motor specification [10].

$$\tau_{motor} = \frac{F_{total}}{g}r \tag{6}$$

- where τ_{motor} = Motor torque.
- F_{total} = Fotal force.
- g = Acceleration due to gravity.
- r = Radius tyre.

$$RPM_{motor} = \frac{Power}{\tau_{motor}} \cdot \frac{60}{2\pi} \tag{7}$$

3.2 The Three Wheels Electric Vehicle Specifications

The methodology to test and clarify the design and manufacturing ability of electric scooters is discussed. This research focuses on the motor transmission mechanism and considering loads that are carried. In the investigation, the effects of motor force transfers and responses are applied [11]. By using the computational ability software, the dynamic responses are calculated and performed. The geometry and weight information are tabled in Table 3.

The design requirements for the said electric vehicle scooter are for a maximum of 30 km/h. The study set a constraint of the maximum speed so that, the desire parameters are accommodated accordingly with the requirements.

The BLDC motor specifications are listed in Table 4 with 800 W power, 1000 rpm speed and 12 Nm torque is chosen for the simulation of 3 wheels electric scooters.

Lead-acid battery with 60 V, 12 Ah is chosen for the simulation of 3-wheel electric scooters as listed in Table 5.

Table 3 Three-wheel electric scooter specifications

Specification	Value
Size	1980 × 1000 × 880 mm
Weight	150 kg
Max load	300 kg
Tire size	18" × 9.5"

Table 4 BLDC motor specifications

Specification	Value
Power	800 W
Max speed	1000 rpm
Max torque	12 Nm

Table 5 Lead-acid battery specifications

Specification	Value
Voltage	60 V
Capacity	12 Ah

These data can be used as the design and manufacturing references of corresponding manufacturers. The modeling data and analysis can be applied as the key parameters for the electric scooter to perform within the allowable dynamic design range.

4 Electric Motor Specification Calculator

For the calculation of motor specifications, Microsoft Excel is used to make the calculation easier. The formula is programmed in Visual Basic to obtain the motor specification such as motor power, torque, and speed motor (Figs. 2 and 3).

It was required input was feed to the calculator and able to calculate the motor power, motor torque and speed for electric vehicle configuration performance prediction.

```
Private Sub CommandButton1_Click()  
Fr = crr.Value * m.Value * 9.81  
Fg = m.Value * 9.81 * Sin((alpha.Value * (3.142 / 180)))  
Fa = 0.5 * cd.Value * ad.Value * (air.Value) * ((v.Value * (1000 / 3600)) * (v.Value * (1000 / 3600)))  
Ft = Fr + Fg + Fa  
Pt = (Ft * (v.Value * (1000 / 3600))) / 0.85  
result.Value = Pt  
result.Value = Format(Val(result.Value), "0.00")  
  
Ac = 2 * 3.142 * rt.Value  
RPMw = (v.Value * (1000 / 60)) / Ac  
RPMmotor = (g1.Value / g2.Value) * RPMw  
rpm.Value = RPMmotor  
rpm.Value = Format(Val(rpm.Value), "0.00")  
  
Torquemotor = 9.5488 * (Pt / RPMmotor)  
tmotor.Value = Torquemotor  
tmotor.Value = Format(Val(tmotor.Value), "0.00")
```

Fig. 2 Coding for the electric motor calculator

Calculator Electric Motor For Electric Vehicle

Mass	<input type="text"/>	kg	Tyre Radius	<input type="text"/>	m
Coefficient of rolling resistance	<input type="text"/>		Driven Gear Teeth	<input type="text"/>	
Angle road	<input type="text"/>	°	Drive Gear Teeth	<input type="text"/>	
Velocity	<input type="text"/>	km/h	Time To Reach Velocity	<input type="text"/>	s
Coefficient of Drag	<input type="text"/>				
Front area of vehicle	<input type="text"/>	m ²			
Air Density	<input type="text"/>	kg/m ³			

Power of Motor	<input type="text"/>	W	Speed	<input type="text"/>	RPM
Torque Motor	<input type="text"/>	N.m			

Fig. 3 Graphical user interfaces the electric motor calculator

5 Specification Calculations and Simulation Results

The three-wheel electric vehicle scooter was configured as a prototype for parametric study and evaluation. The configuration structure is shown in Fig. 4. It is a single wheel at the front and rear double wheel with three-seater capacity.

The specifications were obtained from the calculation and simulation approaches that suit the actuator capacities.



Fig. 4 Electric vehicle structure fabrication and configuration

The motor power, speed and torque were obtained after the required parameters were applied in the calculator in Fig. 5.

When calculating the motor power, the power must be closed to the ideal calculation. Since the desired value of motor power was 800 W, the closed motor power was 748.11 W at the motor speed value was 742.63 rpm as in Table 6. The rated speed was 674.21 rpm as in Table 7.

The simulation was conducted in MATLAB R2020b Simulink. All the results were conducted in 1000 s sample time. From the simulation, the torque produced at 852.59 rpm was 9.62 Nm while the maximum torque was 73.47 Nm as starting torque that shown in Fig. 6 and Table 8.

The maximum velocity of three wheels electric scooter was 13 km/hr. In 1000 s, the 3-wheel electric scooter can travel to 3.3 km as depicted from Fig. 7 and Fig. 8.

Fig. 5 The result after the parameters were applied in the electric motor calculator

Table 6 Parameters of motor specification for the 3-wheel electric motor

Parameters	Value
Power of motor	800 W
Speed of motor	742.63 rpm
Torque of motor	9.62 Nm

Table 7 Rated and No-load speed for BLDC motor

Parameter	Value (rpm)
No-load speed	852.59
Rated speed	674.21

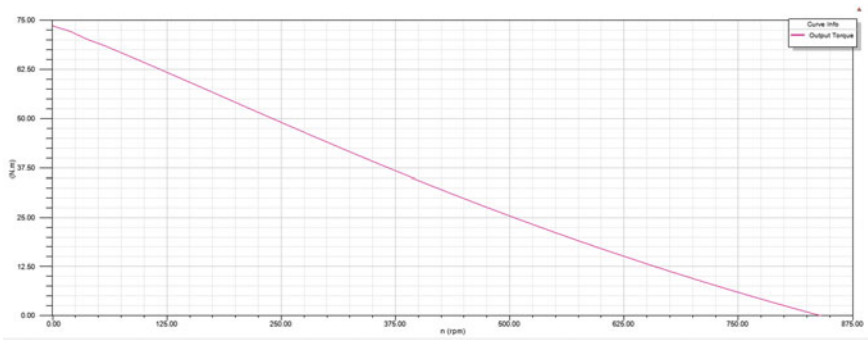


Fig. 6 Output torque vs speed

Table 8 Torque parameter of BLDC motor

Parameter	Value (Nm)
Rated torque	11.34
Cogging torque	679.15
Starting torque	73.47

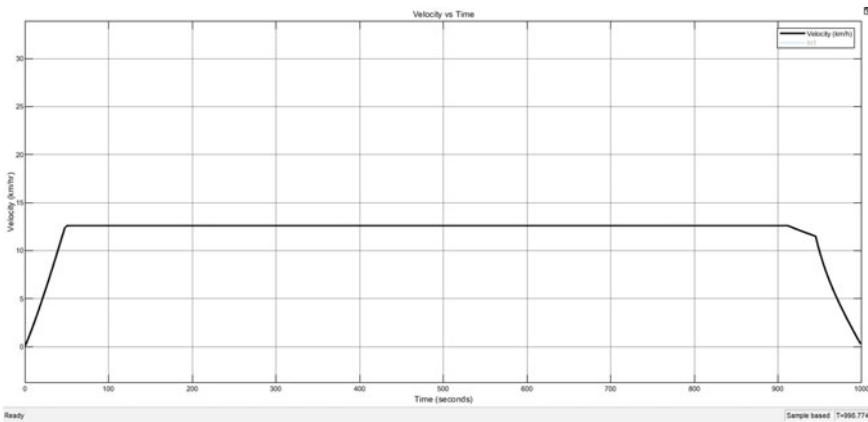


Fig. 7 Velocity vs time graph

These were the simulation outputs of the 800 W brushless DC motor designed based on the manufacturer rating specifications. The study provided a general performance of the motor specifications that suit the three wheels electric scooter requirements.

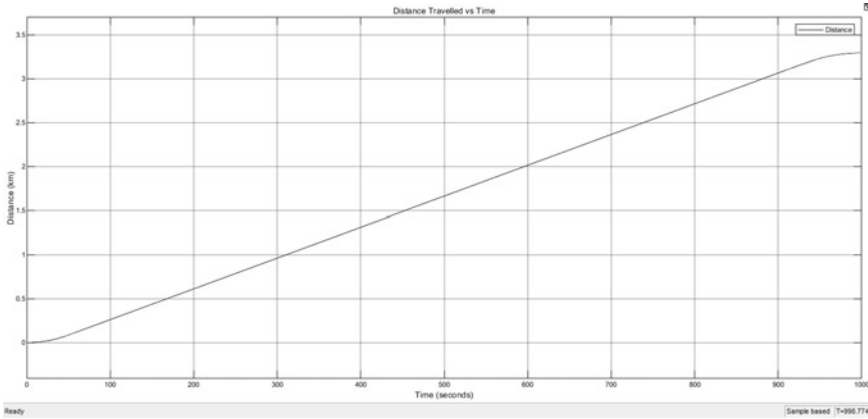


Fig. 8 Distance travelled vs time graph

6 Conclusions and Recommendations

A calculation tool is constructed for the electric vehicle specification. BLDC motor is selected for the 450 kg of the three wheels electric scooter with motor specifications which are 800 W, 1000 rpm and 10 Nm. From the BLDC motor specification, the simulation is carried to obtain the parameters required to operate the electric scooter. The simulation output result shows that the speed, voltage, power and torque parameters obtained also meets the necessary characteristics that are required for 450 kg of three wheels electric scooter. Based on the parameters, the electric scooter was modeled in MATLAB R2020b Simulink. Based on the model, a step-by-step approach was shown to develop various simulation model blocks such as the aerodynamic model, gearbox, motor model and battery model. The simulation result in term of velocity, range, state of charge, voltage and current are presented and analyzed. This paper recommended further studies in the design of a drive system for three wheels electric scooter doing in the experimental method to make the result more valid. This is because the velocity of wind and the road inclination will affect the performance of the electric scooter.

Acknowledgements The authors would gratefully thank Universiti Malaysia Pahang for the financial support through Flagship Grant, RDU172206. Also thank team members who gave high commitment and effort to the project activities.

References

1. Chauhan S (2015) Motor torque calculations for electric vehicles. *Int J Sci Technol Res* 4(8):126–127

2. Guirong Z, Henghai Z, Houyu L (2011) The driving control of pure electric vehicle. *Procedia Environ Sci* 10:433–438
3. Shrivastava K, Pawar MD (2016) A review on types of DC motors and the necessity of starter for its speed regulation. *Int J Adv Res Comput Commun Eng* 5(4):61–63
4. Kumar A, Choudhary SK, Chethan KN (2018) Commercial viability of electric vehicles in India. *Int J Mech Eng Technol* 9(6):730–745
5. Lee JY, Woo BC, Kim JM, Oh HS (2017) In-wheel motor design for an electric scooter. *J Electr Eng Technol* 12(6):2307–2316
6. Mahmoudi C, Flah A, Sbita L (2014) An overview of electric vehicle concept and power management strategies. In: 2014 international conference on electrical sciences and technologies in Maghreb (CISTEM), pp 1–8. IEEE, November 2019
7. Pant N, Racherla V (2019) Thermal management of battery pack for hybrid vehicles using PCM. *IOP Conf Ser Mater Sci Eng* 653(1):012012. <https://doi.org/10.1088/1757-899X/653/1/012012>
8. Porselvi T, Sriharharan MK, Ashok J, Kumar SA (2017) Selection of power rating of an electric motor for electric vehicles. *Int J Eng Sci Comput IJESC* 7(4)
9. Kaushik S (2019) Modelling and simulation of electric vehicle to optimize its cost and range. *Int J Eng Adv Technol* 8(6):415
10. Gu J, Ouyang M, Li J (2010) Vehicle dynamic simulation for efficiency optimization of four-wheel independent driven electric vehicle. *World Electr Veh J* 4(2):319–324
11. Sharmila B (2021). Modelling and performance analysis of electric vehicle. *Int J Ambient Energy* (just accepted) 1–14

Experimentation on Spectra Data Regression Using Dense Multilayer Neural Networks with Common Pre-processing



Kok Pin Chan, Mahmud Iwan Solihin, Chun Kit Ang, and Liew Phing Pui

Abstract Normally preprocess the raw data of near infrared (NIR) spectroscopy is indispensable in the spectroscopy data calibration to reveal useful information from the objects. It will help to remove the significant noises and other undesirable factors to get the more accurate calibration model. However, NIR spectra preprocessing in multivariate calibration is often problematic due to a lack of appropriate prior information and incomplete understanding of the raw data. It depends largely on both the nature of the data and the expertise and experience of the practitioners. This could limit the applications since researchers are unfamiliar with the properties of numerous preprocessing. In this paper, the results of experimentation on NIR spectroscopy calibration (regression) using dense multilayer perceptron neural networks (MLPNN) are presented. This is to investigate whether dense MLPNN can perform robustly for calibration of different datasets with only simple common pre-processing stage. Five NIR spectroscopy datasets are used and pre-processed with common individual pre-processing methods such as Gaussian Smoothing, Savitzky-Golay filter, Min-max Normalization and Extended Multiplicative Signal Correction (EMSC). Those datasets are used to train the dense MLPNN calibration model with different architecture in terms of number of different hidden layer neurons. The results indicates that there is unified pre-processing method that work robustly for all datasets even when dense layer neural networks are used. This could lead to recommend on the use of pre-processing less-affected calibration method such as deep learning where pre-processing effort can be minimized.

Keywords NIR spectroscopy · Intelligent chemometrics · Dense multi layers neural networks · Spectroscopy calibration

K. P. Chan · M. I. Solihin (✉) · C. K. Ang
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, Malaysia
e-mail: mahmudis@ucsiuniversity.edu.my

M. I. Solihin · L. P. Pui
Faculty of Applied Sciences, UCSI University, Kuala Lumpur, Malaysia

1 Introduction

Near infrared (NIR) spectroscopy has been widely used in many fields of applications as a non-destructive method for material fingerprinting such as in food and agriculture [1]–[5], medical application [6], manufacturing [7], etc. The study of the interaction between matter and electromagnetic radiation is known as spectroscopy. It began with the study of visible light scattered by a prism according to its wavelength. NIR spectroscopy is a non-destructive (non-invasive) and quick way of evaluating a substance’s quantitative and qualitative properties. The energy of chemical bonds containing hydrogen (CH, NH, OH, SH) will vary when a sample is exposed to NIR radiation. Due to NIR radiation absorbed by hydrogen from different organic bonds for the analyzed sample with very complex composition. Lambert-Beer’s law states the linear relationship of the absorptive capacity of a dissolved substance is directly proportional to concentration of the solution (c), where l is the length of the optical path and ε is the extinction coefficient (absorption coefficient). It can be expressed as shown in Eq. 1 [8][15], where I_o is the intensity of the initial light, I is the intensity of transmitted light.

$$A = \log_{10}\left(\frac{I_o}{I}\right) = \varepsilon lc \quad (1)$$

Machine learning is often used as calibration tool for spectroscopy data in tandem with advance statistical learning [9]. However, pre-processing is crucial stage for the calibration model to perform well. All signals obtained in analytical instruments are known to be impacted by noise. Interfering physical and/or chemical factors, for example, could cause NIR spectral noises as well as imperfections in the experimental apparatus and/or other random factors. Significant noise would reduce the signal-to-noise ratio (SNR) and resolution of the spectra, compromising the calibration model’s accuracy and precision. Complex backgrounds and baselines, which contribute non-concentration-correlated contributions to spectral data, are also common problems in NIR spectra. Such undesirable spectrum variations would result in an overly complicated calibration model, e.g., a low-accuracy latent-variable model with too many latent components, but it also introduces biases into the model [10].

It is of great importance to pre-process NIR spectra properly for removal of noises, backgrounds, and baselines. Many pre-processing methods exist for transformation of NIR spectra, the most frequently used methods include Gaussian smoothing [11], Savitzky-Golay filter (Derivatives and polynomial order) [12], Normalize spectra (Min–Max normalization) [13] and Extended multiplicative signal correction (EMSC) [14].

Therefore, in this study, we tested the potential of dense multilayer perceptron neural networks (MLPNN) for NIR spectroscopy regression application. This dense MLP is considerably a deeper learning algorithm despite the deep learning term often refers to common algorithm such CNN (convolutional neural networks).

Multilayer perceptron, also called feedforward neural networks, is the most utilized model in neural network applications trained using the back-propagation training algorithm. The definition of architecture in MLP networks is important because a shortage of connections can prevent the network from solving the problem of insufficient customizable parameters, while an excess of connections can cause the training data to be over-fitted [16].

One of the major issues in this research area is optimizing the number of connections and hidden layers for building a multilayer perceptron to solve the problem. The input layer, output layer, and hidden layers between these two layers make up a Multilayer Perceptron. The number of these layers is determined by the complexity of the task [17]. This paper presents the results of experimentation on different MLPNN for calibration, i.e., regression method, for various NIR spectra datasets. Different pre-processing methods of NIR spectra data are also applied in this research. The goal is therefore to observe whether dense MLPNN can perform robustly as calibration method with only simple common pre-processing stage. In the future, a reliable robust calibration model for NIR spectra data will be built based on Deep Learning algorithms with an expectedly trivial pre-processing stage. This experimentation is necessary as normally pre-processing of raw spectra data is the most tedious work when traditional machine learning and traditional statistical tools are implemented for calibration.

2 Calibration Method and Data Collection

Two NIR spectra datasets used were collected in previous studies by the authors and another three datasets were compiled from online repository. These five data sets are namely blood glucose (BG) level, mangoes brix level, dry matter content (DMC) within olive fruit, active substance in a pharmaceutical tablet and CGL (casein, glucose, lactate) of grain protein dataset. The detail about the datasets used are as follows:

The first dataset deals with the regression of the BG level (3.1 to 8.4 mmol/L) of participants. Total of 90 samples of NIR spectra and BG levels were collected from 45 participants. NIR spectra data were collected using micro NIR spectroscopic instrument with wavelength range from 900-1700 nm. The data are described in more detail in the study in [18].

The second dataset deals with the regression of the brix level (9.3 to 18.1%) of the mangoes. Three different type of mango fruits were selected namely Chokonan, Rainbow, and Kai Te. Total of 60 samples were scanned by using handheld NIR spectrometer with a wavelength range of 900 to 1700 nm. The data are described in more detailed in the study in [19].

The third dataset deals with the regression of the dry matter content (%) of intact olive fruit. All fruits were at a range of maturities at the time of scanning. Each side of each fruit was scanned twice, providing two spectra per reference analysis. There were 494 samples collected, as described in the study in [20].

Table 1 Summary of the NIR spectroscopy datasets used in this study

Dataset	Target variable	Wavelength range	Number of samples	Reference
BG level	Blood glucose	900–1700 nm	90	[18]
Mango data	Brix level	900–1700 nm	80	[19]
Olive fruit	Dry matter content (%)	700–1200 nm	494	[20]
Pharmaceutical tablet	Quantitation of the active substance (% per tablet)	7400–10,500 cm^{-1} (~950 nm to 1350 nm)	310	[21]
CGL data	Moisture (wt%)	1100–2500 nm	231	[22]

The fourth dataset deals with the regression of the chemometric quantitation of the active substance in a pharmaceutical tablet (% per tablet). Four different dosage values of this pharmaceutical drug (5, 10, 15, and 20 mg per tablet) were used. In total, 31 batches were used, and from each batch 10 tablets were individually weighed and analyzed. The data are described in more detail in the study in [21].

The fifth dataset deals with regression of the moisture (wt%) from NIR instrument with 231 samples created by Tormod Naes and Tomas Isaakson. As described on the website [22]. The summary of the datasets and their attributes is shown in Table 1.

Normally, the stage of data pre-processing process is crucial in the initial stage of the machine learning model building. This process can significantly affect the prediction accuracy in any type of data [23]. The overall data pre-processing methods used as follow:

- **Gaussian Smoothing:** Gaussian filtering is used to remove noise of the data. Data smoothing is based on the assumption that one is monitoring a variable that is both slowly fluctuating and contaminated by random noise. Then it's occasionally a good idea to replace each data point with a local average of nearby data points. Because close points measure approximately the same underlying value, averaging can minimize noise without (much) biasing the result.
- **Savitzky-Golay Filter:** Typically, a Savitzky-Golay smoothing filter is used to “smooth out” a noisy signal with a large frequency span (without noise). They are also called digital smoothing polynomial filters or least-squares smoothing filters. In some instances, Savitzky-Golay filters outperform ordinary averaging FIR filters, which tend to filter high-frequency data together with noise. High-frequency signal components are better preserved with Savitzky-Golay filters, but noise rejection is less effective.
- **Min–Max Normalization:** One of the most prevalent methods of data normalization is min–max normalization. The minimum value of each feature is converted into a decimal between 0 and 1 for each feature.
- **Extended Multiplicative Signal Correction:** Extended Multiplicative Signal Correction (EMSC) is a model-based preprocessing system that corrects for both undesired additive and multiplicative effects in data. It was first published in 1991 [24]. The EMSC is adaptable in that it can incorporate a priori knowledge

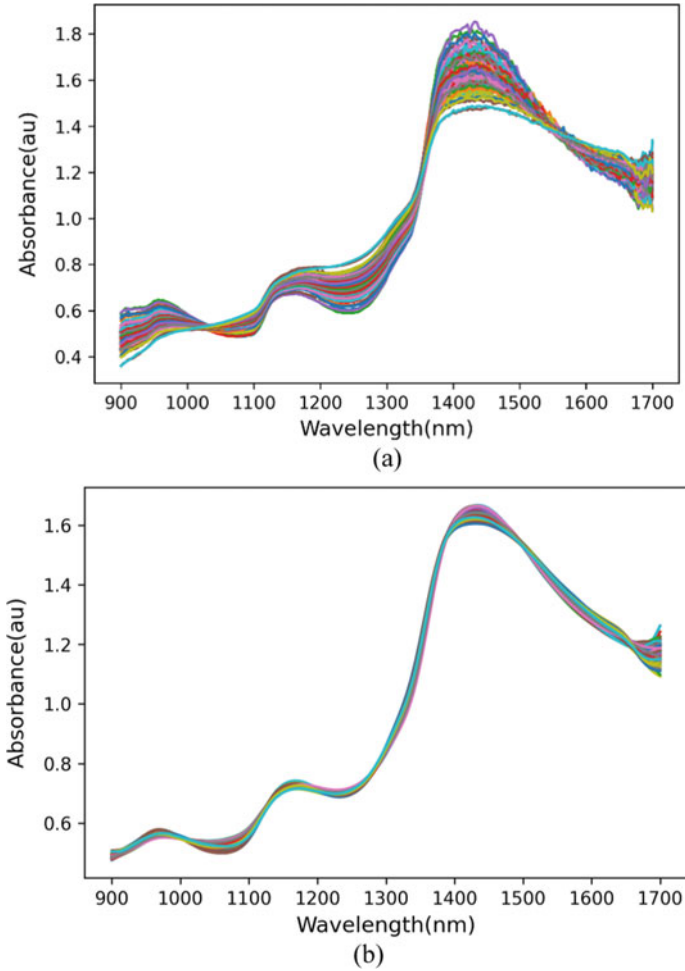


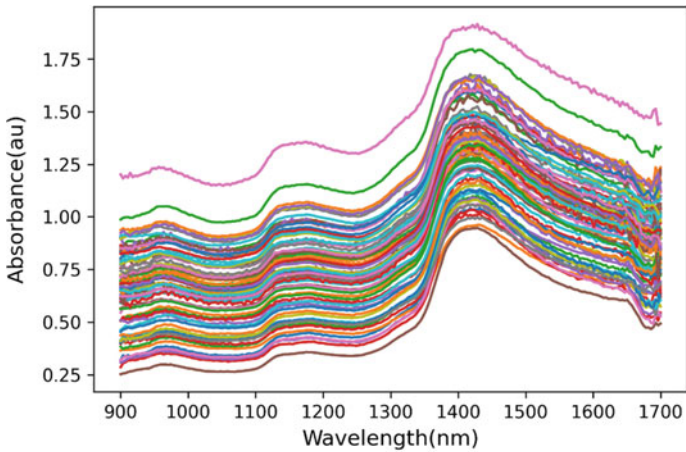
Fig. 1 NIR spectral of BG Level dataset: **a** Raw, **b** Preprocessed with method 4

of chemical and non-chemical patterns into the preprocessing model to improve data quality.

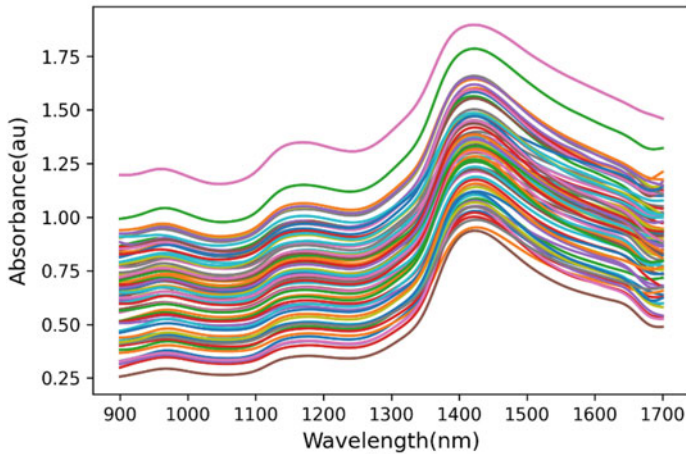
In this paper, those five datasets have been pre-processed by using four individual common methods. They are Gaussian smoothing, Savitzky-Golay filter, normalize spectra and extended multiplicative signal correction (EMSC) methods. Therefore, five sets of spectrum were generated for each datasets as raw data and four sets of pre-processed data with pre-processing methods number 1, 2, 3 and 4 (in Table 2). Sample spectrum of the datasets are shown on Figs. 1, 2, 3, 4, 5. Noted that in the respective Fig. 1(b) to Fig. 5(b), the pre-processing method used are the one

Table 2 Parameter of the pre-processing methods used in this paper

Number of pre-process	Gaussian smoothing	Savitzky-Golay filter		Normalize spectra	EMSC
	Standard deviation	Derivative order	Polynomial order	Min-Max normalization	Polynomial order
1	4	–	–	–	–
2	–	2	2	–	–
3	–	–	–	n/a	–
4	–	–	–	–	2



(a)



(b)

Fig. 2 NIR spectral of Mango dataset: **a** Raw, **b** Preprocessed with method 1

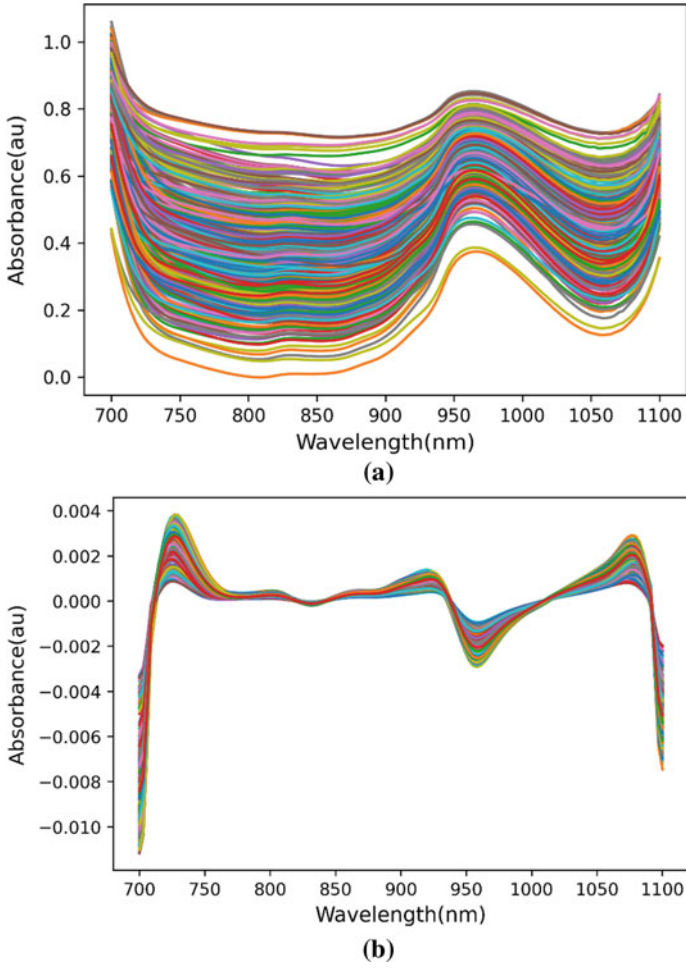


Fig. 3 NIR spectral of Olive Fruit dataset: **a** Raw, **b** Preprocessed with method 2

produces the best performance in the respective experimentation. The parameter of the pre-processing methods is shown in Table 2.

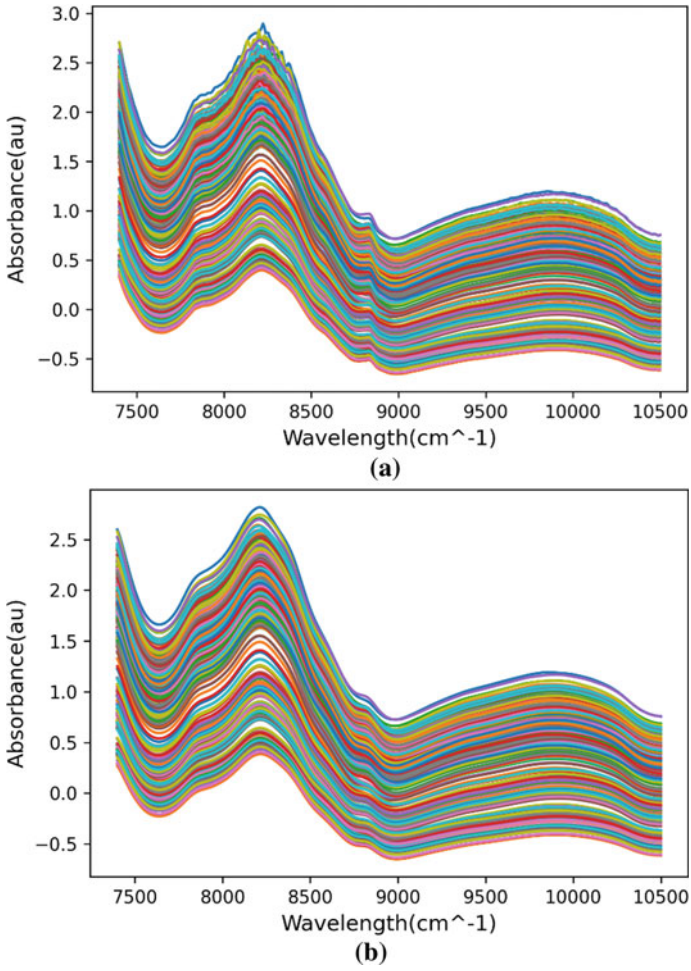


Fig. 4 NIR spectral of Pharma Tablet dataset: **a** Raw, **b** Preprocessed with method 1

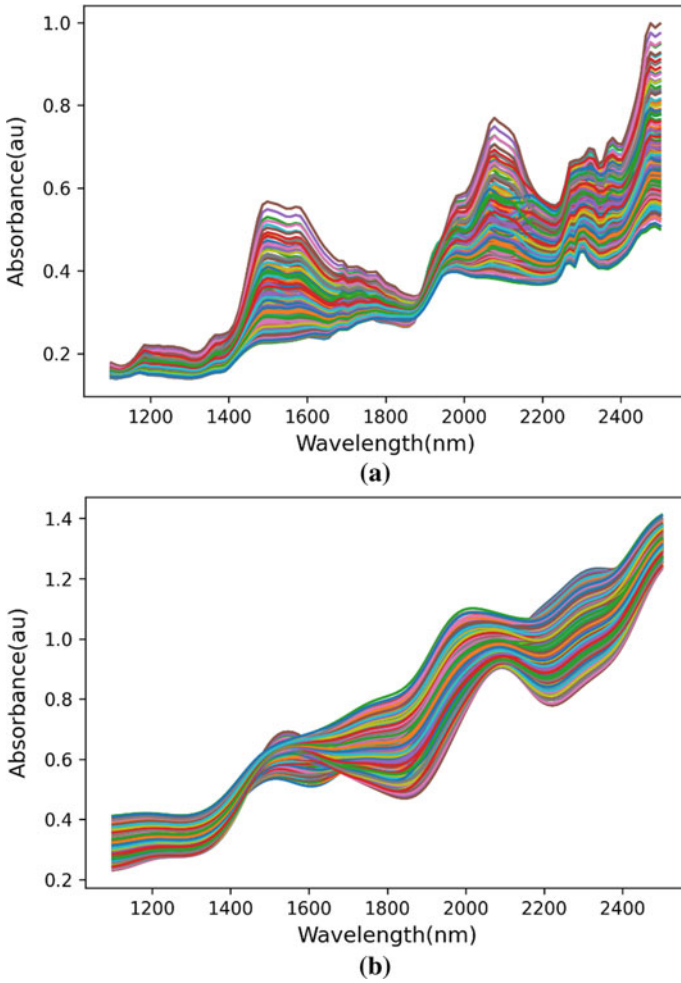


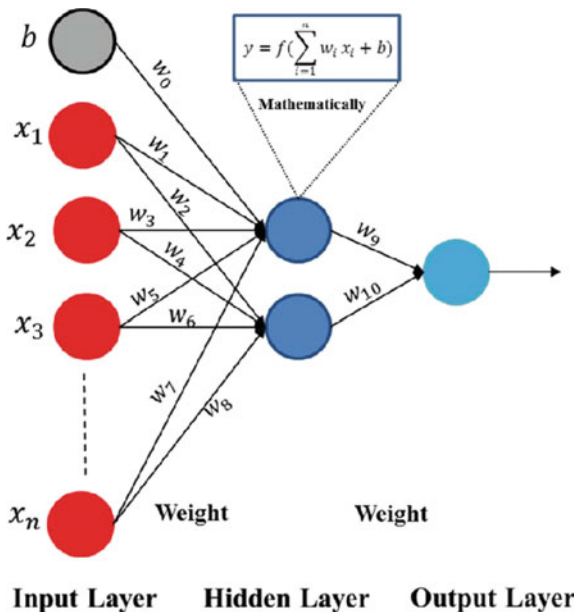
Fig. 5 NIR spectral of CGL dataset: **a** Raw, **b** Preprocessed with method 3

3 Neural Networks for Regression

A multilayer perceptron is a variant of the original perceptron model proposed by Rosenblatt in the 1958 [25]. It has one or more hidden layers between its input and output layers, the neurons are organized in layers, the connections are always directed from upper layers to lower layers, the neurons in the same layer are not interconnected see Fig. 6.

The neural number in the input layer equal to the number of measurement for the predictor variables (x_1, x_2, \dots, x_n) and the neuron number in the output layer equal to the number of output (y_1) , for the choice of hidden layers number (h_1, h_2, \dots, h_n)

Fig. 6 Multilayer perceptron neural network structure



and neurons in each layers and connections called architecture [26]. The neural weights (w_i) and bias (b_i) are the parameters that need to be optimized during MLPNN training to fit the calibration/training dataset.

In this experiment, those datasets have different feature numbers i.e.: 228, 228, 136, 404 and 117. Thus, five hidden layer network structure with 100, 250 and 500 number of neurons (node) are used in the multilayer perceptron (MLP) neural network. For the input layer the number of nodes is equal to the number of features and 1 node are used as output layer. ReLu (rectified linear unit) activation functions are used in hidden layers. Adam optimizer is used in the training process. Number of nodes in hidden layer are shown in Table 3.

Loss or error function is a measure of how good a prediction model does in terms of being able to predict the expected outcome. In this experiment, mean square error (MSE) have been chosen as loss function. Mean Square Error is the most commonly loss function used in regression model [27]. Coefficient of determination (R^2) are

Table 3 Number of hidden neurons in MLPNN

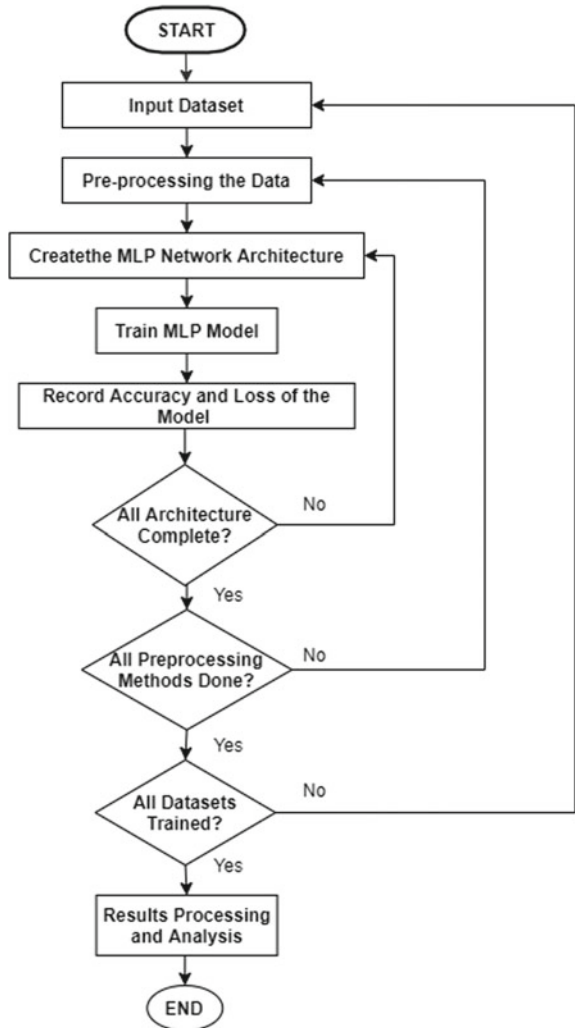
Number of architecture	Number of neurons (node)				
	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Hidden Layer 4	Hidden Layer 5
1	100	100	100	100	100
2	250	250	250	250	250
3	500	500	500	500	500

used to show the regression accuracy in this experiment. The R^2 value is calculated using:

$$R^2 = 1 - \frac{\sum_n (y_n - \hat{y}_n)}{\sum_n (y_n - \bar{y}_n)} \tag{2}$$

The flowchart of the methodology in this study is shown in the Fig. 7.

Fig. 7 Flowchart of methodology



4 Results and Discussion

Total 25 sets of data (from 5 NIR spectra datasets) were used to train different architectures of MLPNN to investigate whether simple individual preprocessing methods can produce robust calibration result for all datasets. The calibration accuracy is evaluated by using coefficient of determination (R^2) as it is regression problem. The experiment was performed using python programming modules.

Each dataset is divided into two subsets where 80% data are used for calibration (training) and 20% are used to test the model accuracy. All the MLPNN models were trained with 500 and 1000 epochs. The mean squared error (MSE) progress against the epoch graph is checked to make sure the convergence and no underfitting or overfitting. Sample of this graph is shown in Fig. 8 for the BG level dataset case with architecture 2 of MLPNN. Showing this graph for all cases of experiment is too lengthy.

For BG level dataset, pre-processed dataset with method 4 has presented the best result with $R^2 = 0.999$ for all three MLPNN architectures as compared to other methods and that without pre-process (raw dataset), as shown in Table 4. Pre-processing method 4 has also presented a good result ($R^2 > 0.91$) in Pharmaceutical Tablet dataset but it gives poor result in Mango dataset and CGL dataset.

For Mango dataset, pre-processed dataset with method 1 has presented the best result with $R^2 = 0.9484$ for architecture 3 with 1000 epochs, as shown in Table 5. In this dataset, pre-processing method 1 has very slightly improved the accuracy as compared to the raw dataset. However, it gives poor result especially in Olive Fruit dataset.

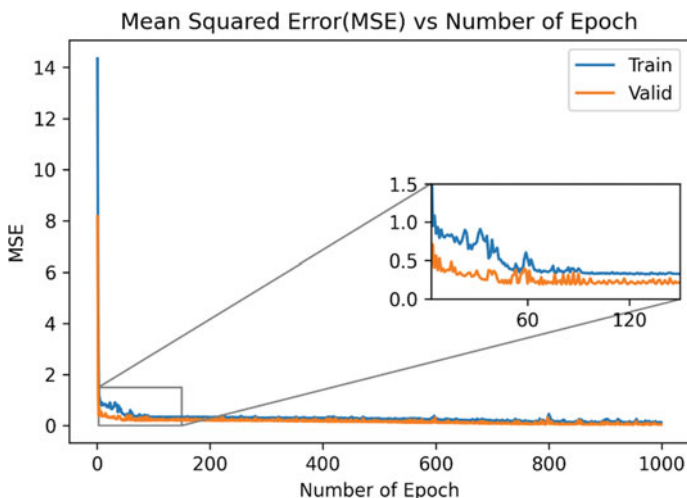


Fig. 8 Sample: MSE versus epoch graph of raw BG Level dataset with architecture 2

Table 4 Testing results of BG Level datasets

No. of Architecture	No. of Epoch	R^2 result with different pre-processing methods				
		Raw	Method 1	Method 2	Method 3	Method 4
1	500	0.5975	0.5815	0.1339	0.4988	0.9996
2	500	0.6574	0.6221	0.9276	0.5371	0.9998
3	500	0.7039	0.6667	0.9324	0.6038	0.9999
1	1000	0.7113	0.6892	0.7015	0.6062	0.9996
2	1000	0.8905	0.8762	0.9446	0.8535	0.9999
3	1000	0.9101	0.889	0.9458	0.8886	0.9997

Table 5 Testing results of Mango datasets

No. of Architecture	No. of Epoch	R^2 result with different pre-processing methods				
		Raw	Method 1	Method 2	Method 3	Method 4
1	500	0.7655	0.7647	0.0029	0.7673	-0.1254
2	500	0.7869	0.7951	0.1318	0.8126	-0.1469
3	500	0.8248	0.8343	0.3109	0.8648	-0.1812
1	1000	0.84	0.8294	0.1387	0.7937	-0.0516
2	1000	0.9291	0.9348	0.3595	0.916	0.0714
3	1000	0.9479	0.9484	0.6702	0.8427	0.2875

In Table 6, pre-processed dataset with method 2 is the only one that produces the better accuracy (R^2) as compared to raw dataset. Testing result of pre-processed dataset with method 1, 3 and 4 are worse than raw dataset especially pre-processed dataset with method 4. The best result of the pre-processed dataset with method 4 is only resulting $R^2=0.3583$ while the worst result of calibration using raw dataset is $R^2=0.4068$.

Furthermore, the testing result of Pharmaceutical Tablet dataset are shown in Table 7. It presents an unexpected result, but it can be, that the calibration using the raw dataset yields the best result as compared to others. However, they have not much difference and all the results show that R^2 is higher than 0.91.

Table 6 Testing results of Olive Fruit datasets

No. of Architecture	No. of Epoch	R^2 result with different pre-processing methods				
		Raw	Method 1	Method 2	Method 3	Method 4
1	500	0.4068	0.3545	0.4571	0.2473	0.1036
2	500	0.5643	0.4783	0.8988	0.2632	0.2241
3	500	0.5434	0.3856	0.9088	0.3734	0.1939
1	1000	0.4953	0.3893	0.842	0.3372	0.2163
2	1000	0.6559	0.5805	0.9079	0.516	0.3104
3	1000	0.6852	0.5038	0.9146	0.5771	0.3583

Table 7 Testing results of Pharmaceutical Tablet dataset

No. of Architecture	No. of Epoch	R^2 result with different pre-processing methods				
		Raw	Method 1	Method 2	Method 3	Method 4
1	500	0.9261	0.925	0.9263	0.9256	0.9301
2	500	0.9257	0.928	0.9248	0.9397	0.9242
3	500	0.9327	0.9292	0.9338	0.9424	0.9267
1	1000	0.9444	0.9475	0.9301	0.9507	0.9303
2	1000	0.9496	0.9543	0.9322	0.9493	0.9185
3	1000	0.9596	0.9524	0.9521	0.9502	0.9246

Table 8 Testing results of GCL datasets

No. of Architecture	No. of Epoch	R^2 result with different pre-processing methods				
		Raw	Method 1	Method 2	Method 3	Method 4
1	500	0.997	0.992	0.9979	0.9942	-0.0573
2	500	0.98	0.976	0.9979	0.979	-0.0421
3	500	0.9959	0.9947	0.9942	0.9946	-0.0225
1	1000	0.9973	0.9965	0.9969	0.9947	-0.0506
2	1000	0.9971	0.9969	0.9985	0.9962	-0.0168
3	1000	0.996	0.9956	0.9988	0.9963	0.1183

For the CGL dataset, the testing results presented in Table 8 have shown that all the results are similar except preprocessed dataset with method 4, i.e. producing accuracy (R^2) of more than 0.97 in the testing result. In this case, the calibration using raw dataset is sufficient to get a good calibration accuracy.

Finally, The relatively better pre-processing methods for the respective dataset are highlighted in grey in the Tables 4, 5, 6, 7, 8. In general, from these experimentations, it can be observed that there is no single pre-processing method that work robustly for all datasets. It is understood that combination of these preprocessing methods are often performed in many studies. However, there are various uncertainties involved including the setting of the parameters in the pre-processing method and the order/sequence of the methods. In some cases, using raw data can even give better calibration accuracy that that using pre-processed data. This has been highlighted in the study in [28].

5 Conclusions

Three different architectures of MLPNN have been used in this research. Five NIR spectroscopy datasets have been used to train MLPNN. These datasets are pre-processed with 4 individual different methods. Based on the results, each pre-processing method only suitable for certain dataset. In some case, calibration using

raw dataset can produce better accuracy in the testing. It can be observed that there is no single pre-processing method that work robustly for all datasets. Further improvement will be carried out on this project such as adding more NIR spectral datasets including classification problem, involving combination of the pre-processing methods, and applying different machine learning algorithms. In addition, deep learning algorithm will be explored toward building robust calibration model with only simple pre-processing work.

Acknowledgements The authors would like to express gratitude to the Centre of Excellence for Research, Value Innovation and Entrepreneurship (CERVIE)—UCSI University - Malaysia for supporting this project. This project is also supported by Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme, with code: Proj-FRGS/1/2020/TK0/UCSI/02/4.


References

1. Silalahi DD, Reaño CE, Lansigan FP, Panopio RG, Bantayan NC (2016) Using genetic Algorithm neural network on Near Infrared spectral data for ripeness grading of oil palm (*Elaeis guineensis* Jacq.) Fresh Fruit. *Inf Process Agric* 3(4):252–261
2. Qu JH, Liu D, Cheng JH et al (2015) Applications of Near-Infrared Spectroscopy in food safety evaluation and control: A review of recent research advances. *Crit Rev Food Sci Nutr* 55(13):1939–1954
3. Magwaza LS, Tesfay SZ (2015) A review of destructive and non-destructive methods for determining avocado fruit maturity. *Food Bioprocess Technol* 8(10):1995–2011
4. Ru YJ, Glatz PC (2000). Application of Near Infrared Spectroscopy (NIR) for monitoring the quality of milk, cheese, meat and fish-review. vol 13. Asian-Aus
5. Alishahi A, Farahmand H, Prieto N, Cozzolino D (2010) Identification of transgenic foods using NIR spectroscopy: A review. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy* 75(1):1–7
6. Sakudo A (2016) Near-infrared spectroscopy for medical applications: Current status and future perspectives. *Clin Chim Acta* 455:181–188
7. Villar A, Gorritxategi E, Aranzabe E, Fernández S, Otaduy D, Fernández LA (2012) Low-cost visible-near infrared sensor for on-line monitoring of fat and fatty acids content during the manufacturing process of the milk. *Food Chem* 135(4):2756–2760
8. Hardesty JH, Attili B, Colledge C (2010) Spectrophotometry and the Beer-Lambert law: An important analytical technique in chemistry
9. Abdel-Aal RE, Al-Haddad MN (1997) Determination of radioisotopes in gamma-ray spectroscopy using abductive machine learning. *Nucl Instrum Methods Phys Res, Sect A* 391(2):275–288
10. Spiegelman CH, McShane MJ, Goetz MJ, Motamedi M, Li Yue Q, Coté GL (1988) Multivariate Calibration. vol 60. Wiley & Sons
11. Xie C, Xu N, Shao Y, He Y (2015) Using FT-NIR spectroscopy technique to determine arginine content in fermented *Cordyceps sinensis* mycelium. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. 149:971–977
12. Jahani S, Setarehdan SK, Boas DA, Yücel MA (2018) Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky-Golay filtering. *Neurophotonics*. 5(01):1
13. Jintao X, Liming Y, Yufei L, Chunyan L, Han C (2017) Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. 179:250–254

14. Skogholt J, Liland KH, Indahl UG (2019) Preprocessing of spectral data in the extended multiplicative signal correction framework using multiple reference spectra. *J Raman Spectrosc* 50(3):407–417
15. Solihin MI, Natarajan E, Ang CK, Khan MKA (2018) “Observation on Multivariable Regression Methods for Various Near Infrared Spectra Data”, 2018 IEEE 4th Int. Symp. Robot. Manuf. Autom. ROMA 2018:1–5
16. Lins APS, Ludermir TB (2005) Hybrid optimization algorithm for the definition of MLP neural network architectures and weights
17. Eğrioğlu E, Aladağ ÇH, Günay S (2008) A new model selection strategy in artificial neural networks. *Appl Math Comput* 195(2):591–597
18. Solihin MI, Shameem Y, Htut T, Ang CK, Bt Hidayab M (2019) Non-invasive blood glucose estimation using handheld near infra-red device. *Int J Recent Technol Eng* 8(3 Special Issue):16–19
19. Abdullah Al-Sanabani DG, Solihin MI, Pui LP, Astuti W, Ang CK, Hong LW (2019) Development of non-destructive mango assessment using Handheld Spectroscopy and Machine Learning Regression. In: *Journal of Physics: Conference Series*. vol 1367. Institute of Physics Publishing
20. Sun X, Subedi P, Walker R, Walsh KB (2020) NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment. *Postharvest Biology and Technology*. 163
21. Dyrby M, Engelsen SB, Nørgaard L, Bruhn M, Lundsberg-Nielsen L (2002) Chemometric Quantitation of the Active Substance (Containing C[N] in a Pharmaceutical Tablet Using Near-Infrared (NIR) Transmittance and NIR FT-Raman Spectra. vol 56
22. Data Sets - Eigenvector. <https://eigenvector.com/resources/data-sets/>. Accessed June 27, (2021).
23. Hsu HH, Hsieh CW, da Lu M (2011) Hybrid feature selection by combining filters and wrappers. *Expert Syst Appl* 38(7):8144–8150
24. Martens H, Stark E (1991) Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *J Pharm Biomed Anal* 9(8):625–635
25. Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol Rev* 65(6):386–408
26. Ramchoun H, Amine M, Idrissi J, Ghanou Y, Ettaouil M (2016) Multilayer perceptron: Architecture optimization and training. *International Journal of Interactive Multimedia and Artificial Intelligence*. 4(1):26
27. Ranković V, Grujović N, Divac D, Milivojević N (2014) Development of support vector regression identification model for prediction of dam structural behaviour. *Struct Saf* 48:33–39
28. Rinnan Å (2014) Pre-processing in vibrational spectroscopy – when, why and how. *Anal Methods* 6(18):7124–7129

A Piecewise Affine System Modeling Approach of Thin McKibben Muscle Servo Actuator



Mohd Akmal Mhd Yusoff , Ahmad Athif Mohd Faudzi ,
Mohd Shukry Hassan Basri, and Mohd Fuaad Rahmat 

Abstract Dynamic characteristic of thin McKibben muscle has not yet been fully investigated. Therefore, the objective of this study is to propose a piecewise affine system to model a thin McKibben muscle servo actuator. The static and dynamic modeling of the actuator has been performed using first principle approach. Different models for extension and retraction operations have been presented. The step input responses of the model have been simulated and compared to the actual system response. Result shows that the model's response is similar to the actual system with steady-state errors of 1.18% for 0.2 MPa input and 0% for 0.25 and 0.3 MPa inputs. The evidence from this study suggests that a piecewise affine system can be used to model a thin McKibben muscle servo actuator, which would facilitate future development of a novel switching controller.

Keywords Piecewise affine systems · Pneumatic actuators · McKibben muscle · Pneumatic artificial muscle · Pneumatic muscle actuators

1 Introduction

McKibben muscle is a type of pneumatic artificial muscle (PAM). An actuator using PAM is known as pneumatic muscle actuator (PMA). Most of the PMA systems use

M. A. Mhd Yusoff
Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia

A. A. Mohd Faudzi (✉) · M. F. Rahmat
School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia
e-mail: athif@utm.my

A. A. Mohd Faudzi
Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia

M. S. Hassan Basri
Engineering Research Centre, MARDI, 43400 Serdang, Selangor, Malaysia

off-the-shelf muscles, especially from Festo [1, 2]. However, there are some systems which use custom-built muscles [3, 4].

Recently, thin McKibben muscles with outer diameter of 1.8 mm have been produced commercially [5]. Different to conventional PAMs which are bulky, the muscle's thinness enables it to be used in tight spaces and to have a higher bandwidth [6]. The muscle has been used in biomimicry eye [7], wearable ankle exercise device [8], soft exoskeleton glove [9], continuum manipulator [10] and hand rehabilitation device [11]. Similarly thin muscles have also been developed and applied in humanoid robot arm [12].

Kurumaya et al. [13] have investigated the thin muscle's static characteristic but a thorough search of the relevant literature yielded no related article on its dynamic characteristic. Piecewise affine (PWA) system has been used to model PAMs [14, 15]. However, the PAMs are conventional rigid muscles which have different characteristics than thin PAMs such as the absence of slack and significantly higher mass. The main novelty of this article then is to propose a piecewise affine (PWA) system to model a linear thin McKibben muscle servo actuator. The system consists of several operating "modes" with one active mode at a time. This approach enables a novel control method to be applied to the servos actuator system, for example a switching model predictive control.

2 Methodology

This section is comprised of three parts. In the first part, the system under study is presented in brief. Then, in the second part, the static and dynamic modeling using first principle approach are elaborated. Finally in the third part, the PWA system is detailed out.

2.1 *Thin McKibben Muscle Servo Actuator*

The CAD drawings of the servo actuator prototype [16] are shown in Figs. 1 and 2. The prototype uses commercial thin McKibben muscles with inner-tube diameter of 1.3 mm and operating range of 0–0.5 MPa [13]. It is a single degree-of-freedom system. Similar to other typical PAM-actuated systems, it consists of two pairs of muscle to obtain a bi-directional motion. An example where this prototype can be implemented is in the pick-and-place application.

2.2 Static and Dynamic Modeling

The static behavior of the force acting on the rod from the muscle can be modeled using Schulte’s formula [17]

$$F = \frac{\pi D_0^2 P}{4} (3 \cos^2 \theta - 1) \tag{1}$$

where F , D_0 , P , and θ are the contractile force, maximum muscle diameter, applied pressure and muscle’s braid angle. In addition, the system under study can be considered a type of PAM-actuated servo system, and therefore the dynamic model similar to Shen [18] has been used. The dynamic model is slightly modified by adding the spring term, Kx to reflect the muscle’s viscoelastic behavior [19]. In addition, different to conventional rigid PAM, the thin McKibben muscle has slack and passive

Fig. 1 Isometric view of the top part in CAD

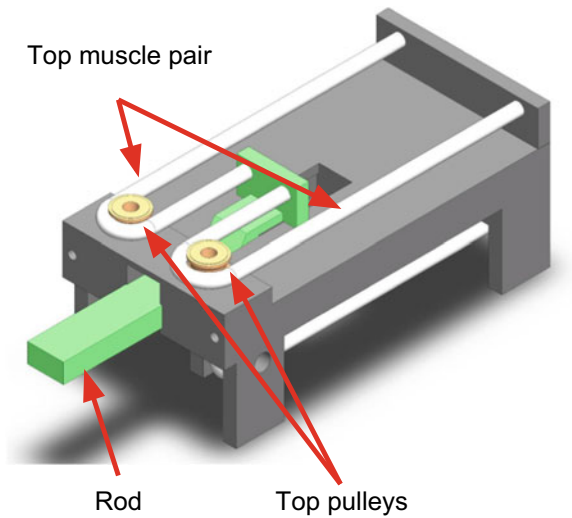
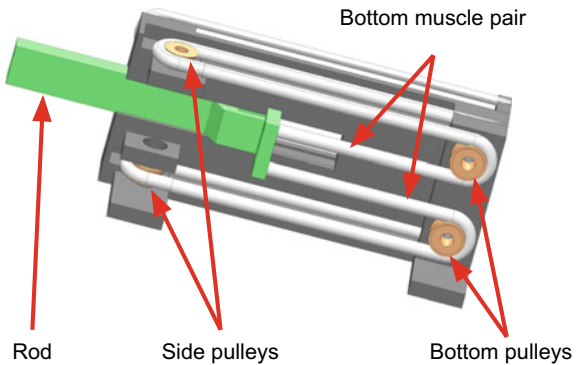


Fig. 2 Cross section view of the bottom part in CAD



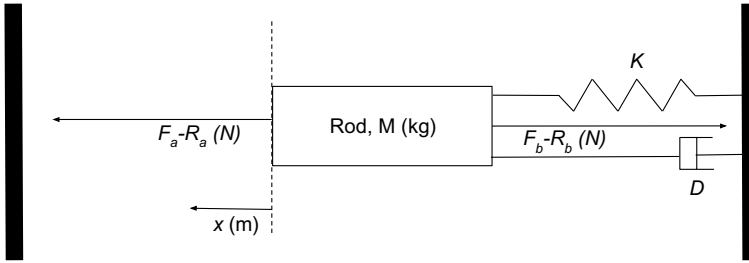


Fig. 3 Dynamic model of the thin McKibben muscle servo actuator

contraction range whereby its contractile force doesn't affect the load displacement. Assuming that the McKibben muscle's mass and friction are negligible, the system can then be represented by

$$M\ddot{x} + D\dot{x} + Kx = F_a - F_b - R_a + R_b \quad (2)$$

where M , D , K , \ddot{x} , \dot{x} and x are the mass, damping coefficient, spring coefficient, acceleration, velocity and position of the rod, and F_a , F_b , R_a and R_b are the contractile force and passive contractile force of the top and bottom muscle pair respectively. The dynamic model is illustrated in Fig. 3. Rearranging (2), the acceleration of the rod for $x > 0$ is given by

$$\ddot{x} = \frac{(F_a - F_b - R_a + R_b - D\dot{x} - Kx)}{M} \quad (3)$$

Even though (1) gives a good background on the McKibben muscle behavior, it does not give a good experimental result especially on higher pressure [6]. Since the McKibben muscle used in this work is similar to the one investigated by Kurumaya et al. [13], their static force model has been used instead. The relationship between the contractile force, F , absolute pressure P and contraction ratio ε is given by

$$F = \alpha(1 - \varepsilon)^2 P + \beta P + \gamma \quad (4)$$

where $\alpha = 6.70 \times 10^{-5} \text{ m}^2$, $\beta = -1.96 \times 10^{-5} \text{ m}^2$ and $\gamma = -8.76 \text{ N}$. The contraction ratio is the ratio between the muscle's contraction and its original length and is given by

$$\varepsilon = \frac{d}{l_0} = \frac{(l_0 - l)}{l_0} \quad (5)$$

where d is the contraction, l_0 is the original length and l is the final length.

For the top muscle pair, its contraction is equal to the displacement of the rod, x whereas for the bottom muscle pair, its contraction is equal to $x_{max} - x$. Therefore, from (4), F_a and F_b are given by

$$F_a = P_a \left[\alpha \left(1 - \frac{x}{l_a} \right)^2 + \beta \right] + \gamma \quad (6)$$

$$F_b = P_b \left[\alpha \left(1 - \frac{(x_{max}-x)}{l_b} \right)^2 + \beta \right] + \gamma \quad (7)$$

where l_a and l_b are the length of each top and bottom muscles and P_a and P_b are the pressure to the top and bottom muscle pairs respectively.

2.3 PWA System

During extension operation, P_b at j -th operating point, $P_{j,b}$ is kept constant. If the bottom muscle's slack length $r_{j,b} > 0$, there will not be any contractile force acting on the load from bottom muscle pair, and the terms with F_b and R_b can be omitted. Substituting (6) into (3) and setting \dot{x} , \ddot{x} , F_b and R_b to zero, and x and P_a to x and P_a at j -th operating point, x_j^{op} and $P_{j,a}^{op}$ respectively yield

$$P_{j,a}^{op} = \frac{(K_j x_j^{op} + R_{j,a} - \gamma)}{\left[\alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta \right]} \quad (8)$$

Similarly for retraction operation, replacing (7) into (3) and setting \dot{x} , \ddot{x} , F_a , D_a , K_a and R_a to zero, and x and P_b to x_k^{op} and $P_{k,b}^{op}$, $k \subseteq \mathbb{Z}^+$ give

$$P_{k,b}^{op} = \frac{(K_k x_k^{op} + R_{k,b} - \gamma)}{\left[\alpha \left(1 - \frac{(x_{max}-x_k^{op})}{l_b} \right)^2 + \beta \right]} \quad (9)$$

To find the state space models of the PWA system, the dynamic model (3) is linearized around multiple equilibrium points [20] $j = \{1, \dots, N\}$, $N \in \mathbb{Z}^+$ for extension and $k = \{N+1, \dots, S\}$, $S \in \mathbb{Z}^+$ for retraction. To do so, we consider small perturbations around x and P_a for extension and x and P_b for retraction.

$$x = x_j^{op} + \delta x, P_a = P_{j,a}^{op} + \delta P_a(\text{extension}) \quad (10)$$

$$x = x_k^{op} + \delta x, P_b = P_{k,b}^{op} + \delta P_b(\text{retraction}) \quad (11)$$

Setting (10) in (3) and dropping the terms associated with bottom muscle pair yield

$$M\delta\ddot{x} + D_j\delta\dot{x} + K_j x_j^{op} + K_j \delta x + R_{j,a} = F_a \left(x_j^{op} + \delta x, P_{j,a}^{op} + \delta P_a \right) \quad (12)$$

for extension, and setting (11) in (3) and dropping the terms associated with top muscle pair yield

$$M\delta\ddot{x} + D_k\delta\dot{x} + K_k x_k^{op} + K_k \delta x - R_{k,b} = -F_b(x_k^{op} + \delta x, P_{k,b}^{op} + \delta P_b) \quad (13)$$

for retraction. To estimate $F_a(x_j^{op} + \delta x, P_{j,a}^{op} + \delta P_a)$ and $F_b(x_k^{op} + \delta x, P_{k,b}^{op} + \delta P_b)$, Taylor series expansions are applied to (6) and (7) around x_j^{op} and $P_{j,a}^{op}$, and x_k^{op} and $P_{k,b}^{op}$ respectively. Taking the first two order terms yields

$$\begin{aligned} F_a(x_j^{op} + \delta x, P_{j,a}^{op} + \delta P_a) &= P_{j,a}^{op} \left[\alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta \right] + \gamma + \left[\alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta \right] \delta P_a \\ &\quad - P_{j,a}^{op} \left[\frac{2\alpha}{l_a} \left(1 - \frac{x_j^{op}}{l_a} \right) \right] \delta x \end{aligned} \quad (14)$$

and

$$\begin{aligned} F_b(x_k^{op} + \delta x, P_{k,b}^{op} + \delta P_b) &= P_{k,b}^{op} \left[\alpha \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right)^2 + \beta \right] \\ &\quad + \gamma + \left[\alpha \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right)^2 + \beta \right] \delta P_b \\ &\quad + P_{k,b}^{op} \left[\frac{2\alpha}{l_b} \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right) \right] \delta x \end{aligned} \quad (15)$$

Substituting (8) into (14) and (9) into (15) yields

$$\begin{aligned} F_a(x_j^{op} + \delta x, P_{j,a}^{op} + \delta P_a) &= K_j x_j^{op} + R_{j,a} + \left[\alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta \right] \delta P_a \\ &\quad - \frac{(K_j x_j^{op} + R_{j,a} - \gamma) \left[\frac{2\alpha}{l_a} \left(1 - \frac{x_j^{op}}{l_a} \right) \right]}{\left[\alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta \right]} \delta x \end{aligned} \quad (16)$$

and

$$F_b(x_k^{op} + \delta x, P_{k,b}^{op} + \delta P_b) = -K_k x_k^{op} + R_{k,b} + \left[\alpha \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right)^2 + \beta \right] \delta P_b$$

$$+ \frac{(K_k x_k^{op} + R_{k,b} - \gamma) \left[\frac{2\alpha}{l_b} \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right) \right]}{\left[a \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right)^2 + \beta \right]} \delta x \quad (17)$$

Substituting (16) into (12) and (17) into (13) yields

$$M \delta \ddot{x} + D_{j,a} \delta \dot{x} + Q_j \delta x = T_j \delta P_a \quad (18)$$

for extension and

$$M \delta \ddot{x} + D_k \delta \dot{x} + Q_k \delta x = T_k \delta P_b \quad (19)$$

for retraction where

$$Q_j = K_j + P_{j,a}^{op} \left[\frac{2\alpha}{l_a} \left(1 - \frac{x_j^{op}}{l_a} \right) \right], \quad (20)$$

$$T_j = \alpha \left(1 - \frac{x_j^{op}}{l_a} \right)^2 + \beta, \quad (21)$$

$$Q_k = K_k + P_{k,b}^{op} \left[\frac{2\alpha}{l_b} \left(1 - \frac{(x_{max} - x_k^{op})}{l_b} \right) \right], \quad (22)$$

$$T_k = -\alpha \left[1 - \frac{(x_{max} - x_k^{op})}{l_b} \right]^2 - \beta. \quad (23)$$

Adding the term $M \ddot{x}_j^{op} + D_j \dot{x}_j^{op} + Q_j x_j^{op} + T_j P_{j,a}^{op}$ to both sides of (18) yields

$$M \ddot{x} + D_j \dot{x} + Q_j x + T_j P_{j,a}^{op} = M \ddot{x}_j^{op} + D_j \dot{x}_j^{op} + Q_j x_j^{op} + T_j P_a \quad (24)$$

whereas adding the term $M \ddot{x}_k^{op} + D_k \dot{x}_k^{op} + Q_k x_k^{op} + T_k P_{k,b}^{op}$ to both sides of (19) yields

$$M \ddot{x} + D_k \dot{x} + Q_k x + T_k P_{k,b}^{op} = M \ddot{x}_k^{op} + D_k \dot{x}_k^{op} + Q_k x_k^{op} + T_k P_b. \quad (25)$$

By definition, the operating points x_j^{op} and x_k^{op} satisfy $\ddot{x}_j^{op} = \dot{x}_j^{op} = 0$ and $\ddot{x}_k^{op} = \dot{x}_k^{op} = 0$ respectively, which results in

$$\ddot{x} + \frac{D_j}{M} \dot{x} + \frac{Q_j}{M} x + \frac{T_j P_{j,a}^{op}}{M} = \frac{Q_j}{M} x_j^{op} + \frac{T_j}{M} P_a \quad (26)$$

and

$$\ddot{x} + \frac{D_k}{M} \dot{x} + \frac{Q_k}{M} x + \frac{T_k P_{k,b}^{op}}{M} = \frac{Q_k}{M} x_k^{op} + \frac{T_k}{M} P_b. \quad (27)$$

The state space models of the system for extension operation near operating points x_j^{op} is thus given by

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{Q_j}{M} & -\frac{D_j}{M} \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{T_j}{M} \end{bmatrix} P_a + \begin{bmatrix} 0 \\ \frac{Q_j}{M} x_j^{op} - \frac{T_j P_{j,a}^{op}}{M} \end{bmatrix}. \quad (28)$$

whereas for retraction operation near operating points x_k^{op} is given by

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{Q_k}{M} & -\frac{D_k}{M} \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{T_k}{M} \end{bmatrix} P_b + \begin{bmatrix} 0 \\ \frac{Q_k}{M} x_k^{op} - \frac{T_k P_{k,b}^{op}}{M} \end{bmatrix} \quad (29)$$

There are S partitions of the polyhedral regions (28) and (29) in the space of state and inputs with distinct affine state-update and output equations. The active mode at each instance is determined by the current rod position, x [21].

3 Results and Discussion

Based on the work by Serres [22], the values of K and D at the i -th operating point, $i \subseteq \mathbb{Z}^+$ can be estimated by

$$K_i = \frac{F_{i,act}}{x_{i,max}} \quad (30)$$

$$D_i = \tau K_i \quad (31)$$

where $F_{i,act} = F_{i,\bar{a}} - F_{i,\bar{b}}$ and $x_{i,max}$ are the active contractile force and the maximum rod displacement at i -th operating point, and τ is the time constant of the displacement response. Based on our previous discussion, at a constant pressure P , F is equal to zero when the muscle is fully contracted. Therefore, for every pressure point, substituting $\varepsilon = \varepsilon_{max}$ in (4) should produce $F = 0$. Any excess value in F can then be considered passive contractile force, R which does not affect the rod displacement. Therefore, the active contractile force at the i -th operating point is given by

$$F_{i,act} = F_i - R_i \quad (32)$$

Using the experiment setup similar to [16], the displacement vs pressure data were collected. The pressure inputs were varied using CKD MEVT500-1-E2-3 electro pneumatic regulator which received the input profiles from the host computer running MATLAB and Simulink. Piston displacements were measured using Gefran PZ-34-A-250 linear transducer and sent to the host computer. F_i and R_i values at P_i were calculated by setting $P = P_i$, and $\varepsilon_i = 0$ and $F_i = 0$ in (4) respectively. $F_{i,act}$ was then obtained by substituting the values of F_i and R_i in (32). Consequently, the

Table 1 Thin McKibben muscle servo actuator PWA modes

Operation	j	Operating points, x_{op} [m, m/s]	Operating range, $x \in$ (m)	K
Extension	1	[0.0020,0]	[0,0.0020]	511
	2	[0.0030,0]	[0.0020, 0.0030]	368
	3	[0.0063,0]	[0.0030, 0.0065]	270
	4	[0.0100,0]	[0.0065, 0.0130]	154
	5	[0.0200,0]	[0.0130,0.0200]	116
	6	[0.0236,0]	[0.0200,0.0030]	112
Retraction	7	[0.031,0]	[0.0291,0.0040]	10,462
	8	[0.029,0]	[0.0241, 0.0290]	9127
	9	[0.0237,0]	[0.0169, 0.0240]	6500
	10	[0.0163,0]	[0.0110, 0.0168]	5179
	11	[0.0230,0]	[0.0070,0.0109]	4567
	12	[0.005,0]	[0,0.0069]	4345

values of spring and damping coefficients at the i -th operating point were calculated using (30) and (31). Since the values are just a rough estimate, some fine tuning were needed to get a good value of both [14]. Based on the results, twelve operating modes have been selected (Table 1). Using Multi-Parametric Toolbox 3.0 [23], YALMIP [24] and Hybrid System Description Language (HYSDEL) [25], the developed model has been simulated with step input pressure of 0.2, 0.25 and 0.3 MPa. The displacement responses were then compared with the actual system responses.

For 0.2 MPa step input, the model achieves a steady-state value of 0.0084 m compared to the actual system’s steady-state value of 0.0085 m, which equates to a steady-state error of 1.18% (Fig. 4). The model takes 4 modes to reach the steady state from the initial position of $x = 0$ m, which agrees with Table 1.

For 0.25 and 0.3 MPA step inputs, the model achieves steady-state values of 0.02 and 0.0285 m respectively which are equal to the actual system value (Figs. 5 and 6). The model takes 5 and 6 modes respectively to reach the steady state from the initial position, which agrees with Table 1.

From the results, it can be inferred that the PWA system can model the actual system with good accuracy. The small errors are mainly due to the mode switching which approximated the nonlinear system, and the estimated values of K and D . It should be noted that the more modes the PWA system has, the more accurate the model would be, with a tradeoff in model complexity.

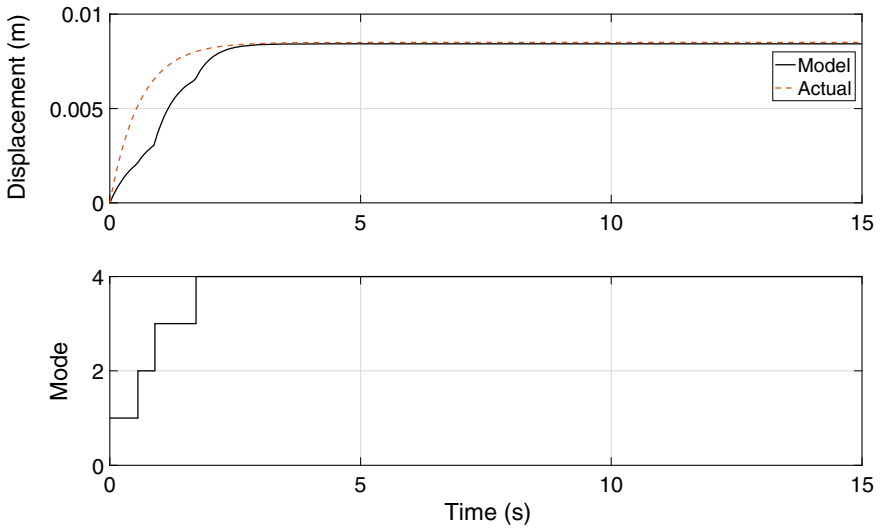


Fig. 4 0.2 MPa step response of the model compared with the actual system

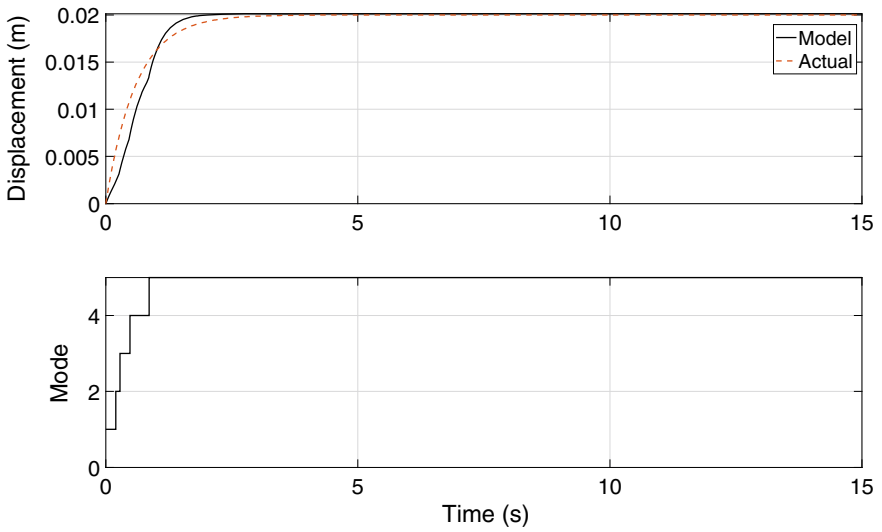


Fig. 5 0.25 MPa step response of the model compared with the actual system

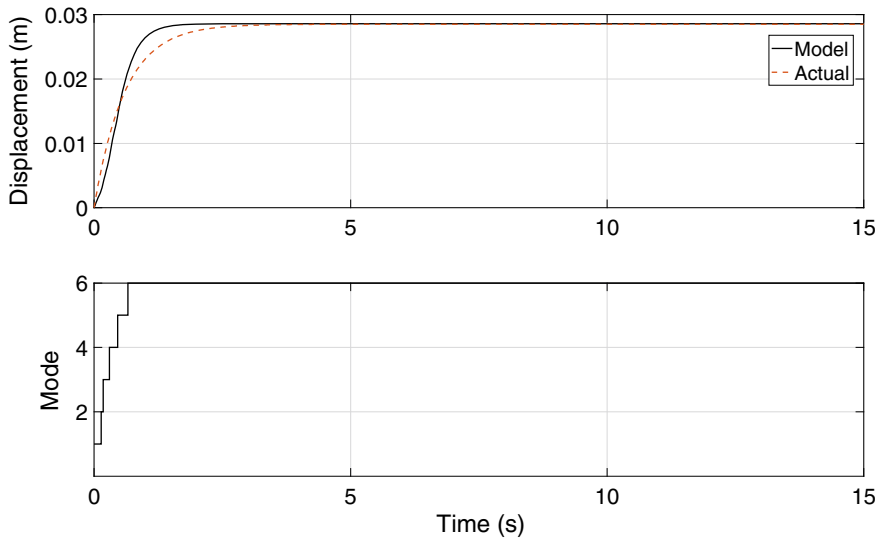


Fig. 6 0.3 MPa step response of the model compared with the actual system

4 Conclusion

The objective of this study is to propose a piecewise affine system to model a thin McKibben muscle servo actuator. It was found that the model produces a similar performance to the actual system in terms of step responses with steady-state errors of 1.18% for 0.2 MPa input and 0% for 0.25 MPa and 0.3 MPa inputs. The evidence from this study suggests that PWA system is a good candidate to model a thin McKibben servo actuator. In the future, a switching controller based on the PWA system would be developed.

Acknowledgements The authors would like to acknowledge the sponsorship provided by Ministry of Higher Education Malaysia (MOHE) through support under Fundamental Research Grant Scheme (FRGS/1/2019/TK04/UTM/02/41). The authors would also like to express appreciation to Universiti Teknologi Malaysia (UTM) and Engineering Research Centre, MARDI for facilities support and all the A2Lab UTM members for their direct or indirect support in making this publication possible.

References

1. Kodama T, Kogiso K (2017) Applications of UKF and EnKF to estimation of contraction ratio of McKibben pneumatic artificial muscles. In 2017 American control conference (ACC). IEEE, Washington, pp 5217–5222
2. Martens M, Boblan I (2017) Modeling the static force of a Festo pneumatic muscle actuator: a new approach and a comparison to existing models. *Actuators* 6(1):33

3. Yi J, Chen XJ, Song CY, Wang Z (2018) Fiber-reinforced origamic robotic actuator. *Soft Robot* 5(1):81–92
4. Faudzi AAM, Endo G, Kurumaya S, Suzumori K (2017) Long-legged hexapod Giacometti robot using thin soft McKibben actuator. *IEEE Robot Autom Lett* 3(1):100–107
5. S-Muscle. <https://www.s-muscle.com>. Accessed 1 July 2021
6. Davis S, Tsagarakis N, Canderle J, Caldwell DG (2003) Enhanced modelling and performance in braided pneumatic muscle actuators. *Int J Robot Res* 22(3–4):213–227
7. Hong WS, Mohd Faudzi AA, Rosli NR (2020) Design and control of biomimicry eye using soft actuator. *PERINTIS eJ* 10(1)
8. Chang T-H, Koizumi S, Nabae H, Endo G, Suzumori K, Hatakeyama K, et al (2020) A wearable ankle exercise device for deep vein thrombosis prevention using thin McKibben muscles. In: 2020 8th IEEE RAS/EMBS international conference for biomedical robotics and biomechatronics (BioRob), New York, USA, pp 42–47
9. Takahashi N, Furuya S, Koike H (2020) Soft Exoskeleton glove with human anatomical architecture: production of dexterous finger movements and skillful piano performance. *IEEE Trans Haptics* 13(4):679–690
10. Mohamed MF, Hanif ASM, Faudzi AA (2020) Segmentation of a soft body and its bending performance using thin McKibben muscle. *Int J Autom Mech Eng* 17(1):7533–7541
11. Ying S, Al-Shammari N, Faudzi A, Sabzehmeidani Y (2020) Continuous progressive actuator robot for hand rehabilitation. *Eng Technol Appl Sci Res* 10(1):5276–5280
12. Fan J, Jin J, Wang Q (2020) Humanoid muscle-skeleton robot arm design and control based on reinforcement learning. In: 2020 15th IEEE conference on industrial electronics and applications (ICIEA), Kristiansand, Norway, pp 541–546
13. Kurumaya S, Nabae H, Endo G, Suzumori K (2017) Design of thin McKibben muscle and multifilament structure. *Sens Actuators A* 261:66–74
14. Andrikopoulos G, Nikolakopoulos G, Arvanitakis I, Manesis S (2014) Piecewise affine modeling and constrained optimal control for a pneumatic artificial muscle. *IEEE Trans Industr Electron* 61(2):904–916
15. Schindele D, Aschemann H (2008) Nonlinear model predictive control of a high-speed linear axis driven by pneumatic muscles. In: 2008 American control conference. IEEE, Washington, pp 3017–3022
16. Mhd Yusoff MA, Mohd Faudzi AA, Hassan Basri MS (2021) Feasibility of Pi control for a double-acting cylinder actuated by McKibben muscles. Springer, pp 327–339
17. Schulte H Jr (1961) The characteristics of the McKibben artificial muscle (1961) The Application of external power in prosthetics and orthotics. National Academy of Sciences-National Research Council, Washington DC, Appendix H, pp 94–115
18. Shen X (2010) Nonlinear model-based control of pneumatic artificial muscle servo systems. *Control Eng Pract* 18(3):311–317
19. Endo N, Kizaki Y, Kamamichi N (2020) Flexible pneumatic bending actuator for a robotic tongue. *J Robot Mechatron* 32(5):894–902
20. Andrikopoulos G, Nikolakopoulos G, Arvanitakis I, Manesis S (2013) Switching model predictive control of a pneumatic artificial muscle. *Int J Control Autom Syst* 11(6):1223–1231
21. Borrelli F, Bemporad A, Morari M (2017) Predictive control for linear and hybrid systems. Cambridge University Press
22. Serres JL (2008) Dynamic characterization of a pneumatic muscle actuator and its application to a resistive training device. Ph.D. thesis, Wright State University, Ohio, USA
23. Hecceg M, Kvasnica M, Jones CN, Morari M (2013) Multi-parametric Toolbox 3.0. In: 2013 European control conference (ECC), Zurich, Switzerland, pp 502–510
24. Lofberg J (2004) YALMIP: a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE international conference on robotics and automation (IEEE Cat No04CH37508), pp 284–289
25. Torrisi FD, Bemporad A (2004) HYSDEL—a tool for generating computational hybrid models for analysis and synthesis problems. *IEEE Trans Control Syst Technol* 12(2):235–249

A Comparison of Type 1 and Type 2 Fuzzy Logic Controller for DC Motor System



Nafrizuan Mat Yahya , Nurainaa Elias ,
and Muhammad Hazim Mohd Nordin

Abstract In today's modern world application, there are high-level uncertainties that are faced and affecting every kind of operation in various industries. Thus, researchers, today are on the rise to find solutions that will able help to reduce these uncertainties in many types of situations especially control system applications. The type-1 Fuzzy Logic Controller is shown not to be able to handle a high level of uncertainties and the new type-2 Fuzzy Logic Controller is now being said to be able to do a better performance than the type-1 especially in controlling a DC motor system. This can be seen by the simulation graph that clearly observes the comparison of both types. The result where FLC type 2 outperforms FLC type 1 with reduced settling time and rising time can be seen. In conclusion, the new type-2 FLC is now able to overcome the limits of what type-1 FLC are able to do and this will give birth to better and improved performance of new Fuzzy Logic Controllers that is well suited as controllers for DC motor system. This paper will briefly discuss the comparison between both of these types of FLC and the benefits.

Keywords Fuzzy Logic Controller · FLC type 1 · FLC type 2 · DC motor system and control system

1 Introduction

Particularly in today's modern industry, direct current or DC motor is still a relevant option, although it is regarded as one of the oldest electric motor designs [1]. This confirms that even when opposed to modern engine models today, the DC motor has a great deal of benefits. The great controllability of the DC motor is one of the key benefits that can be seen when the DC motor is powered by only a change in the input voltage or reversing the leads [2]. Thus, every mechanism which could regulate the direction, speed, or torque of a DC-driven motor in its system and also allows its user to set it according to their will is a DC motor control system. There are several

N. M. Yahya · N. Elias (✉) · M. H. M. Nordin
Universiti Malaysia Pahang, Gambang, Malaysia
e-mail: nurainaa_elias@yahoo.com

types of DC motor controllers that can be used within the system, such as PID, IMC, and Fuzzy Logic controllers, for running the DC motor.

In the face of confusion and imprecision, the Fuzzy logic controller or FLC is intended to be a stable controller that can run. In addition, FLC also provided us with a method for the user-friendly construction of controller algorithms that is relatively similar to human-like thought and perception. The fuzzy control of logic is indicated by L. Zadeh in 1965 and has since been used globally, especially in recent years [3]. It is best represented as a set of rules that can determine the behavior of complex systems that cannot be mathematically specified.

In this research, we propose a simulation-based performance comparison between the two types of generally used FLC in the industry nowadays which are the FLC type 1 and FLC type 2. Type 1 fuzzy logic controllers are commonly used in machine control. Fuzzy stems from the fact that it is not possible to express any reasoning involved in the process as “true” or “false” but as “partly true”. It including the initial phase, an operation phase, and finally a phase of output [4]. The sensor or other inputs will be mapped to the relevant membership functions and their true values in the input process. Next, each appropriate rule is invoked by the processing stage and results are created for each and combined into rule outcomes. Eventually, the final phase will transform an integrated outcome return to the particular output income of the control [5].

Type 2 fuzzy logic controllers. Since it is efficient in different applications, many researchers are now adopting the fuzzy logic of type 2, and it is expected that this growth in the community will continue [6]. In addition to other evidence of interest in Type 2 FLC, the advent of Fuzzy-IEEE type 2 fuzzy logic research since 2016 also exists. Type 2 fuzzy controllers have uncertainties in the second order, enabling fuzzy systems to really deal with the ambiguity of the real world. For more efficient and durable but cheaper and more affordable hardware type 2 fuzzy controllers, we can see the race of development today that poses a large horizon in exploring the uncertainties of the real world [7].

2 Methodology

2.1 Type 1 Fuzzy Logic Controller

In 1965, Professor Lotfi A. Zadeh proposed the idea of Fuzzy logic Controller (FLC) Type-1 with the goal of creating an output that is achieved as an alternative to crisp value via a series of memberships [9]. The FLC has fuzzy rules based on the expert’s experience of a system’s success over time. The fuzzifier, inference engine and defuzzifier are a conventional FLC or Type-1 FLC [3]. The Fuzzifier acts as a converter that transforms the input parameter’s crisp values into fuzzy sets. This also depends on the fuzzy rules that are based on the system’s parameter membership function. Fuzzy outputs are then obtained through the Inference Engine and,

eventually, the defuzzifier transforms all fuzzy inputs into crisp FLC outputs used for control [2].

The definition of FLC is also derived from the fuzzy sets that, by human-like reasoning, are one of the most used methods of overcoming linguistic imprecision [3]. In their fuzzy sets, FLC are also very versatile for the uncertainty as it can also match uncertainty type values that are not crisp values, which makes it more robust logic system, unlike models that require crisp value or threshold in order to have a correct operation [4]. Type-1 Fuzzy set is known as a membership function, μ_A, x , characterized by a set in the universe of discourse X By taking the values at the (0, 1) interval. The membership value of x in the A set is denoted by this membership function. A Type-1 Fuzzy sets that are rendered with all elements in x with membership function, μ_A , are x -cut. This reflects the degree of membership of element $x \in X$.

$$A(\alpha) = \{x, \mu_A(x) | x \in X\}, \tag{1}$$

In which, $\mu_A: X \rightarrow [0, 1]$. Based on their alpha-cuts, the process of addition, subtraction, multiplication or division that the Type-1 Fuzzy sets undergo is specified.

Finally, in the case of the values $[a, b]$ and $[d, e]$, the alpha-cut value of the Type-1 Fuzzy set is set to A_1 and A_2 (i.e., in the case of $0 \leq \alpha \leq 1$). Thus, the alpha-cut process of Type-1 Fuzzy sets of A_1 is added, subtracted, multiplied and divided. And A_2 it will be listed as:

$$A_1 + A_2 = [a, b] + [d, e] = [a + d, b + e] \tag{2}$$

$$A_1 - A_2 = [a, b] - [d, e] = [\min(a - d, b - e), \max(a - d, b - e)] \tag{3}$$

$$A_1 \times A_2 = [a, b] \times [d, e] = [\min(ad, ae, bd, be), \max(ad, ae, bd, be)] \tag{4}$$

$$A_1 \div A_2 = [a, b] / [d, e] = [\min\left(\frac{a}{d}, \frac{a}{e}, \frac{b}{d}, \frac{b}{e}\right), \max\left(\frac{a}{d}, \frac{a}{e}, \frac{b}{d}, \frac{b}{e}\right)] \tag{5}$$

2.2 Type 2 Fuzzy Logic Controller

On the other hand, Type-2 FLC is the latest approach that can provide special features that can solve the drawbacks of the older Type-1 FLC in the event of uncertainties being handled. It is also said that FLC Type-2 can accommodate additional degrees of freedom in different types of uncertainties [8]. Next, Fuzzy Logic Controller Type-2 has a few steps to obtain a crisp value for it, which is firstly to transform crisp inputs into fuzzy inputs through type 2 membership functions, then a set of

fuzzy rules selected will be merged with the outputs forming a single set under the inference mechanism. Under the type reduction operation method, the performance of the inference system will then be converted to Type-1 and will again be converted into crisp value by using many defuzzification techniques that are the same process as in Type-1 FLCC [2].

Professor Lotfi A. Zadeh also introduced Type-2 FL in 1975 10 years after the introduction of Type-1 FL. While most of the FLC breakthrough came after the twentieth century, several researchers such as Dubois and Prade (1978, 1979), Gorzalczany (1987, 1988), Mizumoto and Tanaka (1981), and Norwich and Turksen (1981) have already contributed to this notion (1982). As Type-1 Fuzzy Logic started to be redundant in managing more complex scenarios, Type-2 Fuzzy Logic is mostly suggested [9]. They are also an improved version of the Type-1 Fuzzy Logic and are designed to be capable of handling a more notable degree of modeling imprecision and to be able to manage uncertainties by taking into account unreliability and knowledge vagueness [10].

In 1979, Mizumoto and Tanaka created what would become the basic definition of Type-2 FLC, based on the idea of Zadeh. A type-2 fuzzy package, \tilde{A} . With a Type-2 membership feature, $\mu_{\tilde{A}}(x, u)$ based on [2], can be described as follows:

$$\tilde{A} = \{(x, u), \mu_{\tilde{A}}(x, u)\} \forall x \in X, \forall u \in J_x \subseteq [0, 1] \quad (6)$$

In which, $\mu_A: X \rightarrow [0, 1]$. Where $x \in X, u \in J_x \subseteq [0, 1]$ and $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$. At the same time, \tilde{A} can also be characterized as

$$\tilde{A} = \iint_{x \in X, u \in J_x \subseteq [0, 1]} \frac{\mu_{\tilde{A}}(x, u)}{(x, u)} \quad (7)$$

The union of all Type-2 Fuzzy sets' primary membership is represented by a two-dimensional domain called FOU of \tilde{A} . The FOU is defined by its two boundary functions, the Upper Membership Feature (UMF) and the Lower Membership Feature (LMF). UMF is marked as

$$\bar{\mu}_{\tilde{a}}(x) = \frac{1}{FOU(\tilde{a})}, \forall x \in X \quad (8)$$

Although LMF is marked as

$$\underline{\mu}_{\tilde{a}}(x) = \frac{FOU(\tilde{a})}{1}, \forall x \in X \quad (9)$$

Notice that, J_x . Note The set is an interval;

$$J_x = \{(x, u): u \in [\underline{\mu}_{\tilde{a}}(x), \bar{\mu}_{\tilde{a}}(x)]\} \quad (10)$$

A fuzzy embedded set \tilde{A}_e . X and u are represented in equations for continuous universe discourse

$$\tilde{A}_e = \int_{x \in X} \left[\frac{1}{u} \right] / x, u \in J_x \quad (11)$$

The set, \tilde{A}_e can be seen in the Eq. 11. Embedded in \tilde{A} . As such, at each value of X , the secondary membership function will always be one. Thus, in order to form a Type-2 Fuzzy set, it can be inferred that a large number of embedded Type-1 Fuzzy sets are combined. The Type-2 Fuzzy sets can be viewed as a mixture of several different Type-1 Fuzzy sets in which each of the Type-1 Fuzzy sets is embedded in order to form the FOU.

2.3 DC Motor Control System

One of the most widely used electric motors in our everyday lives is the direct control (DC) motor [11]. In instruments, electrical toys and small appliances, it is primarily used as a small-type DC motor. There is also a broader type of DC engine used in elevators and as a motor for electric vehicles, such as electric cars. Any class of rotary electric motors that work to convert direct current electrical energy into mechanical energy is known as a DC motor. The type most frequently found is the type that relies on the forces generated by magnetic fields. It is mainly used by many users via PID and Fuzzy Logic control for its robust, simplicity and easy to control [10]. The servo motors, the brushed DC motor and the brushless DC motor are many types of DC motor available on the market. A DC engine consists primarily of 2 elements, which are part of the ‘stator’ and part of the ‘rotor.’ The ‘stator’ is the stationary portion of the DC motor, while the ‘servo’ component is the DC motor’s moving part.

The brushless DC motor will be next, which has a higher efficiency, reliability and wide range of spread than both the brushed DC motor and servo motor. Brushless DC motor is used in various kinds of industries, such as in the biomedical and robotic industry, due to its success in being able to handle high torque to weight ration and also being extremely precise, particularly in the industry that needs precision handling control for accuracy. Furthermore, since it acts as a synchronous motor, the brushless DC motor has a major benefit, so it did not experience any form of ‘slip’ that is definitely found in induction-type motors. The Brushless DC motor works by providing three permanent magnetic motor stator coils in which two coils transform the magnetic field and a floating coil is then responsible for the back electro-motive power (EMF). But the brushless DC motor also has its disadvantages because of the electrical switching within the brushless DC motor that would lead to the torque ripple due to the incorrect installation of position sensors. The brushless feedback can be either sensor-based or sensor-less [11].

Table 1 DC motor parameter

Parameter	Value
Armature resistance, R_a	11.2 Ω
Armature inductance, L_a	0.05 H
Rotor inertia, J_m	0.5 kg * m ²
Rotor damping, b	0.002953 N * m/(rad/s)
Moment coefficient, K_i	1.25 Nm/A
Back emf constant, K_b	1.25 V/(rad/s)

As their basic equation for regulating the armature voltage equation is almost the same, brushless DC motor modeling can actually be compared to that of a three-phase synchronous motor. Thus, from here we can conclude that the brushless DC motor speed response is dependent on the parameter of R, B, J (Table 1).

3 Results and Discussion

First, the parameter data of the membership function of input by using the membership function editor will be set. For this variable the file name set to speed. For the speed membership function, set the membership function for lowspeed, optimumspeed and highspeed. For the parameter of the membership function low speed, inserted value of [0 0 150]. Next, inserted the value of membership function of optimumspeed which are [0 150 300]. For the highspeed membership function, inserted the value of [150 300 300]. From here the range speed of the variable is set from 0 to 300 can be seen.

Next will be the second input variable which will be set to voltage that will be used to control speed of the dc motor. For this variable, 3 membership function which are low, optimum and high can be defined which would represent the low voltage, optimum plot and high voltage. The membership function of Low parameters will be set first with its parameters will be set as [0 0 2.5]. Next is the parameters of Optimum that will be set as [0 2.5 5]. Lastly, the parameters for High which are set as [2.5 5 5]. Thus, the range that is used in this variable is from 0 to 5. Next, this fis file exported to workspace so that it can be used as another fis file in the circuit.

When the circuit is set, the fis file is by using fuzzy logic block parameters. The fis file is selected in the FIS structure and run. The result will then be presented in graph form based on the parameters of the membership function that we had set. The file is then is exported in order for it to be used by the circuit. The result when the circuit is run in Simulink program is shown in Fig. 1.

Figure 1 depicts the outcomes of the computer simulation for the step response. The blue color in the graph indicates the benchmark, which represents the system without a fuzzy logic controller, and the green line represents the type-1 fuzzy logic controller. The yellow line indicates this approach has employed the type-2 fuzzy logic controller. According to the graph, the FLC type-2 performs better than the

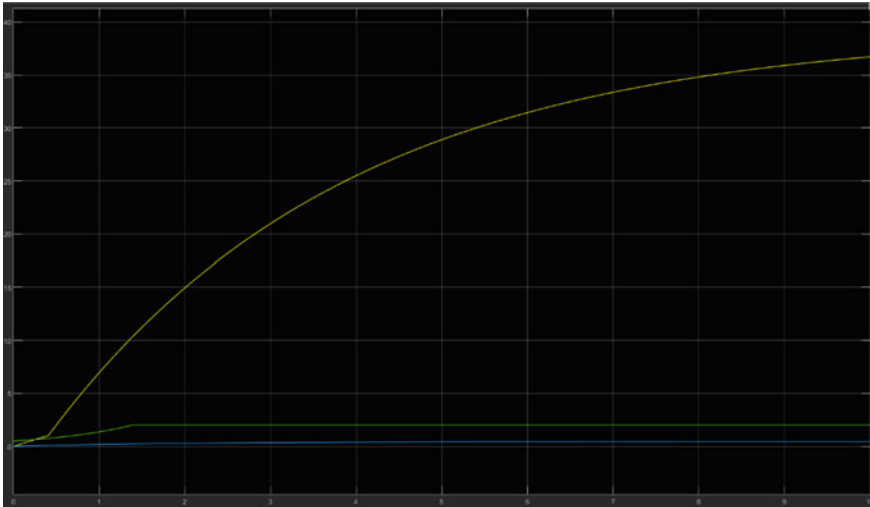


Fig. 1 Simulink graph for FLC type-1, type-2 and without FLC

system that did not use the FLC, while the system that utilized a controller has the best step response performance. The subsystems that employed the fuzzy logic controller had the fastest reaction time in the simulation when compared to the system without a controller. Tables 2, 3, and 4 were tabulated to allow for an easier comparison of the results.

Thus, from the results obtained from the simulation that had been done in the Simulink, it can be seen that there is a significant difference in rise time, T_r value between Tables 2, 3 and 4. There is also difference in value of settling time, T_s in which FLC type 2 give a much lower value than both FLC type 1 and without FLC. From here it can safely assume that DC motor with FLC type 2 offers a much lower rise time value in which in this case the value is 1.034 s and lower settling time of 0.01 s which shows that when DC motor system is paired with FLC type 2 it takes

Table 2 Result of DC motor with FLC type-1

Metrics	Value
Rise time (s)	6.053
Settling time (s)	0.02
Overshoot (%)	None

Table 3 Result of DC motor with FLC type-2

Metrics	Value
Rise time (s)	1.034
Settling time (s)	0.01
Overshoot (%)	0.505

Table 4 Result of DC motor without FLC

Metrics	Value
Rise time (s)	6.015
Settling time (s)	0.02
Overshoot (%)	None

a much shorter time for the system to response thus giving the DC motor system a much better performance in maintaining the speed and the voltage that set same for each types of FLC during the whole simulation.

4 Conclusion

In this paper we had managed to perform the comparison between type 1 and type 2 Fuzzy Logic Controller, simulation has been done in MATLAB/SIMULINK via FIS editor and the speed and voltage of the DC motor has been manipulated in order to see reaction of the voltage that is used by the DC motor. Based on the graph of the simulation, we can verily see the comparison between both type 1 and type 2 FLC and non-FLC DC motor and see that FLC type 2 outperform the FLC type 1 with lower settling time and lower rise time.

Acknowledgements The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2018/ICT02/UMP/02/7 (University reference RDU190178) and Universiti Malaysia Pahang for laboratory facilities.

References

1. Naik KA, Gupta CP (2017) Performance comparison of type 1 and type 2 fuzzy logic system. In: 2017 4th international conference on signal processing, computing and control (ISPC). IEEE, pp 72–76
2. Afshar MR, Alipouri Y (2017) A type 2 fuzzy set model for contractor prequalification. *Autom Constr* 84(2):347–353
3. Ahmed AH, Ali AM (2019) Comparison between fuzzy logic and PI control for the speed of BLDC motor. *Int J Power Electron Drive Syst (IJPEDS)* 9(3):1117–1124
4. Liu J, Zhang P, Wang F (2009) Real-time DC servo motor position control by PID controller using labview. In: International conference on intelligent human-machine system and cybernetics, pp 206–209
5. Elias N, Yahya NM (2019) Comparison of DC motor position control simulation using MABSA-FLC and PSO-FLC. In: 15th international colloquium on signal processing & its applications (CSPA), Penang, pp 39–42
6. Elias N, Yahya NM (2018) Simulation study for controlling direct current motor position utilizing fuzzy logic controller. *Int J Autom Mech Eng* 15(4):5989–6000

7. Parpinelli R, Lopes HS (2011) New inspirations in Swarm Intelligence: a survey. *Int J Bio-Inspired Comput* 3(1):1–16
8. Xue F, Cai Y, Cao Y, Cui Z, Li F (2015) Optimal parameter settings for bat algorithm. *Int J Bio-Inspired Comput* 7(2):125–128
9. Azlan NA, Yahya NM (2019) Modified adaptive bats sonar algorithm with doppler effect mechanism for solving single objective unconstrained optimization problems. In: 15th international colloquium on signal processing & its applications (CSPA), Penang, pp 27–30
10. Krishnan R (2017) Permanent magnet synchronous and brushless DC motor drives. CRC Press, Natick
11. Xia CL (2012) Permanent magnet brushless DC motor drives and controls. Wiley, Singapore

Modelling and PID Control of SEPIC-Boost Converter for 48 V Miniature Mild Hybrid System



M. N. Mamat, M. N. Abdullah, S. Kaharuddin, and D. Ishak

Abstract Modelling and PID control of SEPIC-Boost converter for 48 V automotive miniature mild hybrid system (MHS) is presented in this paper. The proposed design utilizes single switching technique shared by both SEPIC and boost converter in its topology. PID feedback control based on second method Ziegler-Nichols Tuning Rule is used to provide an efficient feedback control to eliminate steady-state error, transient overshoot and to provide fast settling time as its key characteristics. The analysis results show that the proposed topology satisfactorily generated well-regulated output waveforms with ripple less than 10% with excellent PID feedback control.

Keywords SEPIC-Boost · Mild hybrid system · Ziegler-Nichols Tuning Rule

1 Introduction

The automotive world is pushing rapidly towards full electrification, but it will be many years more before full electric vehicle (EV) to monopoly the demand in the automotive industry. Despite the expensive production cost and high maintenance issues, it is no doubt that full electric vehicle offers a lot of advantages especially in reducing greenhouse gases and realize higher performance compared to its internal combustion engine rival. To combine the goodness of internal combustion engine and electric vehicle, automakers have developed many types of hybrid and plug-in hybrid as low-cost solution towards full electrification. One of the common hybrids

M. N. Mamat (✉) · M. N. Abdullah · S. Kaharuddin · D. Ishak
School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Pulau Pinang, Malaysia
e-mail: eenadzri@usm.my

M. N. Abdullah
e-mail: emmnazir@usm.my

S. Kaharuddin
e-mail: suardi@usm.my

D. Ishak
e-mail: dahaman@usm.my

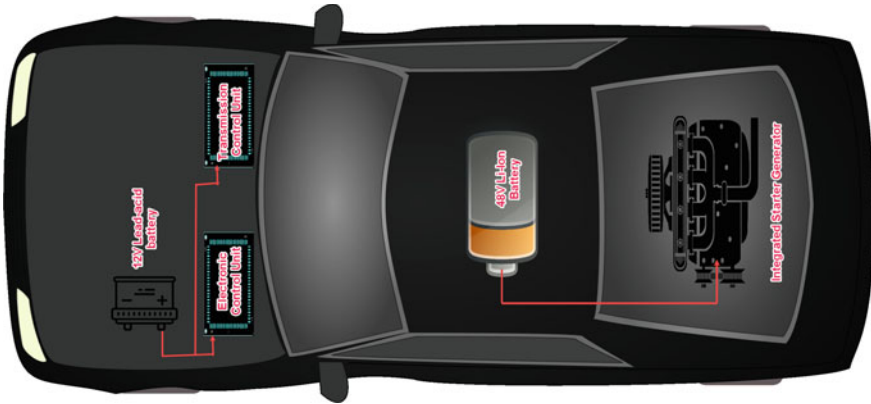


Fig. 1 Conventional automotive mild hybrid system with 12 V and 48 V supply

used nowadays is 48 V mild hybrid system (MHS) where it has been widely used to improve the vehicle performance without negatory of the fuel consumption [1, 2]. A typical 48 V MHS consists of a small electric motor which often known as integrated starter generator (ISG) connected to the drivetrain of an internal combustion engine. ISG assists the internal engine at low speed especially during starting and in-town driving. Some benefits of MHS are better fuel efficiency, improvement in stop-go drive functionality, low production cost and optimal weight reduction [3–5].

Typically, hybrid vehicle uses two different voltage levels to operate namely 12 V and 48 V. The 12 V supply from regular lead-acid battery is used to power the powertrain and drivetrain circuitry through electronic control unit (ECU) and transmission control unit (TCU) whereas the 48 V supply is used to drive the ISG. Figure 1 shows the conventional diagram of a hybrid car operated using 48 V MHS. Typically, ISG is directly controlled by the 48 V battery pack to perform the hybrid task.

In this paper, miniature mild hybrid system operated from a shared 12 V battery is proposed using energy conversion technology. The major advantage of this proposed system is, it uses only one level of voltage to power the vehicle circuitry and the MHS system. An efficient DC/DC converter to convert 12 V to 48 V is introduced in this design which integrates PID control in its closed-loop feedback control. Cascading the proposed design to form a 48 V DC link may be a promising solution to drive the commercial ISG that needed huge ampacity in its operation. The proposed design is a step-up topology where 12 V from the lead-acid battery will be stepped up to 48 V. Many step up topologies where prominent characteristics such as non-isolated, isolated, unidirectional, bidirectional, soft switching and hard switching have been investigated in these articles [6–12] to assist in selecting the best topology for the proposed design stated earlier. It is found out that single-ended primary inductor converter (SEPIC) is the most suitable topology to be used as the converter type. SEPIC has great ability to generate output voltage which is independent of the input voltage state and without polarity reversal. Furthermore, SEPIC is very suitable to be applied in battery storage system where fluctuation in its voltage level is inevitable.

2 Proposed Converter for Mild Hybrid System

2.1 DC/DC Converter Design

The proposed design is based on combination of SEPIC and boost converter where it can produce output voltage at equal, less, or greater than its input voltage. Figure 2 shows the proposed design which uses non-coupled inductors in its topology. It consists of one N-Channel MOSFET switch, two identical inductors L_i and L_s , two ultra-fast recovery diodes D_s and D_o and three capacitors C_s , C_{o1} and C_{o2} . C_s acts as a buffer capacitor to let the generated output voltage be invulnerable regardless of its input voltage level. It is also known as SEPIC coupling capacitor. C_{o1} acts as the boost capacitor while C_{o2} is the output capacitor for the 48 V MHS. The static gain, M for the system is 4.0 and it is considered as a standard static gain.

This proposed design is optimized for MHS application where the rated output voltage is 48 V. To model the proposed converter, the parameter associated in the design are calculated using equations based on the following black box specifications:

- Input voltage: $V_i = 12$ V.
- Estimated efficiency: $\eta = 90\%$
- Total output power: $P_o = 96$ W.
- Rated output: $V_{o2} = 48$ V, $I_o = 2$ A.
- Switching frequency: $f_{sw} = 100$ kHz.

Considering idealities and neglecting the small diode’s voltage drop, the duty cycle of the output voltage can be calculated using the following equation:

Duty cycle,

$$D = \frac{V_{o2} - V_i}{V_{o2} + V_i} \tag{1}$$

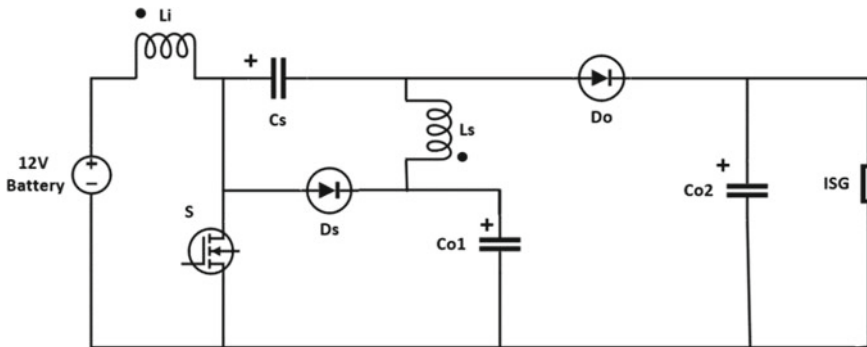


Fig. 2 Proposed SEPIC-Boost converter

The inductor L_i and L_s can be calculated using:
 L_i and L_s inductance,

$$L_i = L_s = \frac{V_i D}{\Delta I_L f_{sw}} \quad (2)$$

where ΔI_L is the peak-to-peak ripple current flowing through the inductor L_i .

For SEPIC converter, the recommended value for ΔI_L is approximately 40% of the maximum input current.

Peak-to-peak ripple current,

$$\Delta I_L = I_i * 0.4 \quad (3)$$

The SEPIC coupling capacitor, C_s and C_{o1} are having similar peak-to-peak ripple voltage. The C_s and C_{o1} capacitance can be calculated using the following equation:

Capacitor C_s and C_{o1} ,

$$C_s = C_{o1} = \frac{I_o}{\Delta V_c f_{sw}} \quad (4)$$

where ΔV_c is the peak-to-peak capacitor voltage ripple and I_o is the total output current. Practically, the ripple value is set between 5 to 10% of the nominal voltage at C_{o1} .

Voltage ripple

$$\Delta V_c = \frac{V_i}{1 - D} * k; \quad k = 0.05 \text{ to } 0.1 \quad (5)$$

The output capacitor for ISG can be calculated using the following equation:

Output capacitor,

$$C_{o2} = \frac{D}{R \left(\frac{\Delta V_{o2}}{V_{o2}} \right) f_{sw}} \quad (6)$$

where ΔV_{o2} is the output ripple voltage at ISG and usually set between 1 to 5% of the output voltage. R is the load impedance of ISG. Table 1 simplifies the calculated parameter used in the proposed design.

2.2 PID Feedback System

The determination of PID gains for converter that integrates MOSFETs and high frequency switching is difficult because the model linearizes to zero. One of the popular techniques to solve this issue is using Ziegler-Nichols Tuning Rule where

Table 1 SEPIC-Boost converter parameters

Parameter	Value	Parameter	Value
Input voltage, V_i	12.0 V	Input ripple current, ΔI_L	3.6A
Input current, I_i	8.9A	Capacitance, C_s and C_{o1}	13.3 μ F
Output voltage, V_{o2}	48.0 V	Capacitor voltage ripple, ΔV_c	1.5 V at k = 0.05
Switching frequency, f_{sw}	100 kHz	Capacitance, C_{o2}	50 μ F
L_i and L_s inductance	20.2 μ H	Equivalent Load	Load = 24 Ω

K_P , K_I and K_D values are estimated using transient step response of a plant [13, 14]. In this proposed design, the Second Method Ziegler-Nichols Tuning Rule is applied to determine the PID gains due to the existence of signal overshoot during the transient period. PID controller block is represented by a combination of three gains which are proportional, integral and derivative gain. The relation between the PID gains is shown in Eq. (7):

$$G_{PID}(s) = K_P + K_I \frac{1}{s} + K_D s = K_P \left(1 + \frac{1}{T_I s} + T_D s \right) \tag{7}$$

where:

- K_P is proportional gain.
- K_I is integral gain.
- K_D is derivative gain.
- T_I is integral time.
- T_D is derivative time.

Table 2 shows the gain estimator chart for Second Method Ziegler-Nichols Tuning Rule where K_U is the ultimate gain and T_U is the steady state oscillation period.

Based on the initial values from open loop response, the selected initial K_P value is 0.008 with oscillation period, T_U equals to 7 ms. The final satisfactory fine tune values for K_I and K_D are 14 and 5.6×10^{-6} , respectively with filter coefficient, N of 100. Figure 3 shows the complete schematic for the simulation using Matlab/Simulink.

Table 2 Z-N gain estimator chart

Controller Type	K_P	T_I	T_D
P	0.50 K_U	∞	0
PI	0.45 K_U	$T_U/1.2$	0
PID	0.60 K_U	$T_U/2$	$T_U/8$

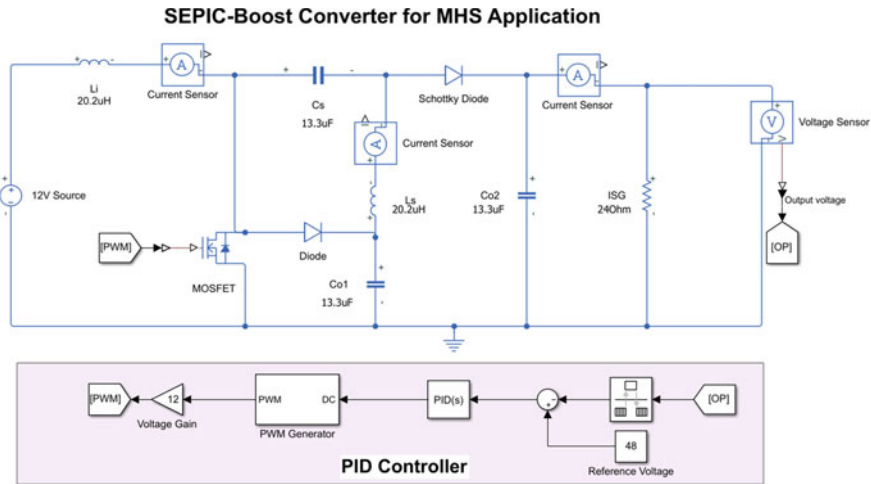


Fig. 3 Simulink schematic of the proposed SEPIC-Boost converter for MHS application

3 Simulation Results and Discussion

The input voltage is set at 12 V with 100 kHz switching frequency applied to the system as depicted in Fig. 3. After the simulation has completed, it is found out that the simulated output waveforms at ISG conclusively show the expected 48 V voltage and 2 A current. Simulation also shows that the design able to operate effectively with duty cycle optimized at 0.6 producing a very minimal peak-to-peak ripple for output voltage which is less than 10%. Figures 4 and 6 show the output waveforms comparison when the proposed design is operated under open loop and PID closed-loop control, respectively. Obviously under PID feedback control, the proposed design exhibits no overshoot and excellent regulation with small peak-to-peak (p-p) ripple in its waveforms.

In Fig. 4, the design suffers extremely high overshoot both for voltage and current waveforms during start-up under open loop system. The voltage overshoot is measured at 82.0 V which is equal to 76.3% whereas the current overshoot is tipped at 3.4A. The settling time for the open loop system is 5.5 ms. Running under steady state condition, the output waveforms cannot optimally regulate as shown in Fig. 5.

Although the peak-to-peak (p-p) ripple for voltage is 2.4 V and 0.09A for current (which is less than 10%), the expected theoretical output waveforms cannot be met using similar duty cycle as in closed loop system. The root mean square (RMS) for the output voltage is 46.5 V and 1.94A for the output current indicating a poor regulated system when operated under open loop control.

Figure 6 shows the output waveforms under PID feedback control. Clearly, the generated output waveforms are excellently regulated with RMS value of voltage at 48.1 V and 2.0A for current. The settling time measured is 6.2 ms which is slightly

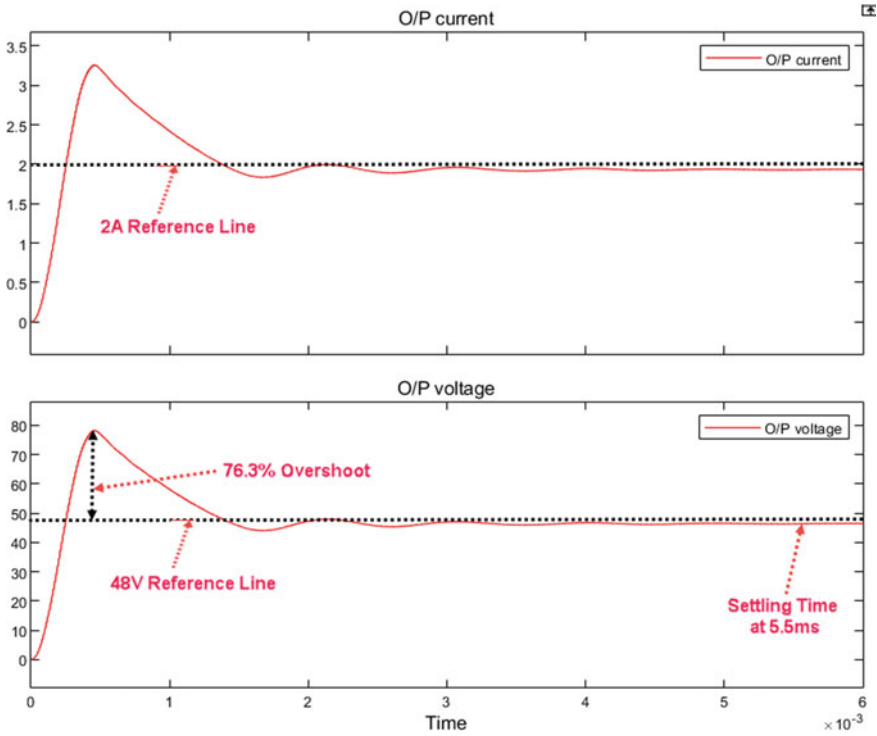


Fig. 4 Output waveforms under open loop control (top) Output current at ISG in ampere (bottom) Output voltage at ISG in volt

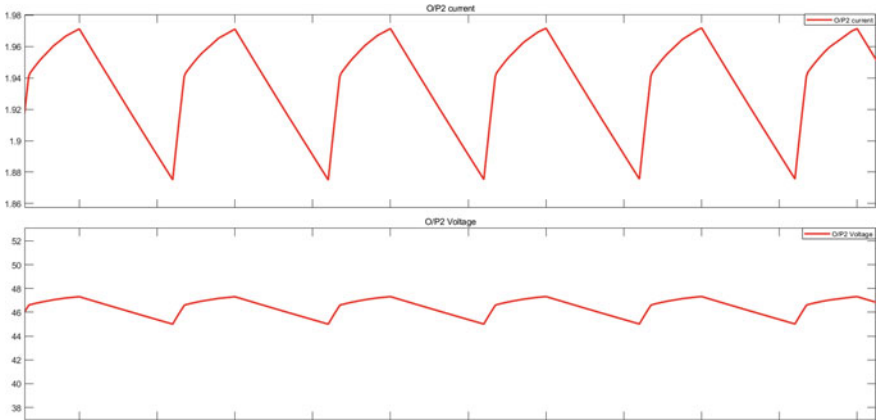


Fig. 5 Peak-to-peak ripple waveforms under open loop control. (top) Output ripple current at ISG in ampere (bottom) Output ripple voltage at ISG in volt

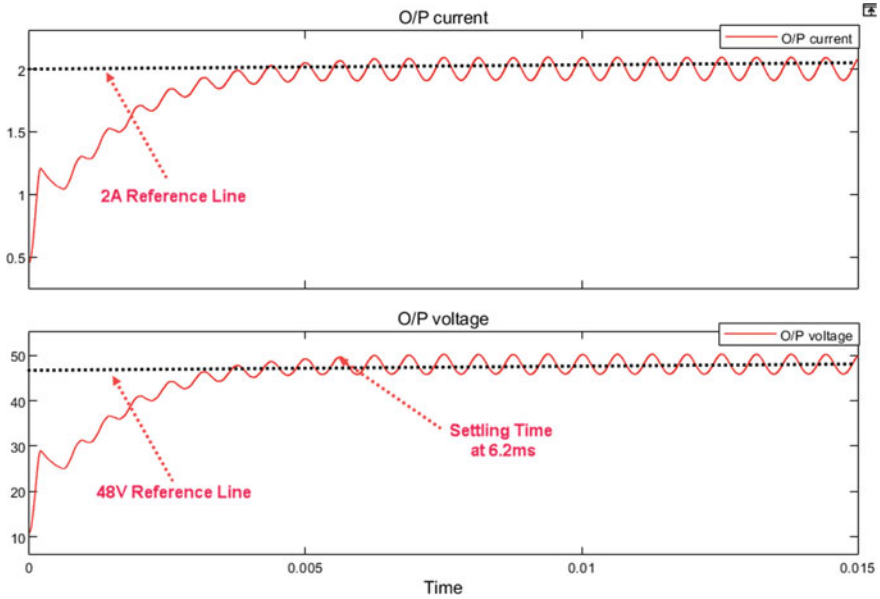


Fig. 6 Output waveforms under PID feedback control. (top) Output current at ISG in ampere (bottom) Output voltage at ISG in volt

Table 3 Dynamic characteristics of the proposed SEPIC-Boost

	Settling time (ms)	Percent Overshoot	Rise time (ms)	Steady-state error
Open loop	5.50	76.3%	0.46	3.1%
PID feedback	6.20	—	1.68	—

slower compared to the open loop system which is normal for PID feedback system to initially adjust its dynamic response.

Table 3 shows the comparison for the dynamic characteristics of the proposed SEPIC-Boost converter under open loop and PID closed-loop control.

Figure 7 shows the peak-to-peak ripple of the output waveforms measured at 4.6 V for the voltage and 0.2A for the current, respectively. The ripple is less than 10% indicating that the proposed design behaves extremely well with the PID feedback control.

4 Conclusion

A 48 V SEPIC-Boost converter with PID feedback control for miniature mild hybrid system is designed in this paper. The converter is based on the conventional SEPIC,

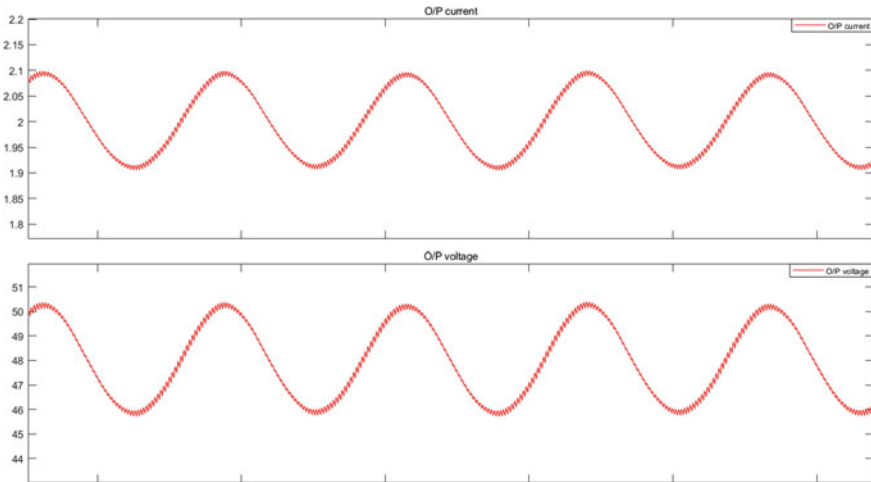


Fig. 7 Peak-to-peak ripple waveforms under PID feedback control. (top) Output ripple current at ISG in ampere (bottom) Output ripple voltage at ISG in volt

but modification based on boost converter is added in the design to broaden its input–output characteristics. By combining the key characteristics from both topologies, the proposed design able to produce well-regulated output waveforms using similar 12 V input supply. Operated by only single switching MOSFET makes the complexity of the proposed topology to be very simple. Furthermore, with the integration of PID feedback control, the system reacts very well resulting in no overshoot and fast settling time in its output waveforms. Finally, this SEPIC-boost converter has been successfully simulated and able to generate well-regulated 48 V, 2A outputs for miniature mild hybrid system with estimated efficiency of 90%.

References

1. Kowalec S (2017) 12V/48V Hybrid Vehicle technologies. In: Applied Power Electronics Conference, pp 1–20
2. Cardoso DS, Fael PO, Espírito-Santo A (2020) A review of micro and mild hybrid systems. *Energy Rep* 6(9):385–390
3. Mckay B (2016) 48V architectures for enabling high efficiency. In: Driving innovation workshop
4. Lee S, Cherry J, Safoutin M, Neam A, McDonald J, Newman K (2018) Modeling and Controls development of 48V Mild Hybrid Electric Vehicles. In: SAE technical paper, vol 2018-April, pp 1–15
5. Lammers MFA (2006) Impact of mild-hybrid functionality on fuel economy and battery lifetime. In: Master thesis report, Technical University of Eindhoven
6. Kummara VGR et al (2020) A comprehensive review of DC–DC converter topologies and modulation strategies with recent advances in solar photovoltaic systems. *Electronics* 9(1):1–41

7. Yadav A, Verma A (2020) Sepic DC-DC converter: Review of different voltage boosting techniques and applications. In: Second international conference on innovative mechanisms for industry applications, pp 733–739
8. Farooq A, Malik Z, Sun Z, Chen G (2015) A review of non-isolated high step-down DC-DC converters. *International Journal of Smart Home* 9(8):133–150
9. Forouzesh M, Siwakoti YP, Gorji SA, Blaabjerg F, Lehman B (2017) Step-Up DC-DC converters: A comprehensive review of voltage-boosting techniques, topologies, and applications. *IEEE Transaction on Power Electronics* 32(12):9143–9178
10. Tytelmaier K, Husev O, Veligorskiy O, Yershov R (2016) A review of non-isolated bidirectional DC-DC converters for energy storage systems. In: International young scientists forum on applied physics and engineering, pp 22–28
11. Dilep G, Singh SN (2017) Selection of non-isolated DC-DC converters for solar photovoltaic system. *Renew Sustain Energy Rev* 76(4):1230–1247
12. Jagadeesh I, V. Indragandhi V (2019) Review and comparative analysis on DC-DC converters used in electric vehicle applications. In: IOP conference series: Materials science and engineering, vol 623(1), pp. 1–15
13. Varsha R, Mukundaswamy MS (2020) Design and simulation of closed loop DC/DC synchronous SEPIC converter using PID Feedback controller. *International Research Journal of Engineering and Technology* 7(7):4510–4516
14. Patel VV (2020) Ziegler-Nichols tuning method. *Resonance* 25(10):1385–1397

Particle Swarm Optimization with Multi-Chaotic Scheme for Global Optimization



Wy-Liang Cheng, Koon Meng Ang, Cher En Chow, Wei Hong Lim, Sew Sun Tiang, Chun Kit Ang, Balaji Chandrasekar, and Eik Cheng

Abstract Particle swarm optimization (PSO) is widely implemented as an optimizer due its characteristics of simple implementation and fast convergence speed. However, it has tendency to cause premature convergence and population diversity loss if the population is not well-initialized due to its randomness. In this research, a PSO with Multi-Chaotic Scheme (PSOMCS) is introduced to generate promising initial population to improve the population diversity and fitness of the candidate solution. Multiple chaotic system is dependent on the initial condition that can cover a broader region of the search space. By using multiple chaotic maps, the introduced method is able to solve different types of problems effectively as each type of chaotic maps has better performance in solving a specific problem. The performance comparisons of PSOMCS and the existing PSO variants are conducted by using the test functions of CEC 2014. The simulation analyses reported that PSOMCS outperforms its competitors with respect to total mean fitness.

Keywords Particle Swarm Optimization · Multi-Chaotic Scheme · Oppositional-based learning · Modified initialization scheme

1 Introduction

Metaheuristic Search Algorithm (MSA) is one of the approachable methods being applied as a decision-making tool due to its characteristic of independent from the gradient information in traditional methods [1]. This characteristic is crucial in optimization tasks because most real-world applications consist of black box optimization problems that does not contain gradient information [2]. MSAs can be categorized based on their search mechanisms and inspiration sources [3], known as (i)

W.-L. Cheng · K. M. Ang · C. E. Chow · W. H. Lim (✉) · S. S. Tiang · C. K. Ang · E. Cheng
Faculty of Engineering, Technology and Built Environment, UCSI University,
56000 Kuala Lumpur, Malaysia
e-mail: limwh@ucsiuniversity.edu.my

B. Chandrasekar
Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology,
Chennai 603203, Tamil Nadu, India

swarm intelligence (SI) algorithms motivated by the group behavior of animal characteristics as in [4–6], (ii) evolutionary algorithms influenced by Darwin’s theory of evolution as in [7–9], (iii) human-based algorithms inspired by multiple human characteristics like thinking, socializing and learning as in [10–12], and (iv) physics-inspired algorithms inspired by the usage scientific method like chemistry, physics or mathematics fundamental as in [13–15]. Based on various objectives, several MSAs variants were established to handle various types of optimization problems more robustly as introduced in [16–24].

Particle Swarm Optimization (PSO) [25] is motivated by the food sources searching behavior of bird’s flock. PSO are widely investigated and proposed with different variants due to its advantage of fast convergence rate and simple execution. However, premature convergence tends to be caused by the disproportional of exploration and exploitation searches. Uncontrol value for exploration tends to decrease the convergence speed of candidate solutions towards the promising regions, whereas the rapid diversity loss in population is caused by excessive exploitation. Furthermore, the qualities, in terms of fitness and diversity levels, of the candidate solutions in initial population are crucial in affecting the optimization performance of PSO [26]. The conventional initialization method generates initial population based on random guess that does not consider the feature of current search environment [27]. This behavior has tendency to cause the candidate solutions being initialized at local optima or inferior regions that are away from global optimum that affects the solution accuracy and convergence rate of the optimizer [27].

To improve the quality of initial population of optimization algorithm, a new PSO variant with Multi Chaotic Scheme (PSOMCS) is introduced in this article. The concept of chaotic system is first adopted to initialize a population with improved diversity level, followed by the implementation of oppositional based learning (OBL) to generate population with broader coverage of search space. The main contributions of the proposed methods are as follows:

- A modified initializations scheme is introduced by hybridizing the fundamental of multiple chaotic systems and OBL to generate the initial population with improved qualities.
- Four different chaotic maps are designed in the modified initialization scheme to initialize the population using different chaotic maps based on a random chaotic variable.
- Extensive performance evaluation is conducted to discover the adequate of PSOMCS in solving all CEC 2014 test problems.

The remaining section of this article are structured as follows. The search mechanisms of conventional PSO are described in Sect. 2. The search mechanisms of proposed PSOMCS are described in Sect. 3. The simulation settings and performance comparisons are reported in Sect. 4. Lastly, the conclusion and future works are described in Sect. 5.

2 Conventional PSO

The conventional PSO [25] introduced in year 1995 adopts the searching method that imitates collective behaviors of bird's flock and fish in locating food sources. Each particle in the population defines a potential solution to the optimization problem. Each i -th particle contain velocity vector $V_i = [V_{i,1}, \dots, V_{i,d}, \dots, V_{i,D}]$ and position vector $X_i = [X_{i,1}, \dots, X_{i,d}, \dots, X_{i,D}]$, where $i \in [1, I]$ and $d \in [1, D]$. Particularly, d denotes the dimensional index and D is the total decision variable number of the solution. The global best position $G_{best} = [G_{best,1}, \dots, G_{best,d}, \dots, G_{best,D}]$, known as the current best solution located during the optimization process, whereas the personal best position of $P_{best} = [P_{best,i,1}, \dots, P_{best,i,d}, \dots, P_{best,i,D}]$ defines the fittest solution of i -th particle found so far. The new velocity $V_{i,d}^{new}$ and new position $X_{i,d}^{new}$ of each i -th particle in every d -th dimension are computed as follows:

$$V_{i,d}^{new} = \omega V_{i,d} + c_1 r_1 (P_{best,i,d} - X_{i,d}) + c_2 r_2 (G_{best,i,d} - X_{i,d}) \quad (1)$$

$$X_{i,d}^{new} = X_{i,d} + V_{i,d} \quad (2)$$

where ω refers to inertia weight; c_1 and c_2 are acceleration coefficients; r_1 and r_2 are two uniform distribution stochastic number in range of 0 to 1. The fitness value of the new position of i -th particle is assessed and competed with its personal best position as well as the global best position. If the former one is fitter, the latter two positions will be replaced. The optimization process of PSO is iterated until the termination condition is fulfilled and the final G_{best} will be used to answer the problem that has encounter.

3 PSOMCS

As mentioned in previous section, the conventional initialization method randomly initializes the population without considering the environment information of the search space. This behavior tends to affect the precision of the initial solutions being generating in inferior regions of the search space. The convergence rate of the optimizer can also be risked due to the initial particles are located far away from the global optimum. In order to overcome the abovementioned drawbacks, a multi-chaotic scheme is designed into the proposed PSOMCS to generate the initial population with better fitness and population diversity.

During the population initialization stage, the particles with improved solution diversity are generated by leveraging the stochasticity and non-iteration attributes of a chaotic map. Define γ_0 as an initial condition of a chaotic variable that randomly generated in every independent run. Meanwhile, γ_x is define as the chaotic variable

generated by the selected chaotic maps at x -th sequence, where $x = 1, \dots, X$ and X refers to the maximum sequence number. The chaotic map is selected to generate the chaotic variable based on 4 conditions: (i) Circle maps if selected if $\gamma_0 \leq 0.25$, (ii) Gauss map is selected if $0.25 \leq \gamma_0 < 0.5$, (iii) Singer map is selected if $0.5 \leq \gamma_0 < 0.75$, and (iv) Sinusoidal map is selected if $0.75 \leq \gamma_0 < 1$. The Circle, Gauss, Singer and Sinusoidal maps selected to generate chaotic population of PSOMCS are described as follows [28], respectively:

$$\text{Circle (CIR): } \gamma_{z+1} = \text{mod}(\gamma_z + b - (a/2\pi) \sin(2\pi\gamma_z), 1) \quad (3)$$

$$\text{Gauss (GAU): } \gamma_{z+1} = \begin{cases} 1 & , \gamma_z = 0 \\ 1/\text{mod}(\gamma_z, 1) & , \text{otherwise} \end{cases} \quad (4)$$

$$\text{Singer (SGR): } \gamma_{z+1} = \mu(7086\gamma_z - 23.32\gamma_z^2 + 28.75\gamma_z^3 - 13.302875\gamma_z^4), \mu = 1.07 \quad (5)$$

$$\text{Sinusoidal (SND): } \gamma_{z+1} = a\gamma_z \sin(\pi\gamma_z), a = 2.3 \quad (6)$$

Suppose that $X_{i,d}^{\min}$ and $X_{i,d}^{\max}$ represent the lower and upper boundaries of i -th chaotic population, respectively, in the d -th dimension, where $d = 1, \dots, D$ and $i = 1, \dots, I$. Given the chaotic variable γ_x produced at the final iteration of x , d -th variable of position vector for the i -th chaotic population member, defined as $X_{i,d}^{CS}$, is formulated as:

$$X_{i,d}^{CS} = \gamma_z(X_{i,d}^{\max} - X_{i,d}^{\min}) + X_{i,d}^{\min} \quad (7)$$

Based on Eq. (7), a chaotic swarm $\phi^{CS} = [X_1^{CS}, \dots, X_i^{CS}, \dots, X_I^{CS}]$ with the swarm size of I is computed.

Although the chaotic map is robust in generating the chaotic population with better population diversity, some particles have tendency to be initialized at the inferior regions that far from the global optimum, leading to the reduction of convergence rate towards the promising zones of the search space. In order to address this issue, the Opposition-Based Learning (OBL) mechanism introduced in [29] is implemented in this framework to produce a better scope of solution space by generating the solution set opposite to ϕ^{CS} . Define X_j^{CS} and X_j^{OBL} as the i -th candidate solution of chaotic population and the corresponding opposite solution of i -th candidate solution produced by OBL, respectively. Given $X_{i,d}^{\min}$ and $X_{i,d}^{\max}$, the value of $X_{i,d}^{OBL}$ associated with the i -th particle in d -th dimension is computed as follows:

$$X_{i,d}^{OBL} = X_{i,d}^{\max} + X_{i,d}^{\min} - X_{i,d}^{CS} \quad (8)$$

Similar with chaotic swarm, an opposite population $\phi^{OBL} = [X_1^{OBL}, \dots, X_i^{OBL}, \dots, X_I^{OBL}]$ with the swarm size of I is generated.

In order to generate an improved preliminary population, the chaotic population ϕ^{CS} and the opposite population ϕ^{OBL} are merged as $\phi^{merged} = [X_1^{merged}, \dots, X_i^{merged}, \dots, X_{2I}^{merged}]$ with the swarm size of $2I$ as follows:

$$\phi^{merged} = \phi^{CS} \cup \phi^{OBL} \quad (9)$$

The objective function value of each member in ϕ^{merged} is assessed. A ranking operator $\Omega(\cdot)$ is employed to reposition all members of ϕ^{merged} from superior to inferior based on their objective function values to produce a ranked population of $\phi^{ranked} = [X_1^{ranked}, \dots, X_i^{ranked}, \dots, X_{2I}^{ranked}]$, known as:

$$\phi^{ranked} = \Omega(\phi^{merged}) \quad (10)$$

Finally, a truncation operator $\gamma(\cdot)$ is adopted to select the top I best particles from ϕ^{ranked} to form the initial population $\phi^{Initial} = [X_1^{ranked}, \dots, X_i^{ranked}, \dots, X_{2I}^{ranked}]$ of PSOMCS as follows:

$$\phi^{Initial} = \gamma(\phi^{ranked}) \quad (11)$$

Based on the positions and fitness values of the solutions in initial population, the personal best position and global best position with their associated fitness values are initialized. During the searching process, the new velocity and position values of each solution are calculated iteratively based on Eqs. (1) and (2), respectively. The objective function value $f(X_i)$ of the new position of each i -th solution in PSOMCS is evaluated and competed with the objective function values of its associated personal best $f(P_{best,i})$ and global best position $f(G_{best,i})$. The latter two positions and their associated fitness values are substituted if the new position X_i^{new} is fitter. Suppose that ε and ε^{\max} refer to the number of fitness evaluations (Fes) consumed and the maximum FEs employed as the termination condition of the searching process. The optimization process is terminated when the termination criteria $\varepsilon > \varepsilon^{\max}$ is fulfilled. The overall framework of PSOMCS algorithm is described in Fig. 1.

4 Comparative Studies

The optimization performance of the proposed PSOMCS is competed with the conventional particle swarm optimization (PSO) [25], particle swarm optimization with sinusoidal map initialization scheme (PSOSND), accelerated particle swarm optimization (APSO) [30] and particle swarm optimization without velocity (PSOWV) [31]. The performance evaluations are conducted using 30 test problems from CEC 2014 [32] at dimension of $D = 30$. For all the compared algorithms, the population size is set as $I = 30$ and the acceleration coefficients are set as

Algorithm 1: PSOMCS	
Input: $I, D, X_i^{\min}, X_i^{\max}, Z, \varepsilon^{\max}$	
01:	Initialize $\Phi^{CS} \leftarrow \{ \}$, $\Phi^{OBL} \leftarrow \{ \}$, $f(G_{best}) \leftarrow \infty$, $\varepsilon \leftarrow 0$;
02:	for each i -th particle do
03:	for each d -th particle do
04:	Randomly generate a chaotic variable $\gamma_0 \in [0,1]$ and set $z = 1$;
05:	if $\gamma_0 < 0.25$ then select Circle map in Eq. (3);
06:	else if $0.25 \leq \gamma_0 < 0.5$ then select Gauss map in Eq. (4);
07:	else if $0.5 \leq \gamma_0 < 0.75$ then select Singer map in Eq. (5);
08:	else if $0.75 \leq \gamma_0 < 1$ then select Sinusoidal map in Eq. (6);
09:	end if
10:	while $z \leq Z$ do
11:	Compute γ_z with the selected chaotic map and $z \leftarrow z + 1$;
12:	end while
13:	Produce $X_{i,d}^{CS}$ and $X_{i,d}^{OBL}$ by using Eqs. (7) and (8), respectively;
14:	end for
15:	Update $\Phi^{CS} \leftarrow \Phi^{CS} \cup X_i^{CS}$ and $\Phi^{OBL} \leftarrow \Phi^{OBL} \cup X_i^{OBL}$;
16:	end for
17:	Produce Φ^{merged} , Φ^{ranked} and $\Phi^{Initial}$ using Eqs. (9) to (11), respectively;
18:	$\varepsilon \leftarrow \varepsilon + 2I$;
19:	for each i -th particle do
20:	Initialize the velocity V_i as zero in all dimensions;
21:	Update $P_{best,i}$, $f(P_{best,i})$, G_{best} and $f(G_{best})$;
22:	end for
23:	while $\varepsilon \leq \varepsilon^{\max}$ do
24:	for each i -th particle do
25:	Compute V_i^{new} and X_i using Eqs. (1) and (2), respectively;
26:	Evaluate $f(X_i)$ and $\varepsilon \leftarrow \varepsilon + 1$;
27:	Update $P_{best,i}$, $f(P_{best,i})$, G_{best} and $f(G_{best})$;
28:	end for
29:	end while
Output: G_{best} , $f(G_{best})$	

Fig. 1 Pseudocode of PSOMCS framework

$c_1 = c_2 = 2.0$. The simulations of all compared optimizers are conducted for 30 independent runs with the maximum number of fitness evaluations $\varepsilon^{\max} = 10000 \times D$ as recommended in [32]. The optimization processes are simulated using Matlab R2020a on a workstation utilized with Intel® Core i7-10875H CPU @ 2.3 GHz. The optimization performance of PSOMCS and all compared algorithms are reported with respect to mean fitness f_{mean} and standard deviation SD , implying the average

search accuracy produced by each independent run and the search consistency of the algorithm in locating the global optimum solution, respectively. The f_{mean} and SD values of PSOMCS and the compared optimizers are reported in Table 1. Particularly, the best f_{mean} value produced for each test function is indicated with boldface, meanwhile the second f_{mean} value is underlined.

Based on Table 1, the proposed PSOMCS is reported to outperform its competitors in solving most of the CEC 2014 test problems by producing the most f_{mean} values. The proposed PSOMCS is reported to have 14 best f_{mean} and 10 second-best f_{mean} values over 30 benchmark functions. Meanwhile, conventional PSO, PSOSND, APSO and PSOWV produce 3, 7, 4 and 2 best f_{mean} values, respectively. For unimodal functions (F1 to F3), the proposed PSOMCS is reported to have two f_{mean} and one second f_{mean} values. For simple multi-modal functions (F4 to F16), the proposed PSOMCS is observed to have best performance among its competitors by producing six best f_{mean} values. Meanwhile, PSOSND is reported to have four best f_{mean} values, implying its competitive performance in tackling multi-modal functions. The proposed PSOMCS has promising performance in solving hybrid functions (F17 to F22) by producing four best f_{mean} and two second best f_{mean} values over six functions, implying that PSOMCS outperforms its competitors. For composite functions (F23 to F30), PSOSND is reported to have the best performance followed by PSOMCS, PSOWV, PSO and APSO which produce 3, 2, 2, 1 and 1 best f_{mean} values, respectively. Although PSOMCS is not outperform PSOSND in tackling composite functions, but PSOMCS is reported to have four second best f_{mean} values, which is competitive among other algorithms. From Table 1, the proposed PSOMCS has shown more competitive search accuracy than that of PSOSND. This observation implies the advantages of using the multiple numbers of chaotic maps over a single chaotic map in initialize the preliminary population with promising solution qualities. The presence of multiple numbers of chaotic maps in PSOMCS has enhanced the robustness of proposed work to tackle the diverse set of optimization problems with different characteristics. The competitive performance of proposed PSOMCS against the APSO and PSOWV also prove that the modification of initialization scheme can serve as a promising approach to achievement notable performance enhancement of PSO.

Table 1 Performance evaluations of the proposed PSOMCS and four optimizers in solving 30 test problems from CEC 2014

<i>f_{mean} (SD)</i>				
	F1	F2	F3	F4
PSO	<u>4.47E+05</u> (5.32E+05)	3.51E+06 (1.07E+07)	3.67E+02 (1.50E+03)	<u>9.73E+01</u> (2.94E+01)
PSOSND	1.23E+06 (2.07E+06)	<u>5.72E+05</u> (1.99E+06)	9.04E-02 (1.77E-01)	8.97E+01 (2.46E+01)
APSO	1.25E+08 (2.02E+07)	3.01E+07 (5.81E+07)	1.97E+05 (3.56E+04)	2.71E+02 (3.29E+01)
PSOWV	4.70E+08 (3.79E+08)	6.94E+10 (1.56E+10)	3.18E+05 (5.70E+04)	8.66E+03 (3.26E+03)
PSOMCS	2.60E+05 (3.61E+05)	2.23E+05 (7.37E+05)	<u>1.31E+00</u> (5.60E+00)	1.03E+02 (3.72E+01)
	F5	F6	F7	F8
PSO	<u>2.01E+01</u> (1.60E-01)	2.88E+01 (3.77E+00)	5.51E-01 (6.85E-01)	1.36E+02 (2.06E+01)
PSOSND	<u>2.01E+01</u> (1.31E-01)	2.96E+01 (4.19E+00)	<u>4.70E-01</u> (6.03E-01)	<u>1.19E+02</u> (1.94E+01)
APSO	2.00E+01 (3.26E-02)	2.54E+01 (3.50E+00)	1.25E+00 (6.16E-01)	1.37E+02 (2.75E+01)
PSOWV	2.09E+01 (9.48E-02)	3.83E+01 (2.41E+00)	5.61E+02 (2.05E+02)	3.87E+02 (3.23E+01)
PSOMCS	<u>2.01E+01</u> (2.54E-01)	<u>2.76E+01</u> (4.96E+00)	1.91E-01 (1.88E-01)	1.14E+02 (1.82E+01)
	F9	F10	F11	F12
PSO	1.49E+02 (2.61E+01)	2.99E+03 (6.56E+02)	<u>3.85E+03</u> (7.53E+02)	<u>5.59E-01</u> (1.81E-01)
PSOSND	<u>1.46E+02</u> (3.10E+01)	2.88E+03 (7.63E+02)	4.16E+03 (7.43E+02)	6.42E-01 (1.58E-01)
APSO	1.98E+02 (2.18E+01)	3.71E+03 (5.04E+02)	4.16E+03 (8.01E+02)	5.61E-01 (2.68E-01)
PSOWV	4.85E+02 (3.67E+01)	7.66E+03 (4.67E+02)	7.56E+03 (4.57E+02)	2.70E+00 (1.22E-01)
PSOMCS	1.36E+02 (3.32E+01)	3.39E+03 (6.06E+02)	3.77E+03 (6.89E+02)	5.03E-01 (1.62E-01)
	F13	F14	F15	F16
PSO	5.33E-01 (1.02E-01)	3.03E-01 (1.60E-01)	2.87E+01 (1.37E+01)	<u>1.21E+01</u> (6.77E-01)
PSOSND	4.34E-01 (8.89E-02)	2.44E-01 (4.40E-02)	3.23E+01 (1.72E+01)	1.22E+01 (6.70E-01)
APSO	5.92E-01 (1.17E-01)	<u>2.57E-01</u> (5.35E-02)	<u>3.16E+01</u> (2.06E+01)	1.28E+01 (6.66E-01)

(continued)

Table 1 (continued)

<i>f_{mean}</i> (<i>SD</i>)				
PSOWV	6.12E+00 (5.58E-01)	1.46E+02 (6.54E+01)	3.58E+06 (7.86E+05)	1.34E+01 (2.05E-01)
PSOMCS	4.93E-01 (7.99E-02)	2.68E-01 (6.42E-02)	4.40E+01 (3.44E+01)	1.19E+01 (7.17E-01)
	F17	F18	F19	F20
PSO	5.89E+04 (9.72E+04)	3.73E+03 (4.97E+03)	2.06E+06 (1.35E+01)	4.68E+02 (4.89E+02)
PSOSND	7.40E+04 (1.14E+05)	3.39E+03 (3.57E+03)	2.41E+01 (1.89E+01)	8.98E+02 (1.47E+03)
APSO	2.05E+06 (2.37E+06)	1.41E+05 (3.12E+05)	2.93E+01 (2.74E+01)	1.02E+05 (4.65E+04)
PSOWV	1.77E+07 (1.08E+07)	5.88E+08 (2.87E+08)	3.64E+02 (1.51E+02)	2.78E+05 (1.88E+05)
PSOMCS	3.98E+04 (2.46E+04)	2.50E+03 (4.48E+03)	2.09E+01 (1.56E+01)	3.18E+02 (2.90E+02)
	F21	F22	F23	F24
PSO	1.82E+04 (1.24E+04)	8.54E+02 (3.28E+02)	3.15E+02 (3.38E-01)	2.33E+02 (7.87E+00)
PSOSND	3.63E+04 (4.85E+04)	8.33E+02 (2.38E+02)	3.16E+02 (1.88E+00)	2.32E+02 (5.91E+00)
APSO	1.16E+06 (9.69E+05)	6.23E+02 (1.73E+02)	3.62E+02 (1.51E+01)	2.50E+02 (2.84E+00)
PSOWV	6.03E+06 (6.96E+06)	1.08E+03 (2.42E+02)	6.74E+02 (8.04E+01)	4.66E+02 (3.90E+01)
PSOMCS	1.80E+04 (1.21E+04)	7.04E+02 (1.46E+02)	3.16E+02 (1.06E+00)	2.33E+02 (7.77E+00)
	F25	F26	F27	F28
PSO	2.18E+02 (7.05E+00)	1.85E+02 (3.61E+01)	9.84E+02 (4.01E+02)	4.57E+03 (7.94E+02)
PSOSND	2.11E+02 (4.66E+00)	1.06E+02 (2.23E+01)	1.15E+03 (2.74E+02)	3.61E+03 (8.62E+02)
APSO	2.22E+02 (2.99E+00)	1.83E+02 (4.59E+01)	8.02E+02 (2.03E+02)	2.50E+03 (5.80E+02)
PSOWV	2.59E+02 (2.21E+01)	1.06E+02 (1.57E+00)	1.28E+03 (3.92E+01)	1.82E+03 (5.12E+02)
PSOMCS	2.13E+02 (3.76E+00)	1.08E+02 (4.08E+01)	9.96E+02 (4.12E+02)	4.16E+03 (7.99E+02)
	F29	F30		
PSO	4.75E+05 (2.12E+06)	4.03E+03 (2.28E+03)		

(continued)

Table 1 (continued)

f_{mean} (SD)				
PSOSND	1.17E+07 (1.30E+07)	<u>3.64E+03</u> <u>(1.22E+03)</u>		
APSO	1.23E+07 (1.88E+07)	9.88E+04 (3.25E+04)		
PSOWV	1.04E+07 (9.14E+06)	3.79E+05 (3.39E+05)		
PSOMCS	7.71E+03 (1.66E+04)	3.05E+03 (1.04E+03)		

5 Conclusion

In this article, an improved PSO variant is introduced by incorporating a multi-chaotic scheme to substitute the conventional population initialization scheme in order to generate a promising preliminary population. By combining the concepts of OBL and various chaotic maps, the proposed multi-chaotic scheme is expected to tackle the premature convergence issue of PSO by producing initial population with better diversity level. Four different chaotic maps are adopted in the multi-chaotic scheme to allow the algorithm in selecting the best chaotic map for certain objective functions. The simulation result reported that the adoption of chaotic maps in initialize population can achieve a better optimization performance compared to conventional initialization scheme, especially when a combination of four chaotic maps is adopted. In overall, the performance evaluations in solving 30 functions from CEC 2014 reported that the proposed PSOMCS has the best performance followed by PSOSND, APSO, conventional PSO, and PSOWV. As the future works, the convergence characteristics of proposed PSOMCS can be investigated. Further modification can be made to the proposed PSOMCS to solve multi-objectives optimization problems, optimization problems with constraints, and real-world engineering application problems. Furthermore, the capabilities of the introduced PSOMCS in optimizing artificial neural networks (ANNs) architectures are worth to be investigated also.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme under Project Proj-FRGS/1/2019/TK04/UCSI/02/1.

References

1. Lee KS, Geem ZW (2005) A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice. *Comput Methods Appl Mech Eng* 194:3902–3933
2. Doncieux S, Mouret J-B (2014) Beyond black-box optimization: a review of selective pressures for evolutionary robotics. *Evol Intel* 7:71–93

3. Mohamed AW, Hadi AA, Mohamed AK (2019) Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm. *Int J Mach Learn Cybernet* 11:1–29
4. Jain M, Singh V, Rani A (2019) A novel nature-inspired algorithm for optimization: squirrel search algorithm. *Swarm Evol Comput* 44:148–175
5. Dhiman G, Kumar V (2017) Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. *Adv Eng Softw* 114:48–70
6. Fausto F, Cuevas E, Valdivia A, González A (2017) A global optimization algorithm inspired in the behavior of selfish herds. *Biosystems* 160:39–55
7. Dhivyaprabha T, Subashini P, Krishnaveni M (2018) Synergistic fibroblast optimization: a novel nature-inspired computing algorithm. *Front Inform Technol Electron Eng* 19:815–833
8. Salimi H (2015) Stochastic fractal search: a powerful metaheuristic algorithm. *Knowl-Based Syst* 75:1–18
9. Civicioglu P (2013) Backtracking search optimization algorithm for numerical optimization problems. *Appl Math Comput* 219:8121–8144
10. Moghdani R, Salimifard K (2018) Volleyball premier league algorithm. *Appl Soft Comput* 64:161–185
11. Shayeghi H, Dadashpour J (2012) Anarchic society optimization based PID control of an automatic voltage regulator (AVR) system. *Electr Electron Eng* 2:199–207
12. Rao RV, Savsani VJ, Vakharia DP (2011) Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. *Comput Aided Des* 43:303–315
13. Kashan AH, Tavakkoli-Moghaddam R, Gen M (2019) Find-Fix-Finish-Exploit-Analyze (F3EA) meta-heuristic algorithm: an effective algorithm with new evolutionary operators for global optimization. *Comput Ind Eng* 128:192–218
14. Kaveh A, Dardas A (2017) A novel meta-heuristic optimization algorithm: thermal exchange optimization. *Adv Eng Softw* 110:69–84
15. Nematollahi AF, Rahiminejad A, Vahidi B (2017) A novel physical based meta-heuristic optimization method known as Lightning Attachment Procedure Optimization. *Appl Soft Comput* 59:596–621
16. Ang KM, Lim WH, Isa NAM, Tiang SS, Wong CH (2020) A constrained multi-swarm particle swarm optimization without velocity for constrained optimization problems. *Expert Syst Appl* 140:112882
17. Ang KM, Lim WH, Isa NAM, Tiang SS, Ang CK, Natarajan E, Solihin MI (2020) A constrained teaching-learning-based optimization with modified learning phases for constrained optimization. *J Adv Res Dyn Control Syst* 12:15
18. Chong OT, Lim WH, Isa NAM, Ang KM, Tiang SS, Ang CK (2020) A teaching-learning-based optimization with modified learning phases for continuous optimization. In: *Science and information conference*. Springer, pp 103–124
19. Choi ZC, Ang KM, Lim WH, Tiang SS, Ang CK, Solihin MI, Juhari MRM, Chow CE (2021) Hybridized metaheuristic search algorithm with modified initialization scheme for global optimization. In: *Advances in robotics, automation and data analytics: selected papers from ICITES 2020*, vol 1350, p 172
20. Suresh S, Elango N, Venkatesan K, Lim WH, Palanikumar K, Rajesh S (2020) Sustainable friction stir spot welding of 6061–T6 aluminium alloy using improved non-dominated sorting teaching learning algorithm. *J Market Res* 9:11650–11674
21. Yao L, Lim WH (2017) Optimal purchase strategy for demand bidding. *IEEE Trans Power Syst* 33:2754–2762
22. Yao L, Chen Y-Q, Lim WH (2015) Internet of things for electric vehicle: an improved decentralized charging scheme. In: *2015 IEEE international conference on data science and data intensive systems*. IEEE, pp 651–658
23. Lim WH, Isa NAM, Tiang SS, Tan TH, Natarajan E, Wong CH, Tang JR (2018) A self-adaptive topologically connected-based particle swarm optimization. *IEEE Access* 6:65347–65366

24. Karim AA, Isa NAM, Lim WH (2020) Modified particle swarm optimization with effective guides. *IEEE Access* 8:188699–188725
25. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *Proceedings of ICNN'95 - international conference on neural networks*, pp 1942–1948, vol 1944
26. Deng Y, Liu Y, Zhou D (2015) An improved genetic algorithm with initial population strategy for symmetric TSP. In: *Mathematical problems in engineering 2015*
27. Tian D, Zhao X, Shi Z (2019) DMPSO: Diversity-guided multi-mutation particle swarm optimizer. *IEEE Access* 7:124008–124025
28. Sprott JC, Sprott JC (2003) *Chaos and time-series analysis*. Citeseer
29. Tizhoosh HR (2005) Opposition-based learning: a new scheme for machine intelligence. In: *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC 2006)*. IEEE, pp 695–701
30. Zhang H, Yang Z (2018) Accelerated particle swarm optimization to solve large-scale network plan optimization of resource-leveling with a fixed duration. In: *Mathematical problems in engineering 2018*
31. El-Sherbiny MM (2011) Particle swarm inspired optimization algorithm without velocity equation. *Egypt Inform J* 12:1–8
32. Liang JJ, Qu BY, Suganthan PN (2013) Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization. Technical report 201311, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Technical report, Nanyang Technological University, Singapore

Flexible Wearable Microstrip Antenna with DGS and Shorting Post for WBAN Application



Liu Yanan, Sew Sun Tiang, Wei Hong Lim, and Xiang Sheng Wong

Abstract A novel flexible microstrip patch antenna with multiple notches, Defected Ground Structure (DGS) and Shorting Post technique (SP) is presented to enhance the antenna performance. The proposed antenna design covers Industrial Scientific Medical (ISM) band, which is 2.45 GHz. A wider impedance bandwidth of around 8.93% (2.32 GHz–2.53 GHz) and return loss of -26.67 dB. The parametric studies showed that the proposed antenna has salient characteristics, low profile and cost-effective compared to existing flexible antennas. The proposed antenna achieves a compact size of area with 19×24 mm. The size and shape are suitable for WBAN application as it has the merits of stable gain and omnidirectional radiation patterns.

Keywords Microstrip · Wearable antenna · Healthcare · WBAN

1 Introduction

A Wireless Body Area Network (WBAN) can be defined as a wireless network of wearable computing devices [1]. It can be used both in medical field, such as healthcare systems and non-medical domain, like the sports and military application [2]. For medical usage, it is introduced to enhance the medical monitoring system, especially the mobility of patients when wearing medical monitoring device. Therefore, a WBAN is made of various sensor nodes and a gateway node which is used to allow connection to external database server. The specialized equipment used in most of the existing medical monitoring system can transfer data using wired or fixed connection such as standard telephone line. However, these systems are not location independent as wired sensors are used. The introduction of wearable antenna can solve the stated problem as the independent monitoring facility are able to be achieved in WBAN [3].

The wearable antennas applications can be categorized into two groups, which are on-body and off-body communication. In on-body communication, the data is

L. Yanan · S. S. Tiang (✉) · W. H. Lim · X. S. Wong
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, Malaysia
e-mail: tiangss@ucsiuniversity.edu.my

Table 1 Frequency band allocation

Frequency bands (MHz)	Maximum EIRP (mW)	Applications
13.5530–13.5670	100/500	SRD/ISM
40.66–40.77	100/500	SRD/ISM
433.05–434.79	50/100	RCD, SD/SRD
863–870	25	SRD/ISM
2 400–2 500	500	SRD/ISM
5 725–5 925	500	ISM
24 000–24 250	500	SRD/ISM
59 300–62 000	500/1000	SRD/ISM
122 020–123 000	500/1000	SRD/ISM
244 000–246 000	500/1000	SRD/ISM

transmitted along the body between body-worn user nodes [4]. There are multiple frequency bands defined for WBAN devices including Medical Implant Communications System band (MICS: 400 MHz), Industrial Scientific Medical Band (ISM: 2.4 GHz and 5.8 GHz) as well as Ultra-Wide Band (UWB: 3 – 10 GHz) [5]. ISM band is widely used in most of the antenna design such as microstrip patch antenna, dipole antenna and planar antenna [6]. Table 1 shows the frequency band allocation.

Antennas are vital component of the WBAN devices [7]. Recently, the rapid development of wearable devices has emphasized the features of flexibility. The key factor of flexibility is using flexible substrates. Fabric materials, FR4 and Polyimide materials are widely used as flexible substrates. Jeans with the permittivity constant 1.7 [8] and felt with the permittivity of 1.22 [9] is chosen as the substrate. In work [10], cotton is used as the substrate. However, these three textile materials have a low permittivity which result in a bulky size as stated in paper [11]. Some researchers proposed FR4 as the substrate as it can achieve compact size [12, 13], however the radiation patterns are omni-directional.

Several problems and challenges have been identified for the antenna design for WBAN. Conventional wearable antennas have been introduced for this application such as microstrip [14–16]. However, microstrip has inherent drawbacks in terms of size and bandwidth [17]. Although Vivaldi antenna exhibits wide-band characteristic [18, 19], the antenna requires larger geometry and is difficult to build. Therefore, this type of antenna is not well-suited for WBAN application. A high gain flexible antenna [9] which is operating from 9.7 to 10.6 GHz has been investigated for healthcare detection and wearable application. Although the design uses fabric materials-zelt as patch and ground, the operation frequency is much higher and does not cover the ISM band. A button like antenna is discussed in the paper [20], the antenna size is smaller as compared with the works mentioned above, it has a good gain and excellent efficiency however the bandwidth is quite narrow. In summary, all the existing designs are a result of trade-off between size and bandwidth. Moreover, microstrip patch antenna has been selected as the baseline design for wearable antenna. One

such drawback is relatively narrow bandwidth of the antenna which can affect the reliability of antenna in terms of data transmission and reception. Due to physical safety, factors such as miniature size and round edge is a matter of concern for the wearable antenna design whereas EM wave and heat dissipation become the operation safety precautions [21]. Thus, modification and shorting post techniques are applied to overcome the challenges. Modification technique is one of the methods to reduce the microstrip patch antenna (MPA) size and this technique can be used to miniaturize the antenna size by modifying their ground plane. There are different modification methods, including reducing the ground plane, inserting different slots in the ground plane, taking an irregular ground structure, make use of defected ground structures (DGSs) and utilize the complementary SRR (CSRR)-based antenna [22]. Modification technique DGS has been integrated on the ground plane with planar transmission line [23–27]. Various novel DGSs have been reported and a lot of applications have been investigated extensively in microwave circuits. Previous studies [28, 29] have showed that using shorting post on the microstrip antenna can reduce the size of the antenna.

In this paper, a novel wearable antenna with DGS and shorting post techniques for bandwidth improvement is proposed. As described in subsequent sections, the initial antenna design and optimization were carried out by simulations. The results are analyzed and discussed in the following sections.

2 Antenna Design

The geometry of the proposed antenna consists of patch radiator on a ground plane with dielectric substrate Polyimide with permittivity of 3.5 and a simple 50-Ω microstrip transmission line. Two shorting posts with the radius of 0.25, are placed on the antenna through the patch to the ground. Figure 1 shows the geometry of the proposed design. The patch radiator is embedded with multiple notches to enhance the bandwidth of the antenna and the size of the proposed antenna is 24 × 19 mm.

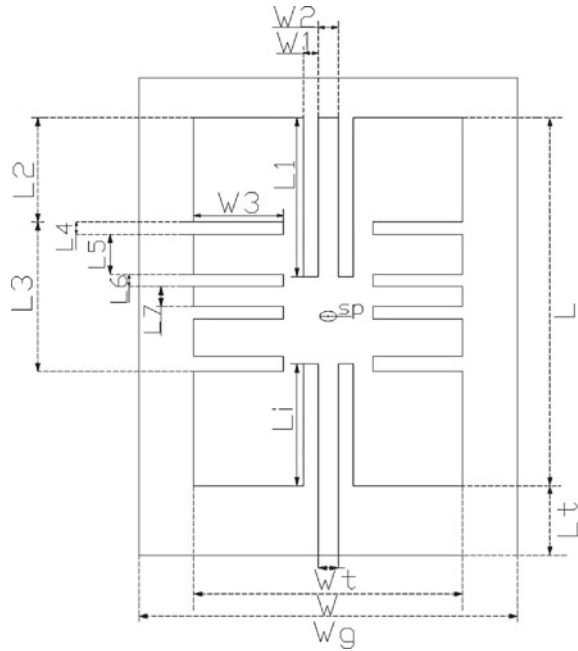
The ground plane is modified with DGS structure for the purpose of bandwidth enhancement. The transmission line of the antenna is connected at the bottom of the antenna from the substrate to the copper patch with the dimension of 3 × 0.75 mm. Inset feed is also introduced with a dimension of 6.5 × 1 mm. The width (W) of the radiating patch is calculated as given equation:

$$W = \frac{c}{2f_0\sqrt{\frac{\epsilon_r+1}{2}}} \quad (1)$$

where, c is the speed of light, f_0 is the resonant frequency and ϵ_r is the dielectric constant of the substrate.

To calculate the length (L) of the patch, the mathematical equation is expressed:

Fig. 1 Geometry of the proposed antenna



$$L = L_{eff} - 2\Delta L \quad (2)$$

$$L_{eff} = \frac{c}{2f_o\sqrt{\epsilon_{reff}}} \quad (3)$$

$$\Delta L = 0.142h \frac{(\epsilon_{reff} + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\epsilon_{reff} - 0.258)\left(\frac{W}{h} + 0.8\right)} \quad (4)$$

$$\epsilon_{reff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left| \frac{1}{\sqrt{1 + 12\left(\frac{h}{W}\right)}} \right| \quad (5)$$

where, L_{eff} is the effective length, ΔL is the length extension and ϵ_{reff} is the effective dielectric constant (Table 2).

Table 2 Optimized parameters of the proposed antenna

Symbol	Dimensions (mm)
Length of antenna, L	19
Width of antenna, W	13
Effective length, L_{eff}	0.035
Length extension, ΔL	1.147
Effective Dielectric Constant, ϵ_{reff}	3.151
Width of ground, W_g	19
Length of ground, L_g	24
Width of inset, W_i	1
Length of inset, L_i	6.5
Width of transmission line, W_t	1
Length of transmission line, L_t	3
Width of notch 1, W_{n1}	0.75
Width of notch 2, W_{n2}	1
Width of notch 3, W_{n3}	4.5
Length of notch 1, L_{n1}	8
Length of notch 2, L_{n2}	5.25
Length of notch 3, L_{n3}	7.5
Length of notch 4, L_{n4}	0.63
Length of notch 5, L_{n5}	2
Length of notch 6, L_{n6}	0.63
Length of notch 7, L_{n7}	1
Radius of SP	0.45
Shorting Post (SP) 1	(0,0)
Shorting Post (SP) 2	(0,8,0)

3 Results

The proposed antenna is in the form of multiple notches, DGS, shorting post and made of polyimide substrate material to achieve its wideband characteristics. DGS is used as an emerging technique for improving the various parameters of antenna design for example its narrow bandwidth, cross polarization, low gain etc. The design of ground plane can be periodic or non-periodic shape. By introducing the single or multiple defects, it affects the current flow and shield current distribution at the ground plane. The etched slots or defects on the ground plane increase the effective capacitance and inductance which varies the distribution of current on the patch. This cutting of edges would result more current paths on the patch and increase the current intensity. Shorting post technique is applied to further increase the bandwidth by changing the current patch. Figure 2 shows the proposed antennas with three different DGS. Figure 3 shows the proposed antennas with three different notches design. Figure 4. shows the proposed antennas with shorting post and without shorting post.



Fig. 2 DGS Design: Design 1 (Left), Design 2 (Middle), Design 3 (Right)

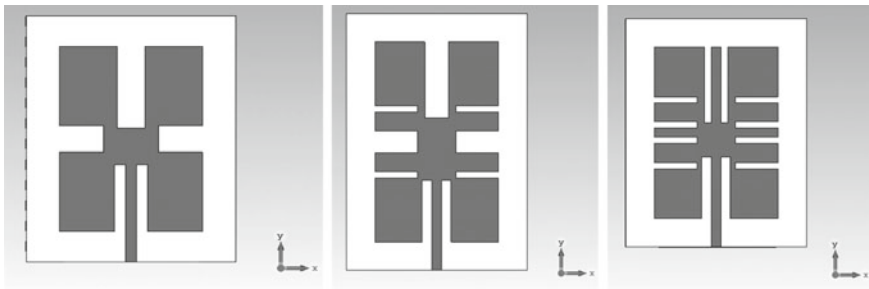


Fig. 3 Notch Design: Design 1 (Left), Design 2 (Middle), Design 3 (Right)

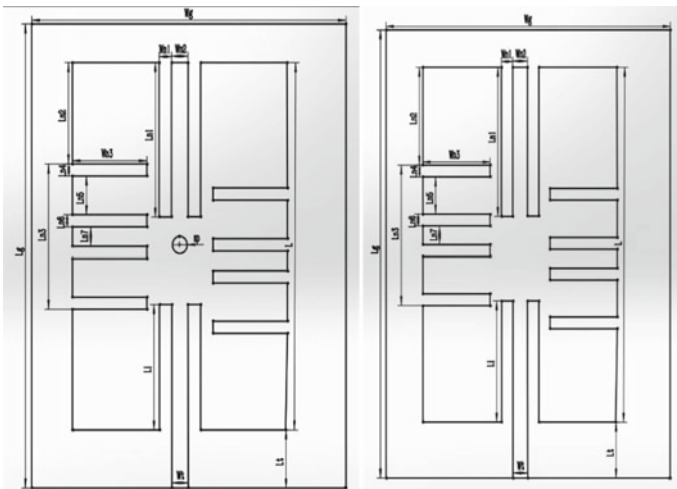


Fig. 4 With shorting post and without shorting post

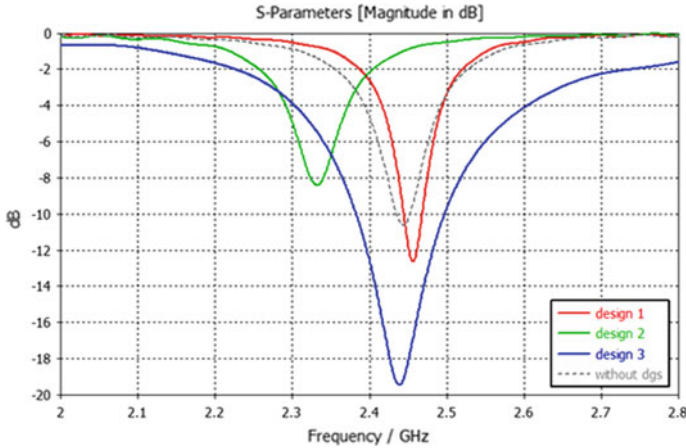


Fig. 5 Reflection coefficients for various notch design

Figure 5 shows the comparison of reflection coefficients for the three designs with different DGS Based on the result obtained. Design 1 contains a sharp resonant frequency with -12 dB at 2.455 GHz with introduction of two etched rectangular slot shapes on the top left and bottom right. Insertion of a circular slot in the middle of the ground plan for Design 2 yields the largest impedance has a better impedance bandwidth of 4.1% with -8 dB at 2.332 GHz. By adding DGS structure, Design 3 can achieve the largest bandwidth of 4.49% among the three designs studied with -19 dB at 2.444 GHz. A ground plane with DGS was selected as the optimum design as it provides significantly wide bandwidth over the frequency range of interest. Apart from that, the result of antenna after adding notches is shown in Fig. 6. Based on the result obtained, Design 1 achieved 2.4059 GHz to 2.5174 GHz with bandwidth of 0.1115 GHz at 4.55%. Design 2 is from 2.395 GHz to 2.5001 GHz with bandwidth of 0.1051 GHz at 4.29%. Design 3 obtained a frequency range of 2.3935 GHz to 2.499 GHz with smaller bandwidth of 4.31%. By comparing with the initial design without notch, the initial bandwidth is only 0.68%.

Figure 7 and Fig. 8 shows the comparison of reflection coefficients for shorting posts with various radius. It can be noticed that with the radius increases, the bandwidth increases. Table 3 shows the bandwidth with various radius. To improve the antenna performance, a parameter sweep was done to find the best position of the shorting post. Move shorting post on X-axis from the from the point of $(-0.3, 0)$ to the point of $(-1.9, 0)$. Move shorting post from the point of $(-0.3, 0)$ to the point of $(-0.7, 0)$, it is found that with the distance between the center to the shorting post increases, the bandwidth and the resonant frequency increases as well. The point of $(-1, 0)$ is a turn point, the bandwidth decreases with the distance increases after this

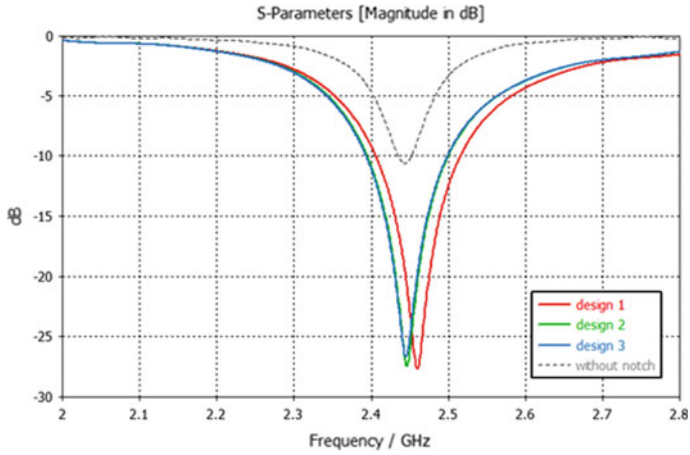
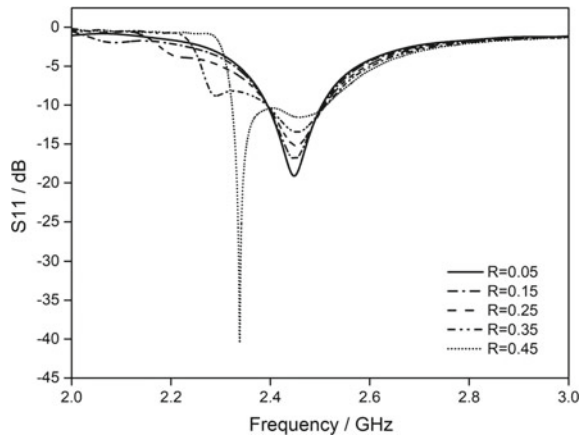


Fig. 6 Reflection coefficients for various DGS design

Fig. 7 Reflection coefficients for shorting posts with various radius



point and the lower resonant frequency of the dual band decreases as well. Figure 9 and Fig. 10 shows the reflection coefficients for shorting post with different locations.

Fig. 8 Reflection Coefficients for shorting posts with various radius

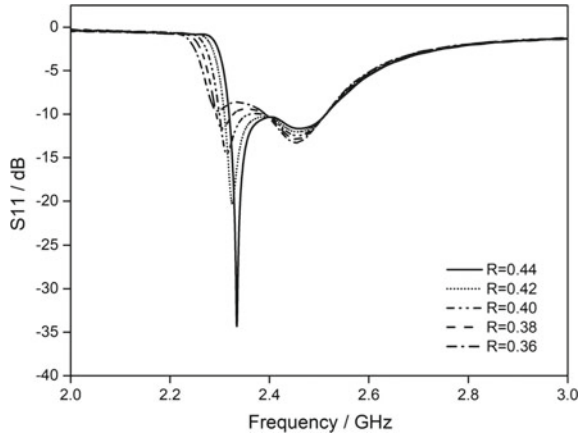
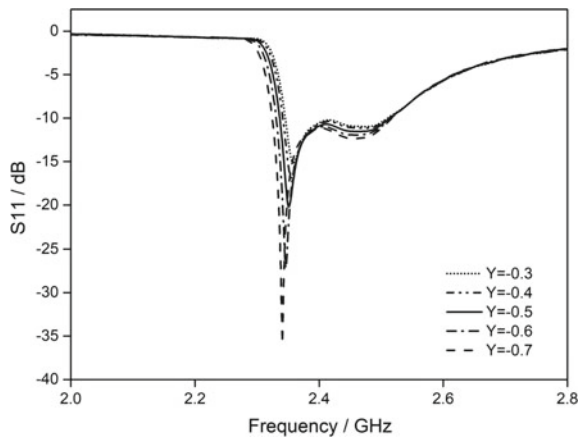


Table 3 Bandwidth with various radius

Radius (mm)	Bandwidth (GHz) (%)	Radius (mm)	Bandwidth (GHz) (%)
0.05	4.83	0.36	5.82
0.15	5.09	0.38	5.81
0.25	5.46	0.4	6.04
0.35	5.84	0.42	9.34
0.45	8.93	0.44	9.08

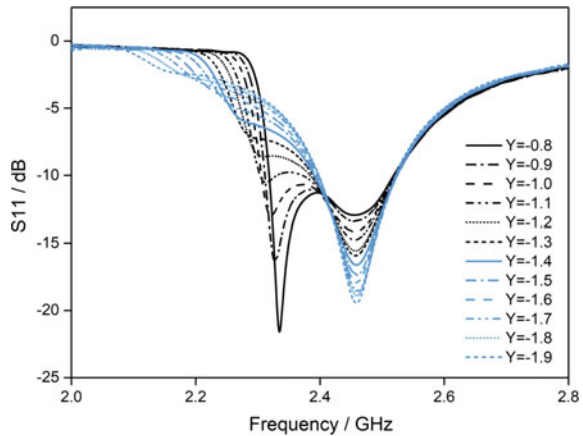
Fig. 9 Reflection coefficients for shorting post with various locations from -0.3 to -0.7



4 Conclusion

A novel, wideband patch antenna with DGS structure, multiple notches and shorting post for WBAN applications has been presented. The proposed antenna has a wide

Fig. 10 Reflection coefficients for shorting post with various locations from -0.8 to -1.9



bandwidth of 8.93%, covering from 2.32 GHz to 2.53 GHz which meets the required WBAN frequency bands. The novel shape and size are suitable for WBAN applications as it has the merits of stable gain and omnidirectional radiation pattern. The introduction of notches, DGS and shorting post improve the radiation efficiency as current path is increased around the slots. The proposed design has small size which makes it a suitable to be embedded for WBAN and healthcare monitoring system.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme under Proj-FRGS/1/2019/TK04/UCSI/02/1.

References

1. Geller T, David YB, Khmelnskiy E, Ben-Gal I, Ward A, Miller D, Bambos N (2019) Learning health state transition probabilities via wireless body area network. In: IEEE International Conference on Communication (ICC), pp 1–6
2. Abidi B, Jilbab A, Mohamed EH (2020) Wireless body area networks: A comprehensive survey. *J Med Eng Technol*:1–11
3. Jamil YK, Mehmet RY (2010) Wireless Body Area Network (WBAN) for Medical Application. *New Developments in Biomedical Engineering* 31:591–628
4. Conway GA, Scanlon WG (2009) Antennas for over-body-surface communication at 2.45 GHz. *IEEE Trans Antennas Propag* 57:844–855
5. 15.6–2012—IEEE Standard for Local and Metropolitan Area Networks—Part 15.6: Wireless Body Area Networks. (2012, February 29). Retrieved May 24, 2018, from <https://ieeexplore.ieee.org/document/6161600/versions/>

6. Wang D Chi HC (2016) Multiband antenna for WiFi and Wigig communication. In: IEEE Antennas and Wireless propagation Letters, vol 15
7. Das D, Maity S, Chatterjee B, Sen S (2020). Publisher correction: Enabling covert body area network using electro-quasistatic human body communication. *Scientific Reports* 10(1). <https://doi.org/10.1038/s41598-020-60493-6>
8. Srinivasan D, Gopalakrishnan M (2019) Breast cancer detection using adaptable textile antenna design. *J Med Syst* 43(6):1–10
9. Jebali N, Gharsallah A, Ribero JM (2020) Textile Ultra-Wide Band Antenna With X Band For Breast Cancer Detection. *Indian J Sci Technol* 13(11):1232–1242
10. Amir NF, Jusoh M, Isa MM, Sabapathy T, Al-Bawri SS, Rahim HA, Yasin MN (2019) A flexible UWB antenna for wearable technologies application. In: 6th International Conference on Space Science and Communication (IconSpace), IEEE, pp. 121–124
11. Mandal K, Sarkar PP (2013) A compact high gain microstrip antenna for wireless applications. *AEU-Int J Electron C* 67(12):1010–1014. <https://doi.org/10.1016/j.aeue.2013.06.001>
12. Karthikeyan M, Sitharthan R, Ali T, Roy B (2020) Compact multiband CPW fed monopole antenna with square ring and T-shaped strips. *Microw Opt Technol Lett* 62(2):926–932
13. Ojaroudi M, Civi ÖA (2016) High efficiency loop sleeve monopole antenna for array based UWB microwave imaging systems. In 2016 IEEE International Symposium on Antennas and Propagation (APSURSI). pp 1781–1782. IEEE
14. Sakib N, Ibrahim SN, Ibrahimy MI, Islam MS, Mahfuz MH (2020). Design of Microstrip patch antenna on rubber substrate with DGS for WBAN applications. In 2020 IEEE Region 10 Symposium (TENSYMP), pp 1050–1053. IEEE
15. Sriramulu D, Saini G, Dhaliwal BS (2020) Design of wearable microstrip patch antenna for WBAN applications to enhance bandwidth using multiple slots and CRR. *International Journal of Advanced Research in Engineering and Technology (IJARET)* 11(8):319–327
16. Yalduz H, Tabaru TE, Kilic VT, Turkmen M (2020) Design and analysis of low profile and low SAR full-textile UWB wearable antenna with metamaterial for WBAN applications. *AEU-Int J Elect Comm* 126:153465
17. Shizham M, Shubham J, Shreyas P, Sumeet B (2016) Design of flexible Microstrip patch antenna of 2.4 Ghz operation frequency using Hfss. *Int Res J Eng Tech (IRJET)* 3(5):638–644
18. Saeidi T, Ismail I, Mahmood SN, Alani S, Ali SM, Alhawari AR (2020). Metamaterial-based Antipodal Vivaldi Wearable UWB Antenna for IoT and 5G Applications. In 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech) pp 14–20. IEEE
19. Abbak M, Akıncı M, Çayören M, Akduman I (2017) Experimental microwave imaging with a novel corrugated Vivaldi antenna. *IEEE Trans Antennas Propag* 65(6):3302–3307
20. Sambandam P, Kanagasabai M, Natarajan R, Alsath MGN, Palaniswamy S (2020) Miniaturized button-like wban antenna for off-body communication. *IEEE Trans Antennas Propag* 68(7):5228–5235
21. Sharizal Fadlie Sabri (2016) Suriani Mohd Sam, Kamilia Kamardin, Salwani Mohd Daud & Nor Asnilawati Salleh. *Jurnal Teknologi R EVIEW OF THE CURRENT DESIGN ON WEARABLE.* 2:111–117
22. Khan MU, Sharawi MS, Mitra R (2015) Microstrip patch antenna miniaturisation techniques: A review. *IET Microwaves Antennas Propag* 9(9):913–922
23. Weng LH, Guo YC, Shi XW, Chen XQ (2008) An overview on defected ground structure. *Progress In Electromagnetics Research* 7:173–189
24. Dautov K, Hashmi M, Naurzybayev G, Nasimuddin N (2020) Recent advancements in defected ground structure-based near-field wireless power transfer systems. *IEEE Access* 8:81298–81309
25. Cao X, Luo B, Zhu Y, Xia Z, Cai Q (2020) Research on the defected ground structure with von Koch snowflake fractals. *IEEE Access* 8:32404–32411
26. Juanda EA, Hariyadi T, Mukhidin U, Yuniar RN (2020) Design and development of band pass filter microstrip cascade trisection with open stub and defected ground structure (DGS) in 1800 MHz frequency. *Journal of Engineering Science and Technology (JESTEC)* 15(1):372–381

27. Sohi AK, Kaur A (2020) UWB aperture coupled circular fractal MIMO antenna with a complementary rectangular spiral defected ground structure (DGS) for 4G/WLAN/radar/satellite/international space station (ISS) communication systems. *Journal of Electromagnetic Waves and Applications* 34(17):2317–2338
28. Kumar P, Singh G (2009) Microstrip antennas loaded with shorting post
29. Souza EA, Oliveira PS, D' Assunção AG, Mendonça LM, Peixeiro C (2019) Miniaturization of a microstrip patch antenna with a koch fractal contour using a social spider algorithm to optimize shorting post position and inset feeding. *International Journal of Antennas and Propagation*

PID Controller Based on Bird Mating Optimizer for Vibration Cancellation of Horizontal Flexible Plate



Muhammad Farith Nazreen Abu Samah, Muhamad Sukri Hadi,
Abdul Malek Abdul Wahab, Nadzirah Mohd Mokhtar,
and Intan Zaurah Mat Darus

Abstract The flexible system has massive benefits compared to the rigid structure, which is lightweight, low energy consumption and high efficiency. However, the lightweight structure causes undesired vibrations on a system that could damage the structure. This paper proposes a modelling of horizontal flexible plate structure by utilizing bird mating optimizer algorithm and implementation of vibration control on the system in simulation environment. There are two main objectives of this project which are to model a horizontal flexible plate system and to design an accurate controller by eliminating the excessive vibration of the system in order to achieve high-performance efficiency. Initially, input–output experiment data is collected from previous researchers and utilized to model system development. Then, the linear autoregressive with exogenous (ARX) model is selected as a model structure for model development using system identification method via bird mating optimizer (BMO) algorithm. Thereafter, the developed models were evaluated by mean square error (MSE), correlation tests and pole-zero map. The finest model is identified based on the minimum MSE, stable in a pole-zero diagram and unbiased in correlation tests. Then, the mathematical transfer function of the best model is utilized for the development of the PID controller. The controller parameter is tuned using trial and error method. Finally, the developed controller was evaluated and accessed using two different disturbances known as single and multiple sinusoidal disturbances. The result shows a significant vibration reduction of 11.18 and 10.13% for single and multiple sinusoidal disturbances, respectively.

Keywords Flexible plate structure · System identification · Swarm intelligence algorithm · Bird mating optimizer · Active vibration control

M. F. N. A. Samah · M. S. Hadi (✉) · A. M. A. Wahab
School of Mechanical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia
e-mail: msukrihadi@uitm.edu.my

N. M. Mokhtar
Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Kuantan, Malaysia

I. Z. M. Darus
School of Mechanical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

1 Introduction

In the past decade, most industries implemented a rigid structure in engineering applications consistently. One of the advantages of the mention structures is less unwanted vibration in the system due to their heavy and bulky characteristics. However, the system characteristics effect limits the operation speed and high energy consumption [1]. Hence, the industry in various engineering applications such as civil, mechanical, and aeronautical, currently replaced it with flexible structure. The latter consists of four elements including shells, beams, plates, and manipulators. Opposite to the rigid structure, the flexible one provides high efficiency, lightweight and low energy performance.

The flexible plate is the most popular element used in industry, such as satellites, solar panels, and conveyors. The benefits of the structure attract the researcher to conduct an experimental study for maximum utilization of the system. However, the lightweight property is usually sensitive to external disturbances that can cause some negative impacts on the plant like noise, fatigue and wear [2, 3]. This impact will consequently reduce the system effectiveness and putting people life at risk. Thus, the undesirable vibration should be controlled or eliminated from the flexible plate structure by implementing robust vibration control system to preserve the effectiveness of the system [4].

In the research area, there are two types of vibration control commonly reported for minimization of excessive vibration on the plate structures; passive and active vibration controls. The passive material in the PVC system plays a role control the characteristic of dynamic system such as stiffness and damping ratio by increasing the structure's weight to eliminate excessive vibration in the structure. However, this method is not suitable for low frequencies, which is expensive and leads to bulky due to larger size and weight, low of versatility and de-tuning treatment process [5].

The next method of vibration control is the active vibration control (AVC) which is the most fitting method for the flexible system. This method minimizes the vibration of the structure by producing secondary sources. This source will interrupt that unwanted disturbance that causes the deduction of the vibration. The efficiency of this method depends on the location of electronic components. The number of secondary sources is subjective to the physical dimension of structure in achieving the accurate output result [6]. In addition, this method usually required a controller to suppress the vibration. The most popular controller is the PID controller because of the numerous advantages such as easy implementation, less complex and minimal computational effort [7]. Moreover, there are only three tuning parameters in the PID controller known as proportional, integral and derivatives gains. However, in order to attain a robust PID controller, an accurate model of plant need to be achieved first [8, 9].

Previously, many researchers using varieties of the method to model the flexible plate system such as finite element (FE) and finite difference (FD) methods [5, 10]. An exact model system is crucial to ensure the controller produces robust performance in the vibration control system. Hence, many researchers currently applying the system identification (SI) method for modelling the system to represent the actual

system at any application. This method produces a mathematical transfer function that represented the actual system by utilization of the vibration input–output data collected from the experiment. Nowadays, researchers implement system identification via intelligence algorithms including firefly, flower pollination, and ant colony optimization to model flexible plate systems [11–13]. However, the implementation of the BMO algorithm in modelling flexible plate structures for vibration control purposes is still limited. The superiority of the BMO algorithm for optimization has been explained by Arram and his colleagues in [14]. Hence, this study proposes the model development of horizontal flexible plate structure using swarm intelligence algorithm via bird mating optimizer algorithm owing to its superiority. Then, the controller was developed based on best model to suppress the undesired vibration. The performance of the controller was validated under single and multiple sinusoidal disturbances.

2 Methodology for Vibration Data Collection

The combination of software and hardware utilizing a data acquisition (DAQ) system is utilized to collect the structural parameter of the plate structure. The DAQ system consists of an actuator, sensor and a computer running DAQ software and signal conditioning hardware. In this research study, the system identification method based on input–output experimental data has been applied to attain appropriate model of the flexible plate structure. However, the vibration data was collected from previous experiment conducted by former researcher. The experimental setup consists of components and instruments namely, flexible plate, power amplifier, function generator, piezo-beam type accelerometer, magnetic shaker, circular shape permanent magnet and DAQ system [15].

This experiment was carried out by referring to the experimental setup conducted by Hadi and his friends in 2018 [15]. The flexible thin, square and flat aluminium plate was clamped at all edges horizontally to ensure the flexible plate vibrate in a vertical condition. The dimension of the flexible plate used in this experiment is 0.07 m on each side and 0.001 m in thickness. The circular shape permanent magnet was attached at 0.01 m parallel to the magnetic shaker at the excitation point of the experimental rig to generate the excitation of the system by the existence of sinusoidal force. The magnetic shaker was connected to the function generator and power amplifier. Furthermore, double pieces of piezo-beam type accelerometer that act as sensors were mounted at different locations which are detection and observation point to collect the acceleration signal that represents the vibration of system. A computer consist of DAQ system was connected to the piezo-beam type accelerometer to analyse the obtained experimental data [15] (Fig. 1).

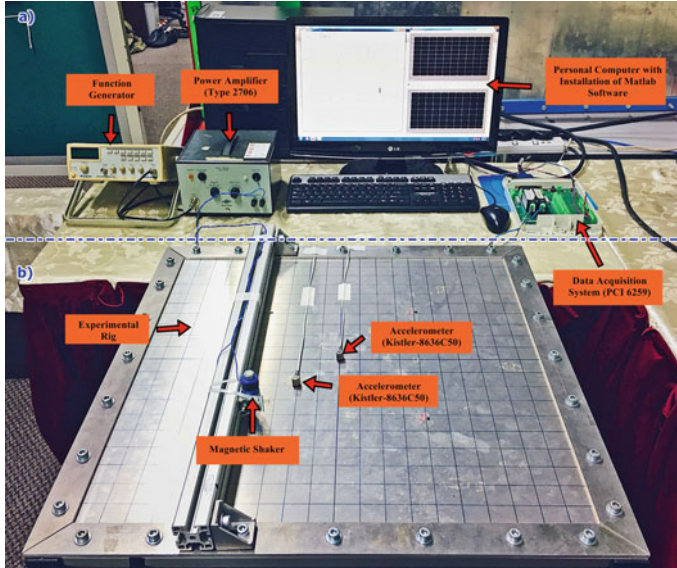


Fig. 1 Experimental layout for input–output data collection, **a** The integration process between the experimental rig and DAQ system, **b** The bottom view of the plate

3 Bird Mating Optimizer Algorithm

Bird mating optimizer (BMO) is a swarm intelligence algorithm inspired by bird mating strategies. The BMO have a population known as society and each society represents a bird which is the feasible solution of certain problems. The society is divided into two groups which are male and female birds. The male birds consist of monogamous, polygynous and promiscuous while female birds have two groups which are parthenogenetic and polyandrous. There are five types of bird mating strategies considered in BMO which are monogamy, polygyny, promiscuity, parthenogenesis and polyandry.

Monogamy is a bird mating strategy that involved a male bird, \vec{x}_M that chooses only a female bird, \vec{x}_i to mate with each other. In monogamy system, most of the bird is monogamous, and the male bird attracts the female bird by an intelligent behaviour. The resultant brood of this system is described in (1) [16]:

$$\begin{aligned}
 \vec{x}_b &= \vec{x}_M + w \times \vec{r} \cdot (\vec{x}_i - \vec{x}_M) \\
 c &= \text{random number between 1 and } n \\
 \text{if } r_1 &> mcf \\
 x_b(c) &= l(c) - r_2 \times (l(c) - u(c)) \\
 \text{end}
 \end{aligned} \tag{1}$$

where \vec{x}_b is the resultant brood of monogamy system, w is a time-varying weight to alter the attractive female bird, \vec{r} is a vector of each element that distributed randomly in $[0, 1]$, mcf is the mutation control varying, which is in range 0 to 1, u and l represented the upper and lower boundary of the element.

Polygyny is a bird mating system in which the male bird chooses to mate with a number of female birds. This male bird involved in this mating system is polygynous. The mating of polygynous and female birds will produce only a brood which is the genes is a combination of female birds. The resultant brood of this mating system is represented in (2) [16]:

$$\begin{aligned} \vec{x}_b &= \vec{x}_{Pg} + w \times \sum_{j=1}^{n_i} \vec{r} \cdot (\vec{x}_{ij} - \vec{x}_{Pg}) \\ c &= \text{random number between 1 and } n \\ \text{if } r_1 &> mcf \\ x_b(c) &= l(c) - r_2 \times (l(c) - u(c)) \\ \text{end} \end{aligned} \tag{2}$$

where n_i is the number of interesting birds, \vec{x}_{Pg} is the polygynous bird and \vec{x}_{ij} indicate the j th attractive female birds.

Next, the promiscuity mating system showed the male and female birds only mate at once and male birds have a high probability to not see the female and his brood for another visit. Polyandry is a mating system in which the female bird tends to couple with a number of male birds. The resultant brood formulation of promiscuity and polyandry is the same as the monogamy system as shown in Eq. (2).

Parthenogenesis is a mating system that the female bird cares for her breed on their own without involving any male bird. The resultant brood of parthenogenesis is outlined as follow [16]:

$$\begin{aligned} \text{for } i &= 1 : n \\ \text{if } r_1 &> mcf_p \\ x_b(i) &= x(i) + \mu \times (r_2 - r_3) \times x(i) \\ \text{else} \\ x_b(i) &= x(i) \\ \text{end} \end{aligned} \tag{3}$$

where, mcf_p is the mutation control of parthenogenesis and μ is the step size.

4 Result and Discussion

4.1 *The Modelling System Based on Bird Mating Optimizer (BMO) Algorithm*

The modelling of the horizontal flexible plate structure based on auto-regressive with exogenous (ARX) model structure is designed via bird mating optimizer (BMO) algorithm to identify the best model system that represent the actual behaviour by a mathematical transfer function. The system model is developed using 5000 vibration input–output data collected from an experiment where the first 2500 data are used to train the model, and the next 2500 data are used to verify the effectiveness of the model. The developed model is considered the best model system after achieving the robustness criteria by obtaining the lowest MSE, stable in pole-zero stability diagram and correlates within 95% confidence level of correlation test.

Since there is no specific approach to tuning the parameter, the heuristic tuning method is applied to obtain the best system model. The tuning of parameters process is considered important to determine the performance of the model. There are 7 parameters that needs to tune properly for this algorithm which are model order, number of maximum iterations, mutation control factor (mcf), lower and upper boundary, mutation size, μ , percentage of each species from the society and weight of time-varying, w .

However, 2 parameters are fixed which percentage of each species from society and time-varying weight to adjust the importance of the interesting female, w . This is because Alireza and his friend (2013) recommend the w to set at 0.001 to improve the global and local search. They also suggested setting the percentage of monogamous, polygynous, promiscuous, polyandrous, and parthenogenetic birds at 50, 30, 10, 5 and 5 of the society, respectively, to obtain the optimum solution for each species [16].

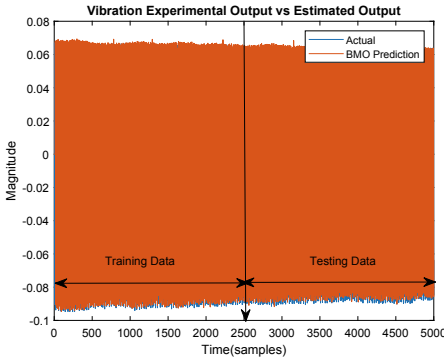
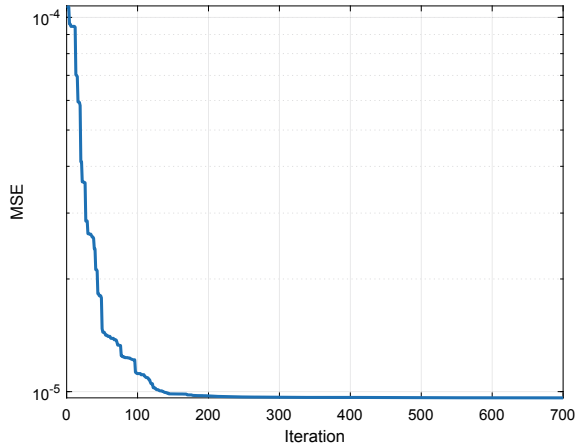
Based on the tuning result, the finest model system is achieved after the developed model is obtained 3.5806×10^{-6} for testing data, which is the lowest MSE, stable in the pole-zero diagram and correlated within 95% confidence level in cross-correlation test. The set of best parameters are utilized in the BMO algorithm during modelling the flexible plate system is indicated in Table 1. The convergence of the BMO algorithm is plotted in the graph, as shown in Fig. 2. Next, the actual and estimated vibration output in time and frequency domain are illustrated in Figs. 3 and 4, respectively, while the error between experiment and estimated output by BMO algorithm is shown in Fig. 5.

Next, the developed model must be validated by a robustness test. If the developed model satisfied the test, the developed model is considered the best model of the system. The performance of the BMO algorithm in pole-zero stability and correlation test is shown in Figs. 6 and 7, respectively. Refer to Fig. 6, and it indicates that all the poles are inside the locus, which considered the system is stable, while Fig. 7 shows the developed model is unbiased in the cross-correlation test. However, this set of parameters does not achieve the 95% confidence level of the correlation test

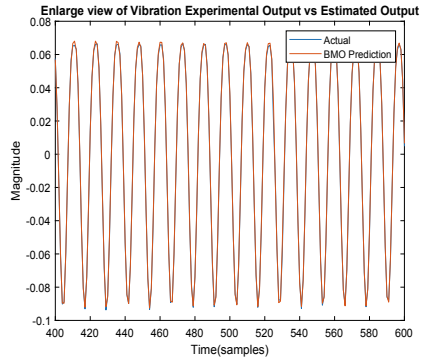
Table 1 Best parameters that utilized to obtain best model based on BMO algorithm

Parameters	Symbols	Value
Number of birds	N_{pop}	200
Time-varying weight	w	0.001
Model order	NA	6
Mutation size	μ	0.006
Mutation control factor	mcf	0.9
Number of maximum iterations	Max_{iter}	700
Lower and upper boundary	$[LB, UB]$	$[-0.4, 0.4]$

Fig. 2 Convergence of BMO algorithm



(i)



(ii)

Fig. 3 Actual and BMO algorithm prediction result in time domain, **i** Vibration output for 5000 data; **ii** Enlarge view of 400 to 600 vibration output

Fig. 4 Actual and BMO algorithm prediction of flexible plate system in the frequency domain

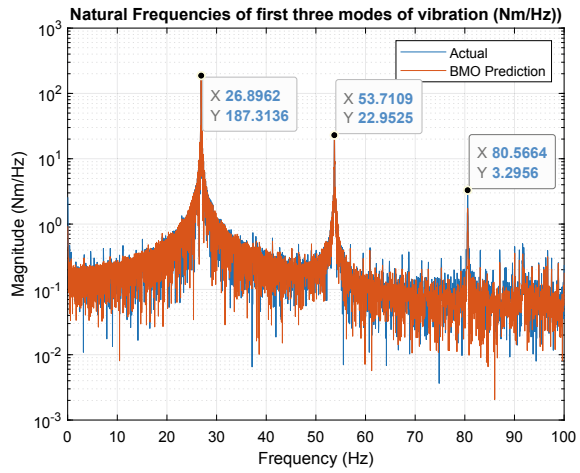
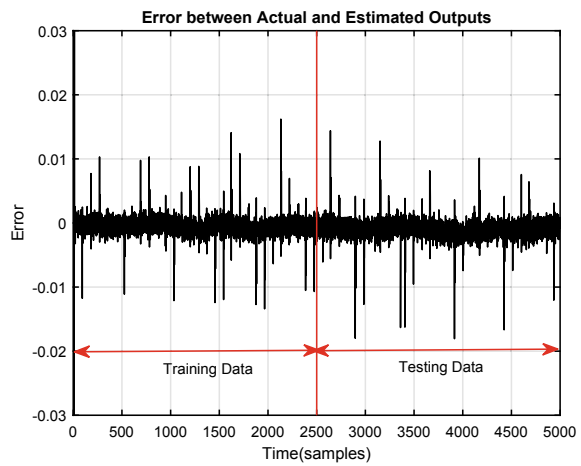


Fig. 5 Error between experiment and estimated output by BMO algorithm



known as biased. In addition, Eq. (4) represents the discrete-time transfer function of the developed model.

$$\frac{y(t)}{u(t)} = \frac{-0.0302z^{-1} + 0.1215z^{-2} + 0.3316z^{-3} + 0.0316z^{-4} - 0.2448z^{-5} - 0.2776z^{-6}}{1 - 0.3122z^{-1} + 0.154z^{-2} + 0.1014z^{-3} + 0.0441z^{-4} + 0.233z^{-5} + 0.1141z^{-6}} \quad (4)$$

Fig. 6 Stability of pole-zero diagram of developed model based on BMO algorithm

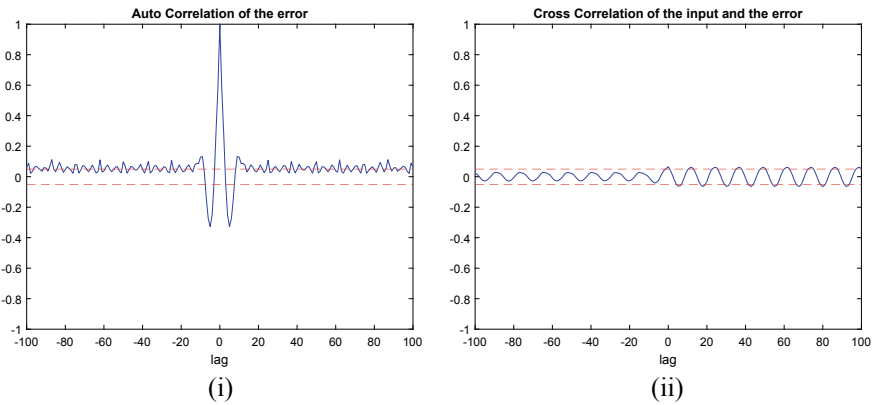
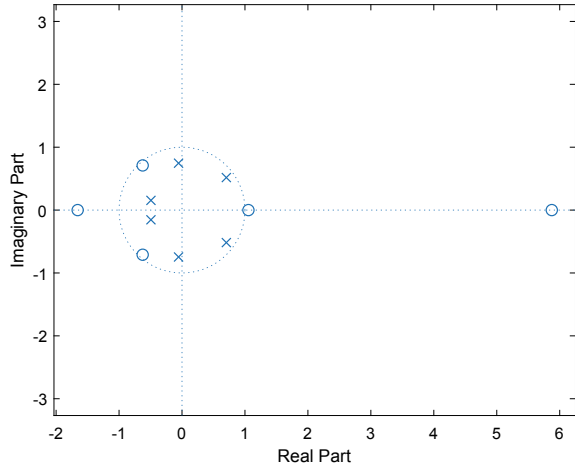


Fig. 7 Correlation test of developed model based on BMO algorithm, **i** Auto correlation test **ii** Cross correlation test

4.2 PID Controller Development in Active Vibration Control System

Active vibration control (AVC) system is applied to suppress the undesired vibration on a flexible plate system. Hence, the parameters of the PID controller are tuned using the heuristic method. Table 2 is illustrating the set of best parameters of the

Table 2 Set of best parameters of PID controller that used in AVC system

Parameter	K_p	K_i	K_d
Value	1.0	0.2	0.007

PID controller obtained from the tuning method.

4.2.1 The Performance of PID Controller Under Single Sinusoidal Disturbance

The AVC system of the flexible plate system is applied under a single sinusoidal disturbance to verify the controller’s performance in suppressing the undesired vibration on the system. The controller is developed by utilization the best parameter of the controller The results of the PID controller are plotted in a graph based on time and frequency domain, as shown in Figs. 8. Next, Table 3 describes the PID controller is attenuated the decibel magnitude by 6.50 dB, which is 11.18% of the reduction percentage.

4.2.2 The Performance of PID Controller Under Multiple Sinusoidal Disturbance

On the other hand, the PID controller is developed under multiple sinusoidal disturbances. Figure 9 illustrates the performance of the PID controller based on the time

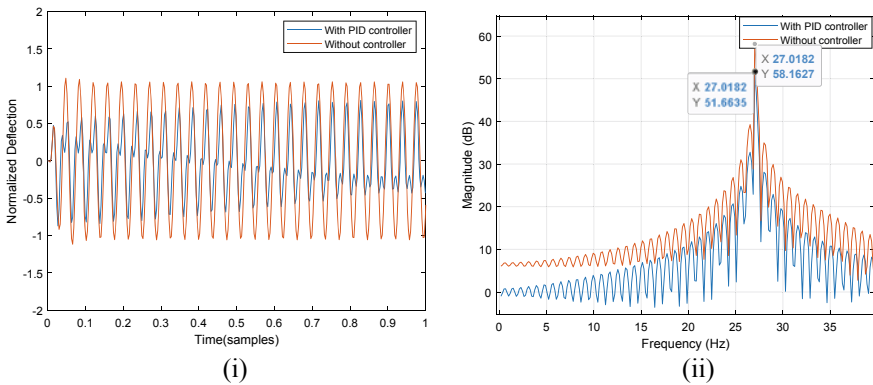


Fig. 8 Result of PID controller under single sinusoidal disturbances, **i** Time domain, **ii** Frequency domain

Table 3 Attenuation level and percentage of reduction under single sinusoidal disturbance

Controller	Decibel magnitude (dB)	Attenuation level (dB)	Percentage of reduction (%)
	1 st Mode	1 st Mode	1 st Mode
Without controller	58.16	References	References
PID controller	51.66	6.50	11.18%

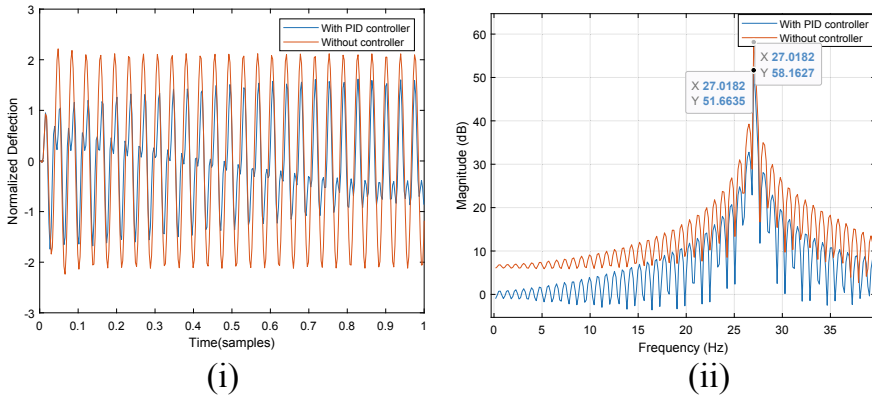


Fig. 9 Result of PID controller under multiple sinusoidal disturbances, **i** Time domain, **ii** Frequency domain

Table 4 Attenuation level and percentage of reduction under multiple sinusoidal disturbance

Controller	Decibel magnitude (dB)	Attenuation level (dB)	Percentage of reduction (%)
	1 st Mode	1 st Mode	1 st Mode
Without controller	64.18	References	References
PID controller	57.68	6.50	10.13%

and frequency domains. Next, Table 4 illustrates the PID controller is attenuated the decibel magnitude by 6.26 dB, which is 10.76% of the reduction percentage. This result proved that the developed controller successfully eliminates the undesired vibration on the flexible plate system.

5 Conclusion

In this paper, an appropriate model of horizontal flexible plate system was developed via the BMO algorithm. The validations of the developed model including mean squared error, pole-zero map and correlation tests were described in this research. The result outlines the best model that represents the dynamic system of the flexible plate structure met all the validation criteria since the developed model achieved the lowest mean squared error of 3.5806×10^{-6} , high stability of pole-zero diagrams and good correlation tests. After that, the best model achieved in the previous stage was employed in the PID controller development to suppress the undesired vibration. The controller design was validated using single, and multiple sinusoidal disturbances.

Based on the obtained result, the developed controller successfully eliminates the undesired vibration up to 11.18 and 10.13% for both single and multiple sinusoidal disturbances, respectively. Hence, the objectives of this research have been achieved successfully.

Acknowledgements The authors would like to express their gratitude to Universiti Teknologi MARA (UiTM), Universiti Malaysia Pahang (UMP), Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education (MoHE) for funding the research and providing facilities to conduct this research. Sponsor file number (RACER/1/2019/TK03/UITM//1).

References

1. Jamali A, Mat Darus IZ, Tokhi MO (2020) Intelligent PID controller of flexible link manipulator with payload. In: 2020 3rd international conference on mechanical, electronics, computer, and industrial technology (MECnIT), pp 118–122
2. Darus IZM, Al-Khafaji AAM (2012) Non-parametric modelling of a rectangular flexible plate structure. *Eng Appl Artif Intell* 25(1):94–106
3. Noack M, Kuhhorn A, Kober M (2021) A new stress-based topology optimization approach for finding flexible structures. *Struc Multidisc Optim* 1–11
4. Lin CY, Chiu WH, Lin JL (2015) Rejecting multiple-period disturbances: active vibration control of a two degree-of-freedom piezoelectric flexible structure system. *J Vib Control* 21(16):3368–3382
5. Tavakolpour AR, Darus IZM, Mailah M (2009) Modeling and simulation of an active vibration control system for a flexible structure using finite difference method. In: Proceedings - 2009 3rd Asia international conference on modelling and simulation, AMS 2009, pp 448–453
6. Rahman TAZ, Darus IZM (2011) Experimental evaluation of active vibration gain controller. In: 2011 IEEE symposium on industrial electronics and applications, pp 382–386
7. Sun JH, Cheng LL (2017) Robust PID controller for AR drone. In: Computer science and technology (CST2016), pp 1213–1221
8. Darus IZM, Tokhi MO (2005) Soft computing-based active vibration control of a flexible structure. *Eng Appl Artif Intell* 18(1):93–114
9. Julai S, Tokhi MO, Salleh SM (2009) Active vibration control of a flexible plate structure using ant system algorithm. In: EMS 2009 - UKSim 3rd European modelling symposium on computer modelling and simulation, no 1, pp 37–42
10. Yuldashev A, Pirmatov ST (2020) Algorithmization of solving dynamic edge problems of the theory of flexible rectangular plates. *Univ J Math Mech* 66:143–157
11. Wang TX, Bo YM, Wang XM, Zhao GP (2013) Optimal actuator and sensor locations for active vibration control: using improved particle swarm algorithm. In: Proceedings - 2013 4th world congress on software engineering, WCSE 2013, vol 1, pp 288–292
12. Maseri SZ, Hadi MS, Jamali A, Yatim HM, Ab Talib MH, Darus IZM (2019) A Single objective flower pollination algorithm for modeling the horizontal flexible plate system. In: Proceedings of the 2019 2nd international conference on applied engineering, ICAE 2019, pp 0–5
13. Jalil NSAA, Julai S, Ramli R (2015) Parametric modelling of flexible plate structures using continuous ant colony optimization. *J Simul* 9(3):223–231
14. Arram A, Ayob M, Kendall G, Sulaiman A (2020) Bird mating optimizer for combinatorial optimization problems. *IEEE Access* 8:96845–96858
15. Hadi MS, Yatim HM, Darus IZM (2018) Modelling and control of horizontal flexible plate using particle swarm optimization. *Int J Eng Technol (UAE)* 7(2):13–19
16. Askarzadeh A (2014) Bird mating optimizer: an optimization algorithm inspired by bird mating strategies. *Commun Nonlinear Sci Numer Simul* 19(4):1213–1228

Hub Angle Control of Flexible Manipulator Based on Bacterial Foraging Optimization



Muhammad Nazri Mohd Salme, Muhamad Sukri Hadi, Annisa Jamali, Hanim Mohd Yatim, Mat Hussin Ab. Talib, and Intan Zaurah Mat Darus

Abstract Flexible manipulator offers industry with less material requirement, lighter in weight thus transportable, consuming less power, require smaller actuators, less control complexity while being able to operate in higher payload to weight. But, due to high flexibility of the flexible manipulator, excessive vibration can be found if the system is implemented. This study aims to simulate an accurate model system using system identification (SI) technique via Bacterial Foraging Optimization (BFO) for control of the hub angle of the flexible manipulator system in simulation environment. It is vital to model the system that represents actual characteristics of the flexible manipulator before precisely control the hub angle of the flexible manipulator's movement. The experimental data obtained from the flexible manipulator system's hub are utilised to construct a model of the system using an auto-regressive with exogenous (ARX) structure. Bacterial Foraging Optimization (BFO) is used to develop the modelling by SI technique to obtain the mathematical models. The generated model's performance is assessed using three methods: minimum mean square error (MSE), correlation tests, and stability test in pole-zero diagram. The model of hub angle constructed using BFO has a minimum mean square error of $1.9694 \cdot 10^{-5}$, a high degree of stability, and strong correlation tests. The model of hub angle constructed using BFO has a minimum mean square error of $1.9694 \cdot 10^{-5}$, a high degree of stability, and strong correlation results. Following that, a PID controller is designed and heuristically tuned to provide accurate hub angle positioning with a short settling time using the BFO model. It is also worth noting that BFO's model successfully regulated the hub angle's positioning with a 0.8% overshoot and a 0.5242 s settling time in the presence of single disturbances.

M. N. Mohd Salme · M. S. Hadi (✉)

School of Mechanical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia
e-mail: msukrihadi@uitm.edu.my

A. Jamali

Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

H. M. Yatim · M. H. Ab. Talib · I. Z. M. Darus

School of Mechanical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

Keywords Flexible manipulator · System identification · Bacterial Foraging Optimization · Swarm intelligence algorithm

1 Introduction

Nowadays, the flexible manipulator system has been offered as a replacement for the rigid manipulator system. A rigid manipulator system is one that has a high vibrational frequency. The system exhibits negligible tip deflection and is capable of maneuver at any desired speed even when the system's single-link mass is large and consumes a large amount of energy [1]. Since the industrial revolution's inception, a manipulator system has improvised the manufacturing and spacecraft industries. In 1984, an ABB Robots rigid manipulator system weighed around 1450 kg and could operate with a weight of only 90 kg. A manipulator system is typically powered by hydraulic actuators and welded with hefty steel plates, which reduces the device's energy efficiency [2]. An industrial robot's performance requirements and motion control should be limitless [3]. However, if a system with a high mass and stiffness that consumes a significant amount of energy to work on a payload can be replaced by a system with a low mass and stiffness that consumes a similar amount of energy to work on a payload, isn't that a good revolution?

In addition to industrial robots, there is a high demand for manipulator systems within spaceships. The inflexible manipulator will spend inefficient amounts of energy and presenting the system in the modern period will be devastating. Thus, a flexible manipulator system is the most appropriate system, as demonstrated by existing space manipulator systems. The system is adaptive as a result of the connections and joints inherent within [4]. The structure is far less stiff, requires less energy, and delivers payloads swiftly and precisely [5]. Regardless of this, flexible manipulator systems are not as prevalent in manufacturing as they are in spaceship manufacturing. The issue exists because the vibration control mechanism for the flexible manipulator system is still in its early stages. Manufacturers are interested in acquiring the system because of the potential benefits that it may give that are much superior than those supplied by competing systems at a lower cost. The flexible manipulator system eliminates the need for a large number of actuators and sensors, which reduces the cost and nonlinearities associated with rigid manipulator systems [6].

Thus, how might a versatile manipulator system contribute more to the industry? Initially, passive vibration control was used to alter the damping coefficient and stiffness of the manipulator system [7]. This control mechanism reduced the system's ability to progress because there were fewer structures to enhance due to the lack of feedback information from the system. Numerous researchers have developed active vibration control for flexible manipulator systems during the last three decades by including sensors such as accelerometers and strain gauges [8]. It is widely agreed that active vibration control with a closed loop system is more suitable for larger scale industrial applications due to the feedback loop's ability to alter the system's structure

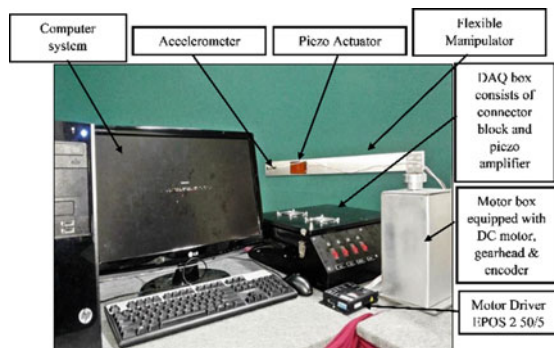
and vibration control more precisely [9]. In a complete scheme of active vibration control, actuators and sensors for vibration detection were attached on the flexible manipulator system. As a result, a proportional-integral-derivative (PID) controller was created to suppress the detected vibration. To complete the set, the controller is also intended to specify a bandwidth, or a high fundamental frequency of vibration, which enables the system to operate at high speeds while avoiding considerable tip deflection [10].

2 Experimental Setup of the Flexible Manipulator System

The experimental setup of the entire system was designed to observe the hub angle and endpoint acceleration and collects its vibration responses of input–output data as referred in Fig. 1. As a mechatronic system, the flexible manipulator is basically a mechanical system be equipped with an instrumentation system and software and computer control [11]. The mechanical parts comprise of a single link thin aluminium alloy pinned one end to a motor and free at the other. The dimensions of the single link aluminium alloy are $600 \times 40 \times 1.5$ mm. The aluminium has 1.125×10^{-11} m² moment of inertia, 71 GPa Young’s modulus and 2710 kg/m² mass density per area.

The instrumentation system is made of actuators, sensors and a software programmable computer that collects the data of structural behaviour by using a data acquisition system (DAQ) that can detect and configure the input output data from sensors and actuators. A single ICP accelerometer is to be used as a sensor to gain vibrational feedback attached at the free end of the flexible manipulator as maximum vibration point. The electrical signal from the accelerometer will be computed by the data acquisition system with the aid of computer. As a patch transducer, the piezo actuator is equipped as it is able to generate mechanical deformation as response when voltage applied and vice versa and also able to generate voltage when applied to stress. These capabilities enhanced the actuator’s principal function in this investigation, which was to generate the control signal. To complete the set, a piezo actuator amplifier is included to boost the voltage signal from the DAQ system

Fig. 1 Experimental setup for the system integration



to the appropriate level for the piezo actuator to function optimally. The flexible manipulator's one end is attached to the actuator system's central hub. At the hub, a 150-W DC motor acts as an actuator for the flexible manipulator system. The configuration enables the hub's speed and angle displacement to be controlled via a PC [10].

3 Bacterial Foraging Optimization

BFO is a mimicry of the foraging procedures used by bacteria, more precisely *Escherichia Coli* (*E. coli*), to locate nutritious areas. The bacteria search for the optimal through a sequence of movements on functional surfaces, including chemotaxis, swarming, reproduction, and finally elimination-dispersal [12]. The modelling of the flexible manipulator system is created using System Identification (SI) technique. By using ARX structure, the model will represent the real characteristics of the experimental flexible manipulator system in mathematical expression. The model will be verified by three robustness tests that are the least mean-squared error in testing data (MSET), stable in the pole-zero diagram and portrays within 95% of correlation tests. The model will be applied in the active vibration control scheme that will be established. Each dynamic response of the flexible manipulator is modelled, including input–output data from the hub angle.

The parameters that are tuned are the step size taken by the bacteria, the chemotactic steps count, the elimination-dispersal steps count and lastly the model order. Chemotaxis is the main part of the BFO algorithm. At chemotactic steps, bacteria move to reach the global optimum. This stage also generates the bacteria's motility. Therefore, it is the most crucial step to determine the convergence of the algorithm to the system. The value of chemotactic steps from 10 to 50 are proven to excel reaching global optimum with acceptable computing time [13]. As crucial as the chemotactic steps, the elimination and dispersal steps direct the algorithm to reach the global optimum solution as the bacteria might skip it in the chemotactic loop [14]. The best model for hub angle is achieved after heuristic approach is conducted on parameters involved in the algorithm.

The implementation steps of BFO in modelling the flexible manipulator system are as follows:

- 1) BFO computation requires the specification of eight parameters, including the dimension of the search space, p and the number of bacteria in the population, S . N_c denotes the number of chemotactic steps per bacteria lifetime, N_s denotes the swimming length of the bacteria, N_{re} denotes the number of reproduction steps, N_{ed} denotes the number of elimination-dispersal steps, P_{ed} denotes the probability of elimination-dispersal, and finally $C(i)$ denotes the step size of the chemotactic steps.
- 2) Eliminate-dispersal step, N_{ed} ($l = 1, 2, \dots$)
Do $l = l + 1$

- 3) Reproduction step, N_{re} ($k = 1, 2, \dots$)
Do $k = k + 1$
- 4) Chemotaxis step, N_c ($j = 1, 2, \dots$)
Do $j = j + 1$. For ($i = 1, 2, \dots, S$),
- i. Conduct a chemotactic step on each bacterium separately.
 - ii. Calculate the nutritional content (J, i, j, k, l) with J as cost function. In this study, MSE is implemented to set up the model prediction error. The cost function J is denoted as:

$$J = \frac{1}{n} \sum_{k=1}^n (e(k))^2 \quad (1)$$

Where n denotes the number of data points and e(k) denotes the prediction error of the model.

Calculate,

$$J(i, j, k, l) = J(i, j, k, l) + J_{cc}(\theta^i(j, k, l), P(j, k, l)) \quad (2)$$

If no swarming effect,

$$J_{cc}(\theta^i(j, k, l), P(j, k, l)) = 0 \quad (3)$$

- iii. Declare $J_{last} = J(i, j, k, l)$ first, it is possible that the further run has better cost.
- iv. Tumble. Produce a random vector, $\Delta(i)$ ranged $[-1, 1]$
- v. Compute,

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C_a(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \quad (4)$$

- vi. Compute nutrient value for every bacterium (J, i, j, k, l) with J as cost function. Calculate,

$$J(i, j+1, k, l) = J(i, j+1, k, l) + J_{cc}(\theta^i(j+1, k, l), P(j+1, k, l)) \quad (5)$$

If no swarming effect, use (3)

- vii. Swim. Set $m = 0$, while $m < N_s$ (if not climbing down too long)

- Count $m = m + 1$
- If $J(i, j+1, k, l) < J_{last}$, means better. Update $J_{last} = J(i, j+1, k, l)$
- Also, let

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C_a(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \quad (6)$$

- As in step v, utilize $\theta^i(j+1, k, l)$ to determine the new $J(i, j+1, k, l)$
 - Else, if $m = N_s$. End the while statement.
- viii. Repeat step ii. for the next bacterium if $i \neq S$
- 5) If $j < N_c$, repeat step 4.
- 6) Reproduction step
- i. Allow

$$J_{health}^i = \sum_{j=1}^{N_c+1} J(i, j, k, l)$$

- be the state of health of bacterium I for the specified values of k and l , and for each $i = 1, 2, \dots, S$. Rank bacteria and chemotactic parameters, $C(i)$ from lowest to highest value of J_{health} (Highest values mean lower health).
- ii. S_r bacteria with highest values of J_{health} dies. The other bacteria reproduced by splitting identically to their parent at the same location.
- 7) If $j < N_{re}$, repeat step 3.
- 8) Elimination-dispersal step
- i. For $i = 1, 2, \dots, S$, and probability, p_{ed} .
- ii. For $m = 1:S$,
- If $p_{ed} < \text{rand}$
 - rand (Generate random number, if the number is less than p_{ed} , bacteria is eliminated and dispersed, generate new random positions for bacteria.
 - Else, bacteria did not eliminate and dispersed hence staying at the current position.
- 9) If $l < N_{ed}$, repeat step 2). Else, end.

4 Result and Discussion

4.1 The Modelling of Hub Angle via Bacterial Foraging Optimization (BFO)

The convergence of BFO modelling of the hub angle is illustrated in Fig. 2, and the model attained an MSE of $1.9694, 10^{-5}$. The output response in time domain for the hub angle predicts the true properties of the hub angle, as illustrated in Fig. 3. The normalised error associated with the BFO forecast of the hub angle is shown in Fig. 4. As demonstrated by the dotted line in Fig. 3 and Fig. 4, 1750 sets of actual input–output data are utilised for training and the remaining 4650 sets of data are used for testing. The model's effectiveness is then tested using robustness tests such as pole-zero stability and correlation tests. The model has high stability in Fig. 6 and

Fig. 2 The convergence of modelling of the hub angle based on BFO

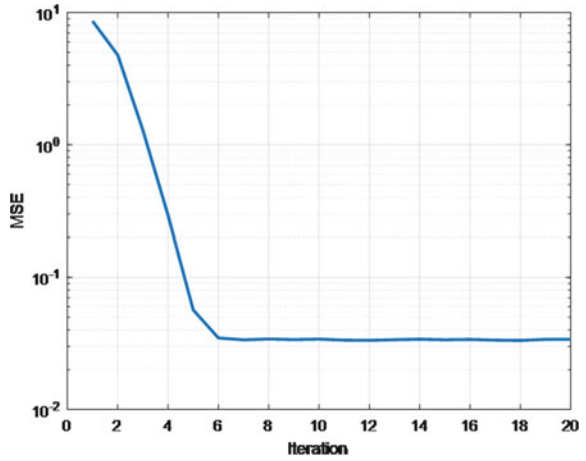
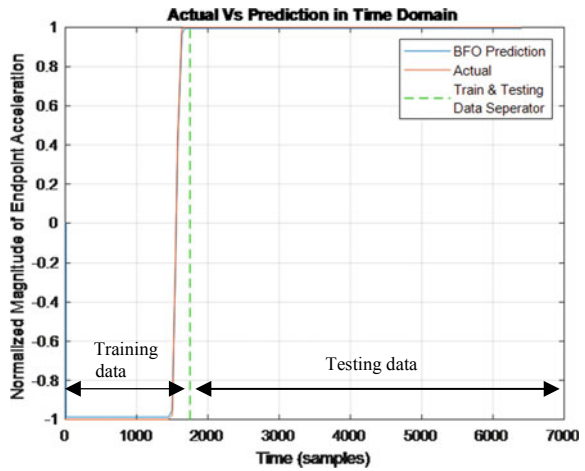


Fig. 3 Actual and BFO model of hub angle output responses in time domain



is also unbiased within 95% of cross-correlation and auto-correlation tests in Fig. 5. The discrete transfer function of the model is as shown in Eq. (7). The parameters used to achieve the best parameter in modeling of hub angle using BFO algorithm as presented in Table 1.

$$\frac{y(t)}{u(t)} = \frac{0.05716z^{-1} - 0.06697z^{-2} - 0.01907z^{-3} + 0.04745z^{-4}}{1 - 0.4855z^{-1} - 0.4269z^{-2} - 0.08995z^{-3} + 0.01039z^{-4}} \quad (7)$$

Fig. 4 The modelling normalized error of hub angle based on BFO

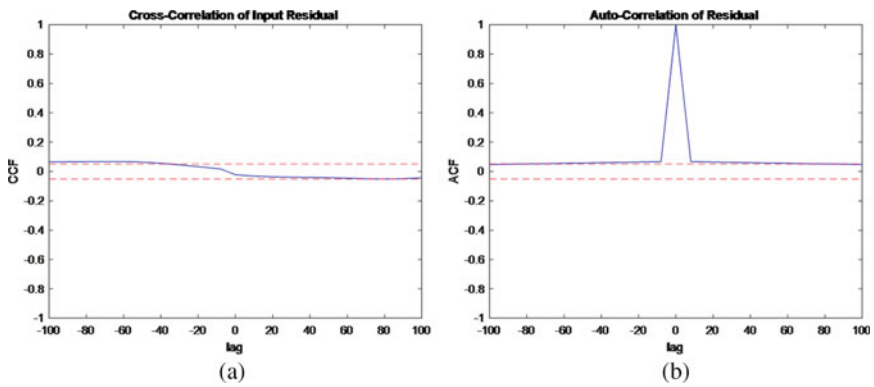
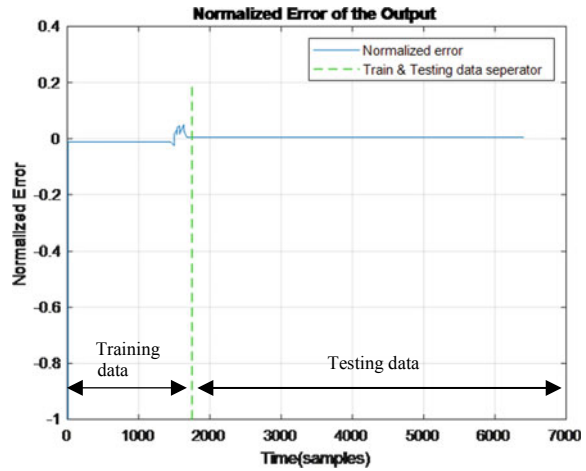


Fig. 5 Correlation results of the hub angle model based on BFO

Fig. 6 Pole-zero stability diagram for the hub angle model based on BFO

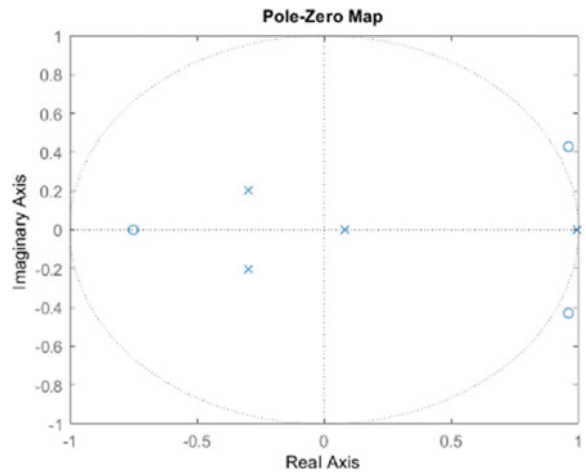


Table 1 The best parameters used in modelling hub angle

Parameters	Symbol	Value
Total number of bacteria	S	60
Chemotactic steps count	N_c	10
Swim steps count	N_s	4
Reproductive steps count	N_{re}	4
Elimination and dispersal steps count	N_{ed}	20
Probability of each bacteria to eliminate or disperse	P_{ed}	0.25
Bacteria that splits per generation	S_r	30
Step size taken by each bacterium	c	0.01
Model order		4
Number of parameters		8

Table 2 Set of parameters applied on the PID controller

Controller	K_P	K_I	K_D
PID	9	105	0.0002

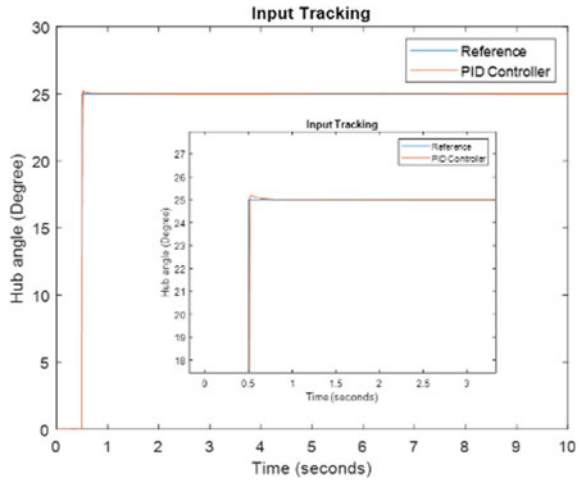
4.2 PID Controller Development in Active Vibration Control

The PID controller for hub angle system is designed in MATLAB/Simulink 2018. For this controller, the reference input has been set to end at 25° as an input of movement of the hub angle. To represent the external disturbances, a sine wave of 2 V amplitude and 1 rad/sec frequency are applied to the hub angle to validate the robustness of the controller developed. The external disturbances assess the performance of the PID controller on its robustness to determine its competitiveness in real-world application. The PID controller’s set of parameters is tuned through trial and error to provide the best possible control of the hub angle with the least amount of overshoot and settling time. The robust controller’s set of parameters is as shown in Table 2.

4.2.1 PID Controller Simulation Result Under Single Sinusoidal Disturbance

Based on the observation made on Fig. 7, the PID controller has flawlessly control the output response of the BFO model to the 25° reference line. With the introduction of single disturbance to the hub angle system, the output response is initially overshoot 0.8% from the reference line but the controller made sure that the response is overlapping with the 25° reference line at 0.5242 s.

Fig. 7 Simulation of the PID controller with single sinusoidal disturbance



4.2.2 PID Controller Simulation Result Under Multiple Sinusoidal Disturbance

Figure 8 below depicts the output response of the BFO model’s hub angle when the PID controller is used in the presence of several sinusoidal disturbances. With an exact proportional, integral, and derivative gain value as the controller and a single disturbance, the PID controller successfully directs the system’s hub angle to 25°, which is initially overshoot by 0.84% and then settles flawlessly at the reference line in 0.5252 s. The settling process ensures that the hub angle is accurately positioned at 25° and that the vibration magnitude is minimised. The results of overshoot and settling time calculations using the controller are shown in Table 3.

Fig. 8 Simulation result of the PID controller with presence of sinusoidal disturbances

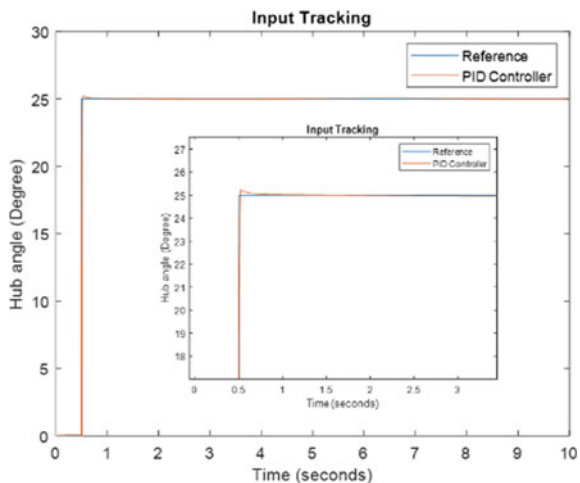


Table 3 Results of overshoot and settling time of PID controller simulations

Disturbances	Overshoot (%)	Settling time (s)
Single sinusoidal	0.8	0.5242
Multiple sinusoidal	0.84	0.5252

5 Conclusion

Bacterial Foraging optimization was successfully used to simulate the hub angle of the flexible manipulator system in this work. This study discussed both the modelling and validation process. The hub angle model was confirmed using a minimum mean squared error, exceptional stability in the pole-zero diagram, and correlations within 95% of confidence lines in both correlation tests. The model was then used to create a PID controller for the hub angle. The existence of external disturbances confirmed the controller's ability to maintain accurate positioning with the least amount of overshoot and settling time.

Acknowledgements The authors wish to express their gratitude to Universiti Teknologi MARA (UiTM), Universiti Malaysia Sarawak (Unimas), Universiti Teknologi Malaysia (UTM), and the Ministry of Higher Education (MoHE) for financially supporting and facilitating the conduct of this research. The sponsor's filing number is (RACER/1/2019/TK03/UITM/1).

References

1. Lakshmi Srinivas G, Javed A (2020) Topology optimization of rigid-links for industrial manipulator considering dynamic loading conditions. *Mech Mach Theory* 153:103979. <https://doi.org/10.1016/j.mechmachtheory.2020.103979>
2. Li Z, Wang C, Quan L, Hao Y, Ge L, Xia L (2021) Study on energy efficiency characteristics of the heavy-duty manipulator driven by electro-hydraulic hybrid active-passive system. *Autom Constr* 125:103646. <https://doi.org/10.1016/j.autcon.2021.103646>
3. Sencer B, Kakinuma Y, Yamada Y (2020) Linear Interpolation of machining tool-paths with robust vibration avoidance and contouring error control. *Precis Eng* 66:269–281. <https://doi.org/10.1016/j.precisioneng.2020.04.007>
4. Jiang X, Yabe Y, Konno A, Uchiyama M (2008) Vibration suppression control of a flexible arm using image features of unknown objects. In: 2008 IEEE/RSJ international conference on intelligent robots and systems IROS, October, pp 3783–3788. <https://doi.org/10.1109/IROS.2008.4651008>
5. Stolfi A, Gasbarri P, Misra AK (2020) A two-arm flexible space manipulator system for post-grasping manipulation operations of a passive target object. *Acta Astronaut* 175:66–78. <https://doi.org/10.1016/j.actaastro.2020.04.045>
6. Boz U, Goyal R, Skelton RE (2018) Actuators and sensors based on tensegrity D-bar structures. *Front Astron Sp Sci* 5:1–15. <https://doi.org/10.3389/fspas.2018.00041>
7. Chen W, Chen S, Hu Z, Tang J, Li H (2020) Dynamic analysis of a bevel gear system equipped with finite length squeeze film dampers for passive vibration control. *Mech. Mach Theory* 147:103779

8. Alasli A, Çetin L, Akçura N, Kahveci A, Can FC, Tamer Ö (2019) Electromagnet design for untethered actuation system mounted on robotic manipulator. *Sensors Actuators A Phys* 285:550–565. <https://doi.org/10.1016/j.sna.2018.11.027>
9. Wang L, Liu J, Yang C, Wu D (2021) A novel interval dynamic reliability computation approach for the risk evaluation of vibration active control systems based on PID controllers. *Appl Math Model* 92:422–446. <https://doi.org/10.1016/j.apm.2020.11.007>
10. Hadi MS, Darus IZM, Talib MHA, Yatim HM, Tokhi MO (2021) Vibration suppression of the horizontal flexible plate using proportional–integral–derivative controller tuned by particle swarm optimization. *J Low Freq Noise Vibr Act Contr* 40. <https://doi.org/10.1177/1461348420934636>
11. Yatim HM, Darus IZM (2018) Development of an experimental single-link flexible manipulator system. *Int J Eng Technol* 7(2):7–12. <https://doi.org/10.14419/ijet.v7i2.29.13116>
12. Pan Y, Xia Y, Zhou T, Fulham M (2017) Cell image segmentation using bacterial foraging optimization. *Appl Soft Comput J* 58:770–782. <https://doi.org/10.1016/j.asoc.2017.05.019>
13. Majumder A, Laha D, Suganthan PN (2019) Bacterial foraging optimization algorithm in robotic cells with sequence-dependent setup times. *Knowl Based Syst* 172:104–122. <https://doi.org/10.1016/j.knosys.2019.02.016>
14. Guo C, Tang H, Niu B, Boon Patrick Lee C (2021) A survey of bacterial foraging optimization. *Neurocomputing* (xxxx). <https://doi.org/10.1016/j.neucom.2020.06.142>

Position and Attitude Control of Quadrotor MAV Using Sliding Mode Control with Tanh Function



Aminurrashid Noordin, Mohd Ariffanan Mohd Basri,
Zaharuddin Mohamed, and Izzuddin Mat Lazim

Abstract This paper presents a Sliding Mode Controller (SMC) with Tanh function for position and attitude stabilization of a quadrotor micro aerial vehicle (MAV). Since this MAV weighs less than 0.1 kg, even minor changes in its environment will affect its performance. Hence a robust controller for position x , y , z , and attitude ϕ , θ , and ψ is required. The nonlinearity factor is used to construct the dynamic quadrotor model, and the Lyapunov stability function is then used to verify and guarantee the proposed control scheme. To evaluate the robustness of the proposed control method, the dynamic model was simulated in MATLAB Simulink both with and without external disturbance. In addition, the proposed control approach is compared to the PID control scheme in terms of integral square error (ISE). The results show that sliding mode control utilizing the tanh function has excellent performance and resilience against perturbation at par with PID.

Keywords Sliding mode control · Lyapunov stability · Quadrotor MAV · Perturbation

1 Introduction

Quadrotor research has grown in popularity during the last two decades. Quadrotors have a wide range of uses, including mapping, transportation, surveillance, and education, despite the fact that they are a highly under-actuated device. To pull off

A. Noordin · M. A. Mohd Basri (✉) · Z. Mohamed
School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, UTM,
81310 Johor Bahru, Johor, Malaysia
e-mail: ariffanan@fke.utm.my

A. Noordin
Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia
Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

I. Mat Lazim
Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, 71800 Nilai,
Negeri Sembilan, Malaysia

achievements such as these, the controller must be able to withstand any perturbations occur. Quadrotors are nonlinear systems with multiple inputs and outputs that have a high degree of dynamic coupling. As a result, a reliable autonomous flight capability is necessary for a quadrotor to complete the desired mission in terms of attitude control and navigation control.

The quadrotor fuselage is simple, but the under-actuated and dynamically unstable system makes designing a full controller difficult. For a very reliable attitude controller, majority of the research relied on standard techniques by developing two virtual inputs of ϕ and θ angles for a position control.

A number of control approaches have lately been implemented to address the increased demands on quadrotor efficiency, such as backstepping control [1, 2], linear feedback control [3, 4], PID control [5–7] and sliding mode control [8, 9].

Sliding mode control (SMC) is a prominent method for dealing with nonlinear quadrotor dynamic uncertainty. In [10], a control rule based on the SMC with Tanh function is proposed to improve quadrotor UAV control performance for altitude and attitude stability. The performance of the SMC utilizing the Tanh function is compared to that of other SMC control laws such as the reaching law, exponential reaching law, and saturation function in the simulation. Without or with the presence of parameter uncertainties and external disturbances, all SMC control approaches show the capacity to stable quadrotor at the specified altitude (z) and attitude (ϕ, θ, ψ).

This paper contributes SMC controller with Tanh function for small scale quadrotor with mass less than 0.1 kg focus on a position control with the sense of Lyapunov theorem that guarantee the stability of the nonlinear system.

2 Quadrotor Systems Modelling

Generalized quadrotor coordinates consist of six degrees of freedom in which (x, y, z) describe the absolute positions and (ϕ, θ, ψ) are roll, pitch and yaw angle, respectively, used to describe the orientation. Hence, the mathematical modelling of a quadrotor [11] can be divided into two sections using the Newton–Euler formulation as translational dynamics (absolute position) are represented by (1), and rotational dynamics (orientation) by (2),

$$\begin{aligned} m\ddot{x} &= U_1(S\psi S\phi + C\psi S\theta C\phi) \\ m\ddot{y} &= U_1(-C\psi S\phi + S\psi S\theta C\phi) \\ m\ddot{z} &= mg + U_1(C\theta C\phi) \end{aligned} \quad (1)$$

$$\begin{aligned} I_{xx}\ddot{\phi} &= (I_{yy} - I_{zz})\dot{\psi}\dot{\theta} - (J_r\Omega_d)\dot{\theta} + IU_2 \\ I_{yy}\ddot{\theta} &= (I_{zz} - I_{xx})\dot{\psi}\dot{\phi} + (J_r\Omega_d)\dot{\phi} + IU_3 \\ I_{zz}\ddot{\psi} &= (I_{xx} - I_{yy})\dot{\theta}\dot{\phi} + U_4 \end{aligned} \quad (2)$$

where, m is the mass, g is the gravitational coefficient, l is the lateral arm length, I is the moments of inertia for the quadrotor, the diagonal matrix 3-by-3 is defined as $I = diagonal[I_{xx} I_{yy} I_{zz}]^T$, J_r is the rotor inertia, Ω_d is total rotor speeds generated from the two pairs of rotors. U_1 is the total thrust force generated by four rotors, U_2 , U_3 and U_4 are the torque for the respected DOF of roll, pitch and yaw angles. Finally, S and C in the equation represent sin and cos function respectively.

2.1 Quadrotor State Space Representation

In order to carry out a systematic control system design procedure and for clarity of notation, the dynamics described in (1) and (2) are rewritten as a standard system of second order state space equations:

$$\ddot{X} = f(X) + g(X)U \tag{3}$$

where X and U represent the state vectors and input vectors respectively described as follows:

$$X = [x_1 x_2 x_3 x_4 x_5 x_6]^T = [x \ y \ z \ \phi \ \theta \ \psi]^T \tag{4}$$

and

$$U = [U_1 U_2 U_3 U_4]^T \tag{5}$$

The nonlinear function of $f(X)$ and $g(X)$ can be rewritten as:

$$f(X) = \begin{pmatrix} 0 \\ 0 \\ g \\ a_1 \dot{\psi} \dot{\theta} + a_2 \Omega_d \dot{\theta} \\ a_3 \dot{\psi} \dot{\phi} + a_4 \Omega_d \dot{\phi} \\ a_5 \dot{\theta} \dot{\phi} \end{pmatrix}, g(X) = \begin{pmatrix} u_x m^{-1} & 0 & 0 & 0 \\ u_y m^{-1} & 0 & 0 & 0 \\ u_z m^{-1} & 0 & 0 & 0 \\ 0 & b_1 & 0 & 0 \\ 0 & 0 & b_2 & 0 \\ 0 & 0 & 0 & b_3 \end{pmatrix} \tag{6}$$

where $a_1 = (I_{yy} - I_{zz})I_{xx}^{-1}$, $a_2 = -J_r I_{xx}^{-1}$, $a_3 = (I_{zz} - I_{xx})I_{yy}^{-1}$, $a_4 = J_r I_{yy}^{-1}$, $a_5 = (I_{xx} - I_{yy})I_{zz}^{-1}$, $b_1 = l I_{xx}^{-1}$, $b_2 = l I_{yy}^{-1}$, $b_3 = I_{zz}^{-1}$, $u_x = U_1(S\psi S\phi + C\psi S\theta C\phi)$, $u_y = U_1(-C\psi S\phi + S\psi S\theta C\phi)$, and $u_z = U_1(C\theta C\phi)$.

3 Control Design

Sliding mode based on reaching law include reaching phase and sliding phase. The reaching phase drive system to stable manifold, the sliding phase drive system slide to equilibrium. Chattering can be caused by a discontinuity in the sign function in the SMC law. As a result, the discontinuous sign function can be substituted by the continuous Tanh function to eliminate chattering. The sign function can be approximated by the Tanh function depending on its steepness [12]. Following is a description of the control design formulation for sliding mode control utilizing the Tanh function. A sliding surface and it derivative is chosen as:

$$s = \dot{e} + \lambda e, \dot{s} = \ddot{e} + \lambda \dot{e} \quad (7)$$

where, $\lambda > 0$, to satisfy the Hurwitz condition. Then tracking error and its derivative value is defined as:

$$e = X_d - X, \dot{e} = \dot{X}_d - \dot{X}, \ddot{e} = \ddot{X}_d - \ddot{X} \quad (8)$$

Replacing (8) and (3) into (7) gives

$$U = g^{-1}(\ddot{X}_d - f + \lambda(\dot{X}_d - \dot{X}) - \dot{s}) \quad (9)$$

For the reduction of high frequency chattering, \dot{s} can be replaced by continuous tanh function as follows:

$$\dot{s} = -\eta s - k.tanh(s\epsilon^{-1}), \eta > 0, k > 0, \epsilon > 0 \quad (10)$$

where $-\eta s$ is an exponential term, and its solution is $s = s(0)e^{-\eta t}$. Clearly, by adding the proportional rate term $-\eta s$, the state is forced to approach the switching manifolds faster when s is larger. The steepness of the tanh function determines how the tanh can approximate the sign function. The hyperbolic tangent function is defined as:

$$\tanh(x\epsilon^{-1}) = \frac{e^{\frac{x}{\epsilon}} - e^{-\frac{x}{\epsilon}}}{e^{\frac{x}{\epsilon}} + e^{-\frac{x}{\epsilon}}}, \epsilon > 0 \quad (11)$$

where, the steepness of the tanh function is determined by ϵ value [12].

3.1 Stability Analysis

For stability, Lyapunov function $V = 2^{-1}s^2$ is chosen. Thus, we get

$$\dot{V} = s\dot{s} \quad (12)$$

Replace (10) into (12) we get

$$\dot{V} = -k.s.tanh(s\epsilon^{-1}) - \eta s^2 \quad (13)$$

For stability $\dot{V} < 0$, therefore (13) can be split into two terms and can be described as

$$\dot{V} = \dot{V}_1 + \dot{V}_2 \quad (14)$$

For the first term of (14)

$$\dot{V}_1 = -k.s.tanh(s\epsilon^{-1}) \quad (15)$$

According to Lemma 1.1 [13], we have

$$|s| |\tanh(s\epsilon^{-1})| \geq 0 \quad (16)$$

For the second term of (14)

$$\dot{V}_2 = -\eta s^2 \quad (17)$$

replace $s^2 = 2V$ into (17) gives

$$\dot{V}_2 = -2\eta V_2 \quad (18)$$

Use Lemma 1.2 [13], and choose $\alpha = 2\eta$ and $f = 0$; we obtained

$$V_2(t) \leq e^{-2\eta(t-t_0)} V_2(t_0) \quad (19)$$

Therefore, if η is a positive constant value, $V_2(t)$ tends to be zero exponentially with the value of η . Therefore, when $\dot{V}_1 \leq 0$ and $\dot{V}_2 \leq 0$, it proves that $\dot{V} \leq 0$ guarantees the stability of the system.

3.2 Attitude and Altitude Control

From (9), the control law for altitude and attitude using tanh function can be described as follows:

$$\begin{aligned} U_1 &= mu_z^{-1}(\ddot{z}_d - g + \lambda_z(\dot{z}_d - \dot{z}) + \eta_z s_z + k_z tanh(s_z \epsilon_z^{-1})) \\ U_2 &= b_1^{-1}(\ddot{\phi}_d - a_1 \dot{\psi} \dot{\theta} - a_2 \ddot{\theta} + \lambda_\phi(\dot{\phi}_d - \dot{\phi}) + \eta_\phi s_\phi + k_\phi tanh(s_\phi \epsilon_\phi^{-1})) \\ U_3 &= b_2^{-1}(\ddot{\theta}_d - a_3 \dot{\psi} \dot{\phi} - a_4 \Omega_d \dot{\phi} + \lambda_\theta(\dot{\theta}_d - \dot{\theta}) + \eta_\theta s_\theta + k_\theta tanh(s_\theta \epsilon_\theta^{-1})) \end{aligned}$$

$$U_4 = b_3^{-1} \left(\ddot{\psi}_d - a_5 \dot{\theta} \dot{\phi} + \lambda_{\psi} (\dot{\psi}_d - \dot{\psi}) + \eta_{\psi} s_{\psi} + k_{\psi} \tanh(s_{\psi} \epsilon_{\psi}^{-1}) \right) \quad (20)$$

3.3 Position Control

The position controller's purpose is to follow a predetermined path while maintaining zero angular velocities. The commanded x and y accelerations are used to calculate the desired pitch and roll angles due to the coupling between x dynamics and the θ angle and also between y dynamics and the ϕ angle. The following sliding mode with tanh function is used to determine the required acceleration x and y .

$$\begin{aligned} \ddot{x}_d &= -\eta_x s_x - k_x \cdot \tanh(s_x \epsilon_x^{-1}) \\ \ddot{y}_d &= -\eta_y s_y - k_y \cdot \tanh(s_y \epsilon_y^{-1}) \end{aligned} \quad (21)$$

To acquire the necessary roll and pitch angles, the desired accelerations (6) are inverted with small angles approximation applied to ϕ and θ .

$$\begin{aligned} \phi_d &= m u_1^{-1} (\ddot{x}_d \sin(\psi_d) - \ddot{y}_d \cos(\psi_d)) \\ \theta_d &= m u_1^{-1} (\ddot{x}_d \cos(\psi_d) - \ddot{y}_d \sin(\psi_d)) \end{aligned} \quad (22)$$

4 Simulation Results

The quadrotor model is simulated in MATLAB Simulink. The performance of SMC with Tanh controller is compared to classical PID controller. A PID controllers are found in a wide range of applications in control system and has proven to be robust in many linear or nonlinear applications [14]. Therefore, the robustness of the proposed system is compared to it. The parameters value used for quadrotor's dynamics are tabulated in Table 1. A few tests were conducted in order to evaluate the performance of the designed SMC with Tanh. Integral Square Error (ISE) is chosen as the index performance. Initially the quadrotor is set at $\zeta_i = 0 [i = x, y, z, \phi, \theta, \psi]$. The PID, proportional gain, derivative gain, and integral gain were manually tuned to obtain satisfactory results and settling times. The desired circle trajectory given to the systems as follows:

$$\begin{aligned} x_d &= A_x \cos(\pi t / 30) \\ y_d &= A_y \sin(\pi t / 30) \end{aligned} \quad (23)$$

where $A_x = A_y = 1$, and $t = 60$ s.

Table 1 Parameters' for x -mode configuration quadrotor

Specification	Parameter	Unit	Value
Quadrotor mass	m	kg	0.063
Lateral moment arm	l	m	0.0624
Thrust coefficient	b	Ns^2	0.0107
Drag coefficient	d	Nms^2	0.78264×10^{-3}
Rolling moment of inertia	I_{xx}	kgm^2	0.0000582857
Pitching moment of inertia	I_{yy}	kgm^2	0.0000716914
Yawing moment of inertia	I_{zz}	kgm^2	0.0001
Rotor moment of inertia	J_r	kgm^2	0.1021×10^{-6}

4.1 Experiment 1: Circle Trajectory Without Perturbation

The simulation results presented in Fig. 1, demonstrate the greater performance of the proposed SMC with Tanh compared to Classical PID Control. Both controllers were able to follow the desired trajectory, however SMC with Tanh gave 0.01516, 1.1630, and 1.6160 at x , y , and z , respectively, based on the index performance in Table 2. While the PID controller shows 0.1733, 1.5303, and 1.6208 on x , y , and z respectively. In addition, clearly shows in Y position at Fig. 1, there is slightly overshoot produces by PID compared to a smooth response by SMC with Tanh.

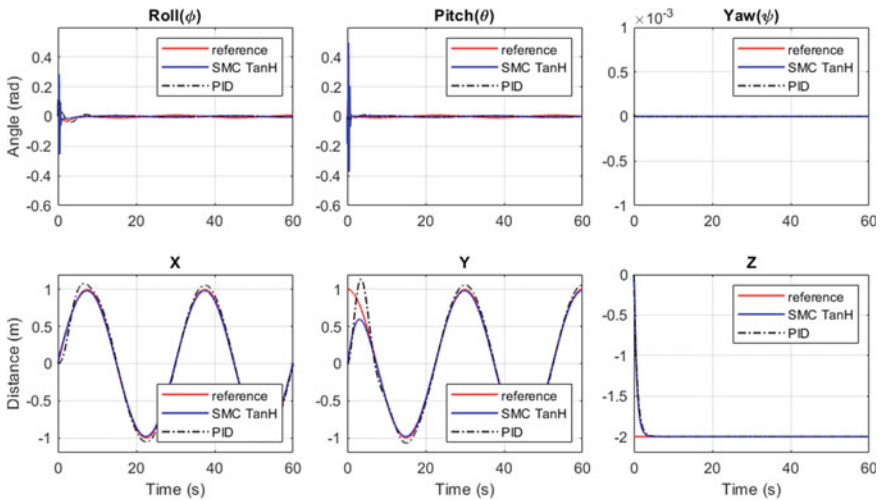


Fig. 1 Position and attitude tracking without perturbation

Table 2 Performance index using IAE without and with perturbation on Altitude

ISE	Controller	X	Y	Z	Roll	Pitch	Yaw
Without perturbation	SMC TanH	0.01516	1.1630	1.6160	0.01552	0.03924	$6.949e^{-34}$
	PID	0.17330	1.5303	1.6280	$0.384e^{-3}$	$21.490e^{-6}$	$33.209e^{-12}$
With perturbation	SMC TanH	0.01532	1.1640	1.6190	0.01502	0.03919	$9.734e^{-07}$
	PID	0.17340	1.5299	1.6343	$0.3836e^{-3}$	$0.2546e^{-5}$	$0.2283e^{-6}$

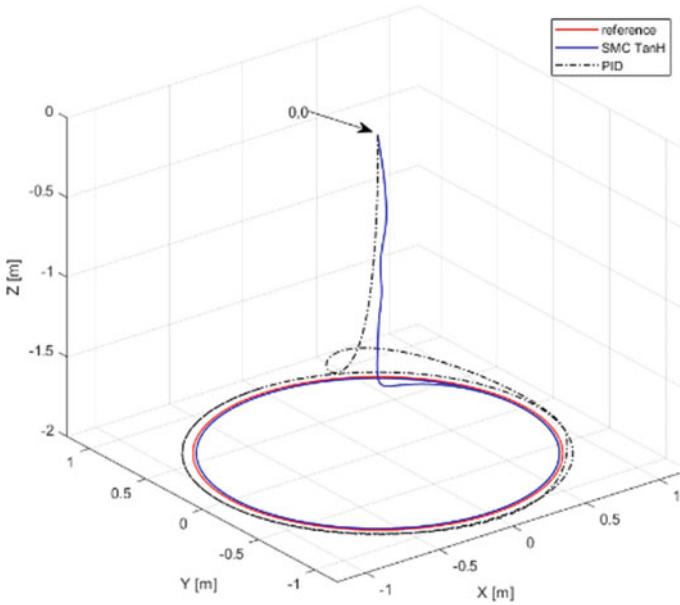


Fig. 2 Circle trajectory without perturbation

Figure 2 describes the circle trajectory performed by both controllers which indicates that using SMC with Tanh, the quadrotor able to hover better on Z axis compared to PID controller before smoothly follow the circle trajectory. Comparison of input control signal between PID and SMC with Tanh to perform the circle trajectory within 60 s is shown in Fig. 3.

4.2 Experiment 2: Circle Trajectory with Perturbation

From state space, with uncertainties and disturbances, (4) can be written to

$$\ddot{X} = (f_n + \Delta f) + (g_n + \Delta g)U + d = f_n + g_n U + w \tag{24}$$

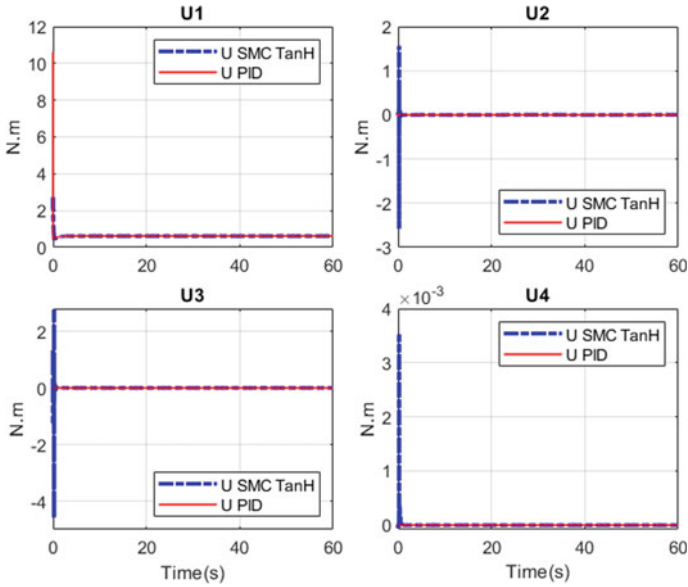


Fig. 3 Control input signal for the quadrotor without perturbation

where d is the external disturbance i.e. wind, Δf and Δg are the system uncertainties, w is the lump of uncertainties and defined as $w = \Delta f + \Delta g u + d$ with the assumption $|w| \leq W$, a positive constant. Hence, for this simulation the external disturbance for attitude ϕ , θ , and ψ d is taken as normal Gaussian noise [12].

$$d = [\mathcal{N}(0, 0.5)\mathcal{N}(0, 0.5)\mathcal{N}(0, 0.5)\mathcal{N}(0, 0.5)] \tag{25}$$

In addition, after 15 s hovering, a pulse of four Newton is injected to z-axis every 20 s to see the robustness of the SMC controller to handle slightly force for small scale MAV (less than 0.1 kg).

Figure 4 clearly shows the quadrotor with SMC Tanh is robust against perturbation and able to withstand the noise on the roll, pitch and yaw while having force applied on z-axis during movement. Figure 5 shows the circle movement, while Fig. 6 shows the perturbation on the input signal. In addition, there is less chattering phenomenon appear shows the greatest of SMC with Tanh function compared to other SMC techniques such as reaching law, exponential reaching law, and saturation function [13].

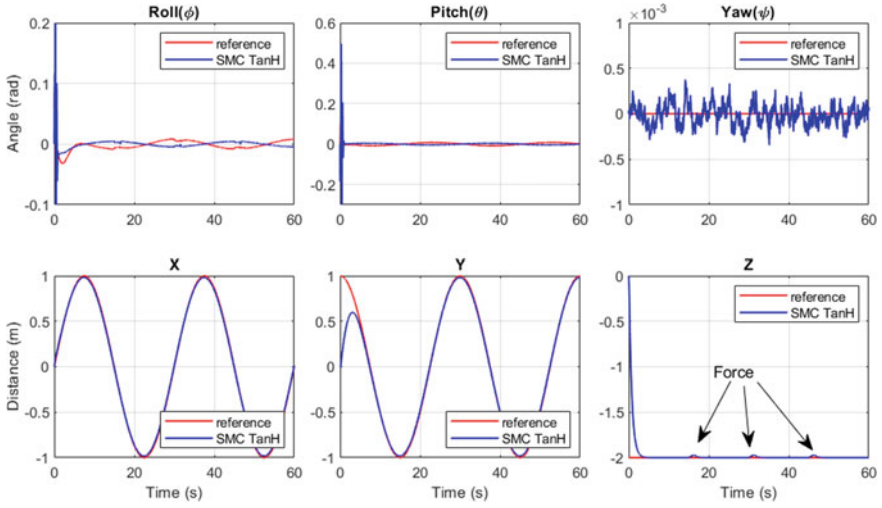


Fig. 4 Position and attitude tracking with perturbation for SMC TanH

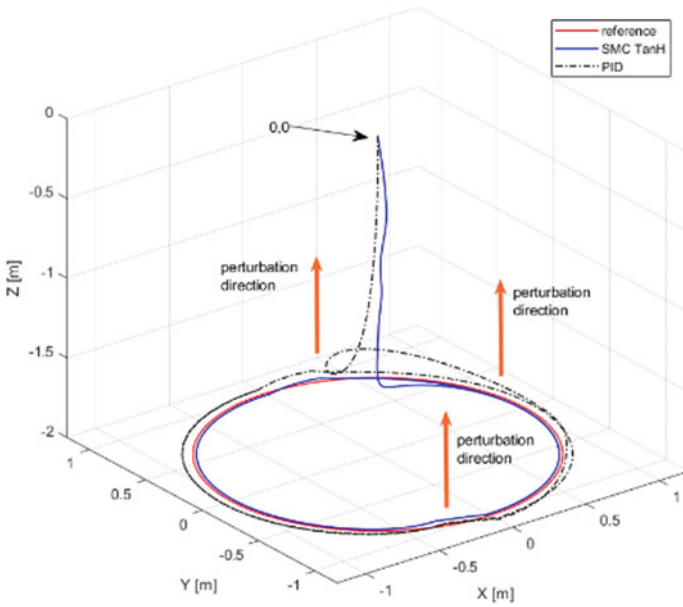


Fig. 5 Circle trajectory with perturbation

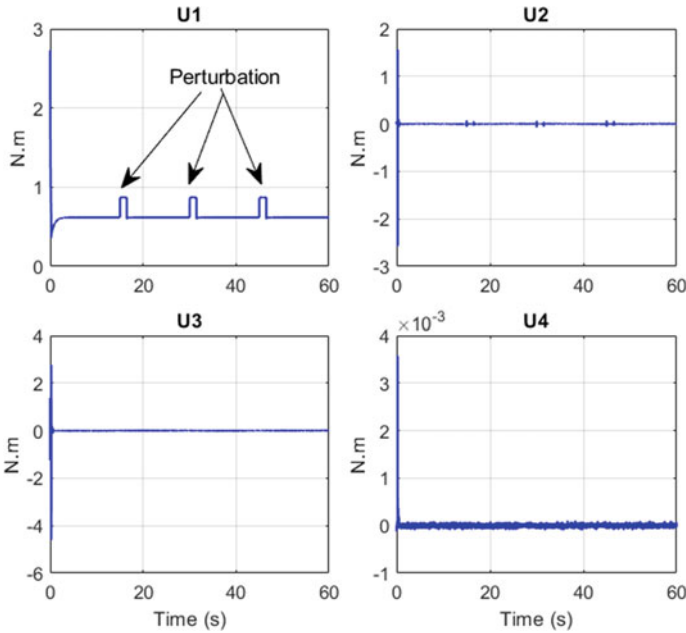


Fig. 6 SMC Tanh control input signal with perturbation

5 Conclusion

In this paper, a control law using Sliding Mode Controller with Tanh function as sliding manifold is proposed for full control of quadrotor MAV in position and attitude. The simulation of the proposed SMC system was conducted in a MATLAB Simulink environment, where the results demonstrated a greatest performance without or with perturbation (slightly force) applied on altitude z and white noise applied on attitude ϕ , θ , and ψ respectively. The performances of the proposed control technique are measured using ISE. In the presence of perturbation, the index for position x , y , z give 0.01532, 1.1640, 1.6190 for SMC with Tanh while 0.17340, 1.5299, 1.6343 for PID. Then the index for orientation ϕ , θ , ψ give 0.01502, 0.03919, 9.734×10^{-7} for SMC with Tanh while 0.3836×10^{-3} , 0.2546×10^{-5} , 0.2283×10^{-6} for PID. In the future, optimization techniques can be applied to optimize the controller’s parameter to obtain better performance in the presence of perturbation. Furthermore, investigation of other controller such as super-twisting algorithm or adaptive control scheme can be investigated on a MAV quadrotor.

Acknowledgements The authors would like to thank Universiti Teknologi Malaysia (UTM) under the Research University Grant (R.J130000.2651.17J42), Universiti Teknikal Malaysia Melaka (UTeM), and Ministry of Education Malaysia for supporting this research.

References

1. Basri MAM, Husain AR, Danapalasingam KA (2014) Enhanced backstepping controller design with application to autonomous quadrotor unmanned aerial vehicle. *J Intell Robot Syst Theory Appl* 79:295–321. <https://doi.org/10.1007/s10846-014-0072-3>
2. Mohd Basri MA, Husain AR, Danapalasingam K (2015) Intelligent adaptive backstepping control for MIMO uncertain non-linear quadrotor helicopter systems. *Trans Inst Meas Control* 37:345–361. <https://doi.org/10.1177/0142331214538900>
3. Lazim IM, Rashid Husain A, Adilla Mohd Subha N, Ariffanan Mohd Basri M (2018) Intelligent observer-based feedback linearization for autonomous quadrotor control. *Int J Eng Technol* 7:904. <https://doi.org/10.14419/ijet.v7i4.35.26280>
4. Lazim IM, Husain AR, Ariffanan M, et al (2018) feedback linearization with intelligent disturbance observer for autonomous quadrotor with time-varying disturbance. *Int J Mech Mechatron Eng IJMME-IJENS*
5. Monroy JA, Campos E, Torres JA (2017) Attitude control of a micro AUV through an embedded system. *IEEE Lat Am Trans* 15:603–612
6. Yenugula BR, Ziz-ur-Rahman M (2017) Stability control structure of hovercraft prototype utilising PID controller. *Bull Electr Eng Inform* 6:348–350. <https://doi.org/10.11591/eei.v6i4.864>
7. Cedro L, Wiczorkowski K (2019) Optimizing PID controller gains to model the performance of a quadcopter. *Transp Res Procedia* 40:156–169. <https://doi.org/10.1016/j.trpro.2019.07.026>
8. Noordin A, Basri MAM, Mohamed Z (2019) Sliding mode control for altitude and attitude stabilization of quadrotor UAV with external disturbance. *Indones J Electr Eng Inform* 7:203–210. <https://doi.org/10.11591/ijeei.v7i2.1149>
9. Castañeda H, Gordillo JL (2019) Embedded flight control based on adaptive sliding mode strategy for a quadrotor micro air vehicle. *Electronics* 8:793. <https://doi.org/10.3390/electronics8070793>
10. Noordin A, Basri MAM, Mohamed Z (2021) Sliding mode control with tanh function for quadrotor UAV altitude and attitude stabilization. In: Bahari MS, Harun A, Zainal Abidin Z, et al (eds) *Intelligent manufacturing and mechatronics*. Springer, Singapore, pp 471–491
11. Noordin A, Basri MAM, Mohamed Z (2020) Simulation and experimental study on pid control of a quadrotor MAV with perturbation. *Bull Electr Eng Inform* 9:1811–1818. <https://doi.org/10.11591/eei.v9i5.2158>
12. Jinkun L (2017) *Sliding mode control using MATLAB*. Academic Press, Elsevier
13. Noordin A, Basri MAM, Mohamed Z (2019) Sliding mode control for altitude and attitude stabilization of quadrotor UAV with external disturbance. In: Bahari MS, Harun A, Zainal Abidin Z, et al (eds) *Indonesian journal of electrical engineering and informatics*. Springer, Singapore, pp 203–210
14. Noordin A, Basri MAM, Mohamed Z, Abidin AFZ (2017) Modelling and PSO fine-tuned PID control of quadrotor UAV. *Int J Adv Sci Eng Inf Technol* 7. <https://doi.org/10.18517/ijaseit.7.4.3141>

On the Partial Transmit Sequences (PTS) Optimization Using Firefly Algorithm for PAPR Reduction in OFDM Systems



Hazmarini Husin, Rosmiwati Mohd-Mokhtar, and Aeizaal Azman A. Wahab

Abstract A system called Orthogonal frequency division multiplexing (OFDM) is frequently applied in communication system because of its advantages in terms of high spectral performance, resistance to inter-symbol interferences (ISIs), fast implementation, and multipath transmission. Unfortunately, OFDM signals generate a high peak-to-average power ratio (PAPR) and that is the main weakness in this system. In this project, Partial Transmit Sequence (PTS) optimization using the of Firefly Algorithm (FA) is applied to decrease the high PAPR. The analysis of PAPR minimization by the proposed method will be carried out by using different number of subblocks, subcarriers and iterations. Two different modulation types will be used which are 16 Quadrature Amplitude Modulation (QAM) and Quadrature Phase Shift Keying (QPSK). The CCDF and BER plots are evaluated at the end of the process. The performance of OFDM signal and PTS-FA through AWGN channel is examined. The result of this project has shown that conventional PTS only managed to reduce PAPR values by 4 - 5 dB, while using FA, it effectively reduced PAPR by 7 - 10 dB. The computational complexity is observed from simulation of different iterations and low search complexity is preserved by adjusting the FF algorithm parameters.

Keywords OFDM system · PAPR reduction · 16QAM · QPSK modulation

1 Introduction

1.1 Research Background

In wireless communication system, the most prominent multicarrier modulation technique that being used is known as Orthogonal Frequency Division Multiplexing (OFDM) [1, 2]. OFDM is extensively utilized in various applications such as Long-term evolution (LTE), LTE-Advance, Digital Audio Broadcasting (DAB),

H. Husin · R. Mohd-Mokhtar · A. A. A. Wahab (✉)

School of Electrical and Electronic, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia

e-mail: aeizaal@usm.my

High-Definition Television (HDTV), Digital Video Broadcasting (DVB) and power line communications [1]. Other than that, the crucial present standards of wireless communication such as IEEE 802.16 m, 802.20, 802.22 are implementing the OFDM scheme as well due to its advantages [3]. In conclusion, it has been chosen as the best approach to be applied due to the better channeling of rate data and great power efficacy [4]. However, this multicarrier system generates high peak-to-average power ratio (PAPR) which degrades its performance [4]. Since OFDM signals are multicarrier, it typically has amplitude variance in the time domain and a larger range of dynamic, which introduces the PAPR [2]. This high PAPR restrained the potential of the OFDM because of the distortion produced by nonlinear characteristics of the High-Power Amplifier (HPA) and the Analog-to-Digital Converter (ADC) [5]. Apart from that, the high PAPR causes the power amplifier to become drenched and produces interferences in subcarriers. Thereby, it makes trouble on the signal of the spectrum and lowering the BER output. It can be solved by decreasing the high peak power of the signal [1].

1.2 Problem Statement

As previously mentioned, the huge drawback of OFDM system is suffering from the excessive PAPR. The outrange PAPR will cause the OFDM to be clipped when it travels through a nonlinear high-power amplifier (HPA) and in consequence, the in-band distortion and out-of-band radiation will be existed which can reduce the performance. As a result, the OFDM transmitters requisite large dynamic range with a costly linear HPA [2]. Besides, the nonlinear distortions that caused by this high PAPR is introduced the inter-modulation, spreading of the spectral, and alterations in the constellation of the signal which makes the power efficiency reduced significantly. In order reduce the PAPR, effective techniques are crucial for the fastest speed of wireless communication systems. Aside from that, the elimination of the non-linear issue and the increased of power amplifiers proficiency are also aided by an alleviation in PAPR [6]. Among various techniques, PTS is mostly used since it shows better performance in reducing the high PAPR [1, 4, 5].

Partial Transmit Sequence (PTS) has less distortion and more effective to alleviate PAPR since it employs an iterative method to evaluate the optimal phase factors without limiting the number of subcarriers. In conventional PTS algorithm, however, the complexity is increased since it necessitates a comprehensive search of the authorized phase factors [2]. PTS scheme of low-density parity check coded OFDM systems is developed for PAPR reduction. But the computational complexity is not considered by the drawback of this scheme. To reduce the complexity, PTS sub-blocking method is introduced which is using only partial IFFT's. Although it attains the complexity minimization, it lessens the efficiency and increases the complexity of power amplifier due to the similarity of original PTS scheme and PAPR output [7]. Therefore, in this project, a swarm intelligence algorithm for phase optimization

based on the firefly PTS (FA-PTS) algorithm has been proposed for PAPR reduction. In contrast to traditional PTS, phase optimization scheme of PTS-FA is able to accomplish a superior PAPR minimization performance besides it needs fewer computational complexity for a huge sub-blocks value [6].

2 Literature Review

M. Hosseinzadeh Aghdam and A. A. Sharifi [2] was introduced PTS based on advanced PSO to minimize the higher PAPR in the OFDM system and minimum computational complexity. The simulation was performed with different PTS optimization such as genetic algorithm (GA) and PSO. The result showed that both techniques managed to reduce PAPR from 10.7 dB to almost 7.16 dB for GA and 6.59 dB for proposed PSO. It can be summarized that PSO is the good tool to decrease PAPR as well as optimizing the search complexity. However, PSO is not good compared to the other evolutionary techniques [6]. A. A. Sharifi and H. Emami [4] proposed the improved flower pollination (IFP) algorithm that originated from pollination behavior of the flowers to search the maximum PTS phase rotation factors to minimize the PAPR of the ACO-OFDM signals. The proposed technique obtained better out-of-band radiation, and achieved an outstanding overall presentation compared to companding and iterative clipping techniques.

3 Methodology

3.1 Methodology

Project begins with research and studies on concept of OFDM, PAPR and PTS-Firefly Algorithm technique. The next step is identifying which parameters are suitable for this project and then the proposed technique is applied together with the chosen parameter. Simulation will be conducted using the proposed technique and the result will be observed and evaluated. The flowchart for the overall project can be seen in Fig. 1.

3.2 Software Description

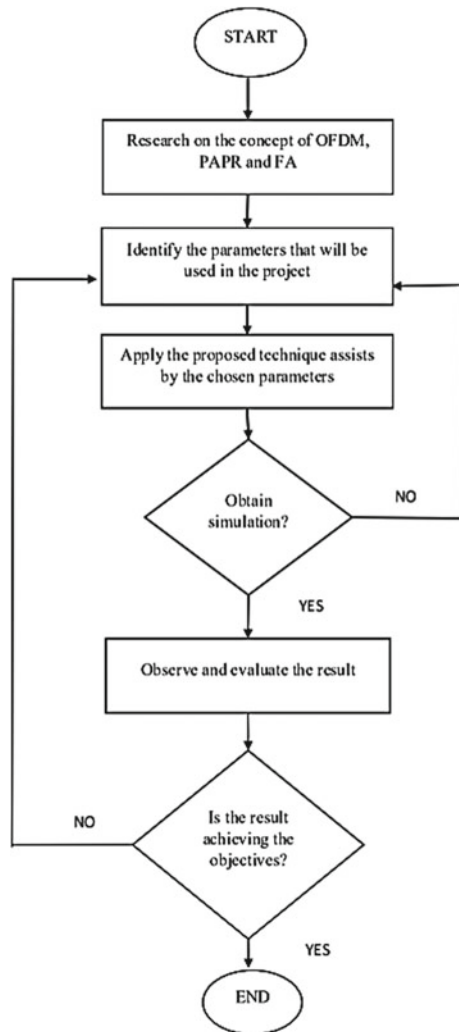
In this project, MATLAB is used for simulation of OFDM in two different modulation which are QPSK and 16QAM. The main objective is to minimize the high PAPR in OFDM by using PTS-Firefly Algorithm method. The difference before and after using PTS-FA of PAPR values will be shown on the command window.

3.3 Parameters Implementation

Some variable parameters are manipulated for the analyzing of the results. The input parameters that will be used are the values of sub-carriers, sub-blocks and iterations. The parameters to be analyzed are the value of PAPR, CCDF and BER.

Table 1 shows the parameters that will be used for PTS-FA. The previous researcher had been conducted a project that related to Firefly algorithm [6], and a relevant population size of $Gn = 10$ was chosen since it achieved better tradeoff for the PAPR minimization and complexity. Besides, by setting randomization of 0.4, minimum attractiveness of 0.20 and absorption of 1.0, the algorithm can converge

Fig. 1 Flowchart for the whole project



with good quality solution and it is outperformed as opposite to the ordinary PTS. Moreover, few variables of iteration value are selected to observe its ability in diminishing the PAPR as well as the computational complexity. Dimensions of fireflies also affects the result. The larger the dimension value, the graph can be seen clearer and better. Hence $d = 1000$ is set to observe the better graph. Finally, the cooling factor in most applications 15 used range of 0.95 to 0.97 as the best value since it is correlated to randomness which also help to speed up the convergence.

Table 2 and Table 3 shows the parameters used in 16 QAM and QPSK modulation respectively. Three different number of sub-blocks and sub-carriers are used as input data to examine the PAPR performance. OFDM blocks of 1000 is selected to get the better performance and to reduce the extended execution time. The oversample factor is adjusted by 4 for precise computation of PAPR. The constellation size used

Table 1 FA-PTS simulation parameters

Simulation parameters	Type/Values
Dimension of fireflies (d)	1000
Population of fireflies (G_n)	10
Randomization parameter (α)	0.40
Firefly attractiveness co-efficient (β)	0.20
Absorption co-efficient (γ)	1.0
No. of iteration (K)	50, 100, 200
Cooling factor (δ)	0.97

Table 2 16QAM modulation parameters

Simulation parametrs	Type/Values
Number of sub-blocks (M)	4, 8, 16
Number of sub-carriers (N)	64, 128, 256
Bit per symbol	4
Oversampling factor	4
OFDM blocks	1000
Phase factor	(+1 -1)
Constellation size	16-QAM

Table 3 QPSK modulation parameters

Simulation parametrs	Type/Values
Number of sub-blocks (M)	4, 8, 16
Number of sub-carriers (N)	64, 128, 256
OFDM blocks	1000
Oversampling factor	4
Phase factor	(+1 -1)
Modulation type	QPSK

Table 4 Comparison of improvement of PAPR reduction between conventional PTS and PTS-FA when $M = 8$

N	IMPROVEMENT OF PAPR REDUCTION IN OFDM SYSTEM	
	PTS	PTS-FA
64	4.1793 dB	10.727655 dB
128	4.21119 dB	10.749754 dB
256	4.21096 dB	10.7604436 dB

in QAM modulation is 16 and bit per symbol is set to 4 to ensure low possibility of data errors.

3.4 Project Requirement

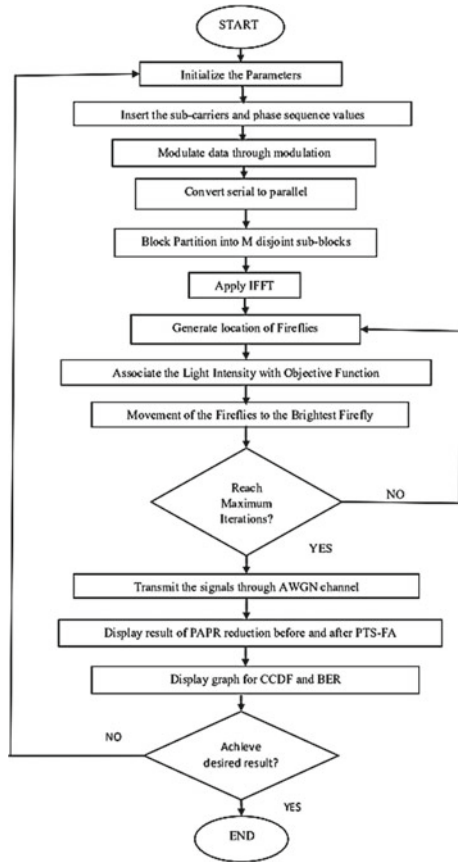
This project is fully software simulation and MATLAB R2021a is preferred to simulate the algorithm of the presented method. MATLAB is appointed for this project due to its advantages such as easy to implement algorithm and perform substantial data analysis such as calculations and matrix supports. It also useful specially to observe the PAPR values after implementing the PTS-FA. Besides, the plots of CCDF can be observed in two different modulation schemes as well as BER. The effects also will be analyzed.

3.5 Project Design

The project simulation is done in two different modulation schemes which are 16QAM and QPSK. Different modulation type has slightly different parameters and the PTS-FA parameters are used in both modulations. The simulation process is the same for both except QPSK will display the results of bit error rate (BER).

Figure 2 shows the flowchart for overall project design and the implementation of algorithm is displayed in Fig. 3. The simulation starts by initializing the selected parameters. Next, the input data: subcarrier value and subblock value will be keyed in into the system. Once those values are inserted, the OFDM signals will be modulated. Next, OFDM signals will be converted from serial to parallel. Input data block splits into dissociate sub-blocks of M and they are orthogonal among each and another. The IFFT operation is applied to frequency-domain OFDM symbols and converts them to time-domain symbols. Once signals have been changed to domain symbols, the location of firefly will be generated and the population of fireflies are initialized by 10 fireflies. The association occurs between light intensity and objective function, where the objective function is to find signal with slightest PAPR values. The fireflies will move to the attractive (brighter) firefly. This step will repeat until it achieves the maximum iteration. The signals then will be transmitted through the AWGN channel

Fig. 2 Flowchart for project design



and FFT is applied to change it into frequency domain. The next process is converting signals into the initial signals by demodulation and the results of PAPR reduction will be displayed. It will be verified using CCDF. The desired results are obtained in 16QAM and QPSK modulation. The BER performance also will be observed in QPSK modulation.

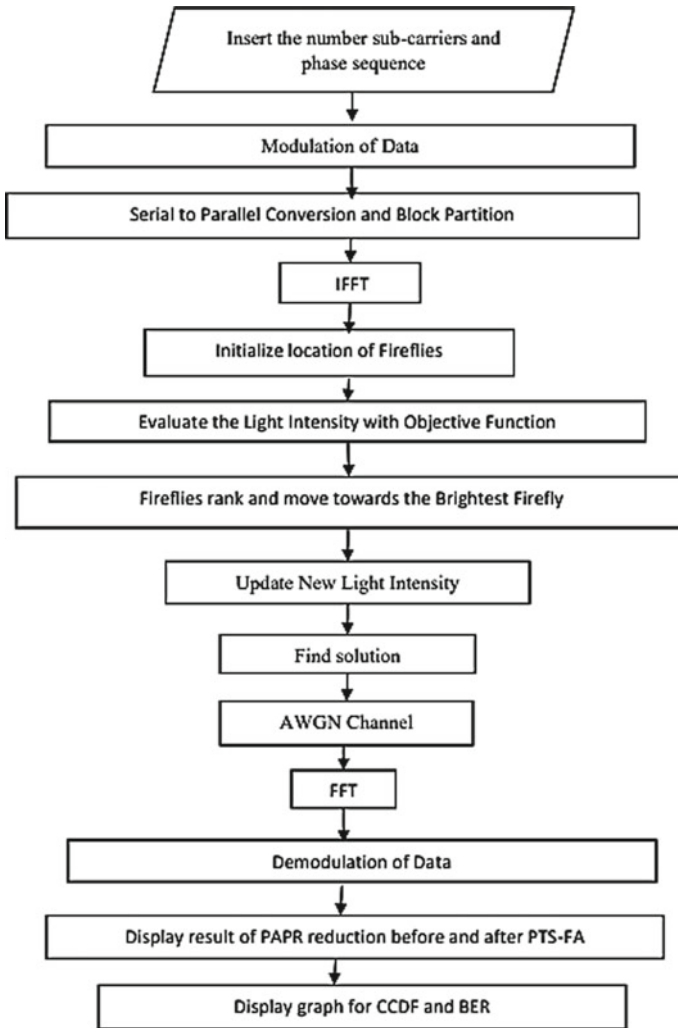


Fig. 3 Implementation of PTS optimization using FF algorithm

4 Results and Discussions

There are five cases to examine the performance of PAPR. First case is using differ subcarriers and constant subblock. Second case is applying varied subblocks and fixed subcarrier. The third case is employing both subcarriers and subblocks and its values are increasing subsequently. Fourth case is utilizing different iterations and last case is by observing BER performance through AWGN medium.

4.1 Simulation Results in 16QAM Modulation

All diagrams above represent the PAPR reduction using fixed subblock which is $M = 8$ and using subcarriers of 64, 128 and 256. The propose of using various subcarriers in the simulation is to observe and analyze the minimization of PAPR in different subcarriers while maintaining the same number of subblock. Figure 4, Fig. 5 and Fig. 6 shows the reduction of PAPR in conventional PTS and implementation of proposed method. Figure 10 displays differentiation in respect of improvement of PAPR reduction using conventional PTS and PTS-FA. By using the conventional PTS, it is able to decline PAPR value from the OFDM signals by 4.179 dB for $N = 64$, 4.211 dB for $N = 128$ and 4.210 dB for $N = 256$ respectively with CCDF

Fig. 4 PAPR performance when $N = 64$

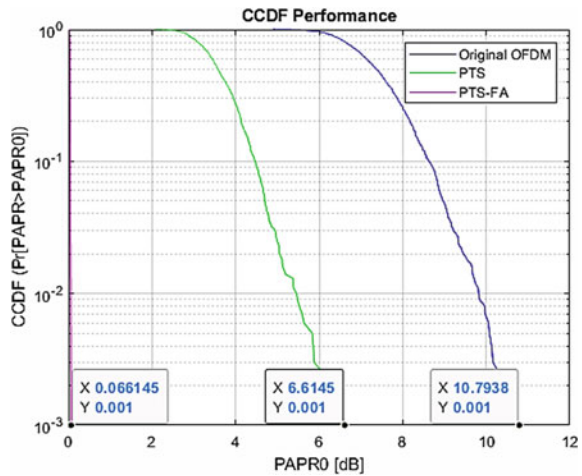


Fig. 5 PAPR performance when $N = 128$

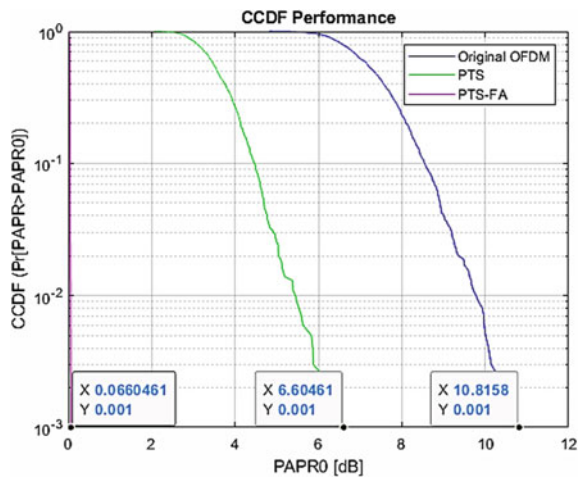


Fig. 6 PAPR performance when $N = 256$

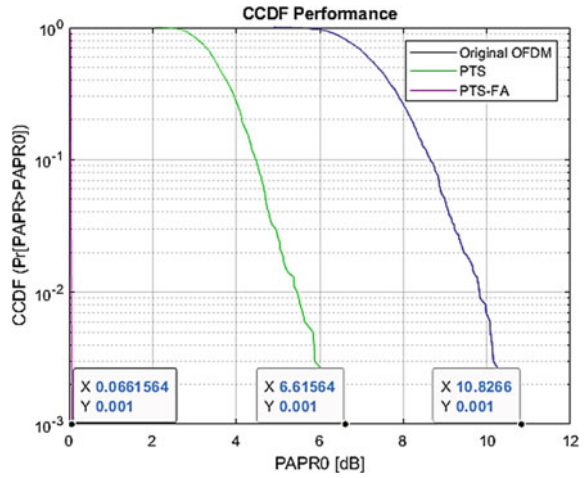
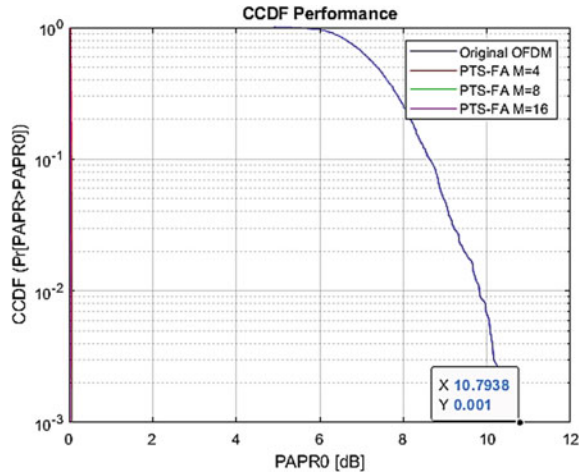


Fig. 7 PAPR performance using PTS-FA for $N = 128$ and $M = 4, 8, 16$ subblocks



of 10 – 3. Meanwhile, by implementing firefly algorithm in PTS, it displays better reduction in PAPR values by 10.73 dB for $N = 64$, 10.75 dB for $N = 128$ and 10.76 dB for $N = 256$. Based on the observation, the PAPR value increases when the number of subcarriers set is larger. This is because more data to be transmitted and the high possibility of noise to be occurred during transmission besides the subblock used is constant value, thus, the values of PAPR reduction for every subcarrier is almost similar. However, the PTS-FA technique provides better enhancement in the PAPR performance in contrast to the traditional PTS. From the simulation, it could be concluded that the PTS-FA is the productive method to decrease the PAPR value even with large number of subcarriers.

Figure 7 represents the CCDF graph of the PAPR performance in PTS-FA and Fig. 8 is zoomed in version for the same subcarrier of $N = 128$ and different number

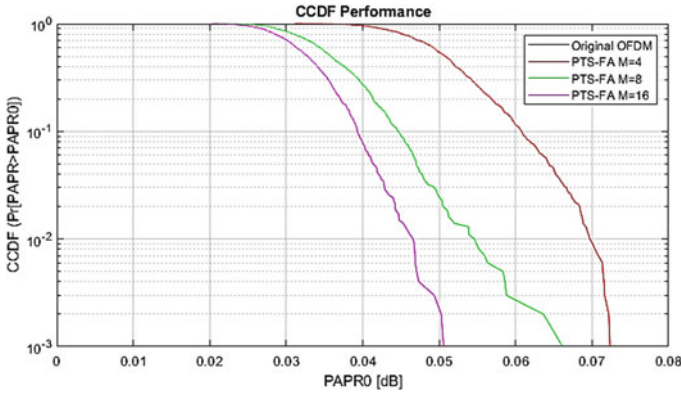


Fig. 8 Magnified view of PAPR performance using PTS-FA for N = 128 and M = 4, 8, 16 subblocks

Table 5 Improvement of PAPR reduction in PTS-FA when N = 128

M = 4	M = 8	M = 16	OFDM	IMPROVEMENT		
				M = 4	M = 8	M = 16
0.0738	0.0679	0.0515	10.7938	10.72	10.7259	10.7423

of subblocks of M = 4, 8 and 16. Meanwhile, Table 5 displays the improvement of PAPR minimization with various subblock values. This simulation is constructed to examine and evaluate the performance of PAPR in different subblock values. As can be seen in the table, subblock value of M = 16 made the most improvement in reduction of PAPR by 10.74 dB. Whereas M = 8 and M = 4 made an enhancement in decreasing the PAPR values by 10.73 dB and 10.72 dB respectively. Based on the analysis, the larger value of subblocks, the finer the PAPR reduction. It is because the higher number of subblocks, the effectiveness of PTS-FA in transmitting data also increases. However, it takes longer processing time for higher subblock value since it involves the complexity.

Figure 9. indicates the PAPR reduction when iteration values are varying with input data of N = 64 and M = 4. From the graph, iteration of 200 attained smallest PAPR by 0.611 dB, whereas K = 100 and K = 50 established PAPR values by 0.73 dB and 1.47 respectively. When the number of iteration value increases, PTS-FA accomplished preferable PAPR performance and lessen the computational complexity as well, unfortunately, the simulation time increases exponentially.

Fig. 9 PAPR reading in CCDF of PTS-FA using various iterations

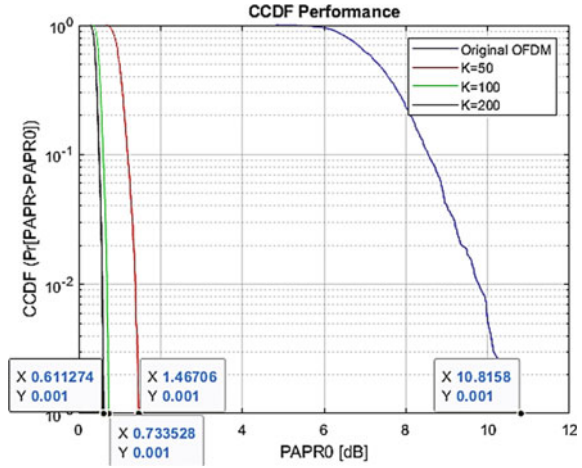
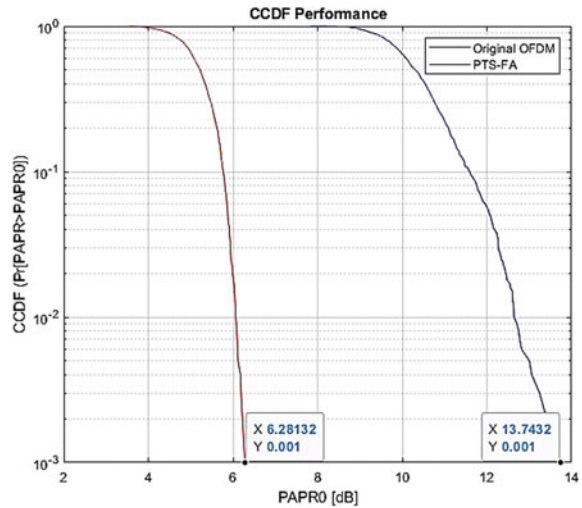


Fig. 10 PAPR performance of PTS-FA when N = 64 and M = 4



4.2 Simulation Results in QPSK Modulation

The purpose of conducting the simulation in QPSK modulation is so observe the PTS-FA performance and compare the results with the previous modulation scheme. In this section, the PAPR reduction will be examined by inserting various number of subcarriers and subblocks simultaneously.

Figures 10, 11 and 12 show the CCDF graph over 10 – 3 of PAPR performance when N = 64, M = 4, N = 128, M = 8 and N = 256, M = 16 respectively. Meanwhile, table 6 displays the improvement of PAPR reduction in PTS-FA. When the subcarrier is N = 64 and subblock is M = 4, the PAPR decreases by 7.46 dB. Whereas, for

Fig. 11 PAPR performance of PTS-FA when $N = 128$ and $M = 8$

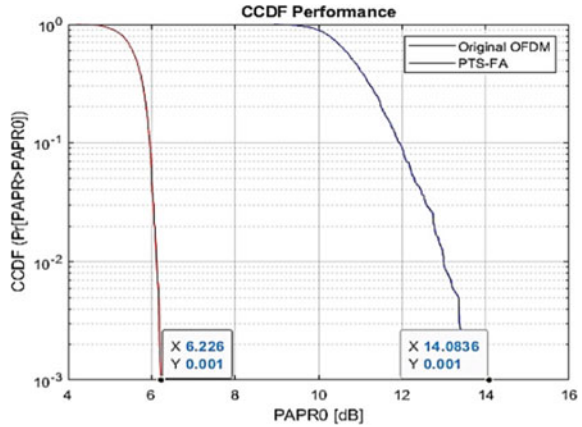


Fig. 12 PAPR performance of PTS-FA when $N = 256$ and $M = 16$

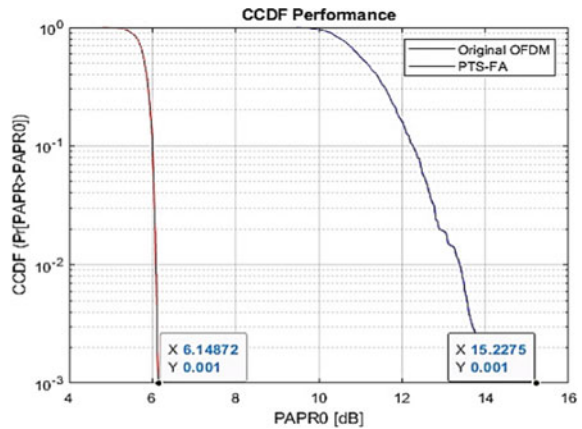
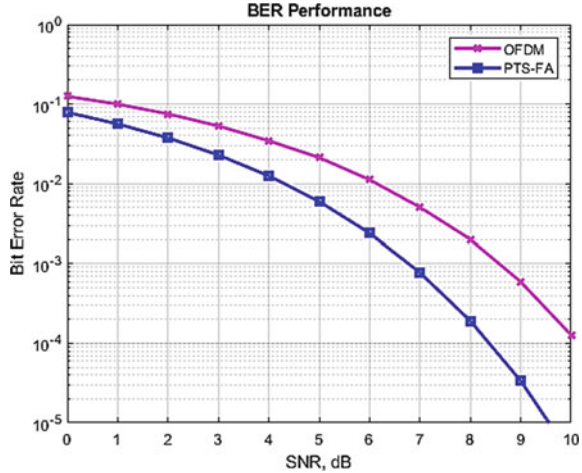


Table 6 Improvement of PAPR reduction in PTS-FA (dB)

	$N = 64$ $M = 4$	$N = 128$ $M = 8$	$N = 256$ $M = 16$
	6.28132	6.226	6.14872
OFDM	13.7432	14.0836	15.2275
IMPROVEMENT	7.46188	7.8576	9.07878

$N = 128$ and $M = 8$, the reduction of PAPR value is improved by 7.86 dB. Lastly, when $N = 256$ and $M = 16$, the better PAPR reduction is achieved by 9.08 dB. As can be seen in the figures, when the number of subcarriers and subblocks increase, the PAPR value dropped a lot. There are more data can be transmitted efficaciously if the number of subcarriers and subblocks are larger. However, it leads to longer processing time. From the observation, the implementation of PTS-FA technique is

Fig. 13 BER Performance for OFDM and PTS-FA



capable to minimize the higher number of PAPR and the complexity can be preserved as well.

4.3 Performance of Bit Error Rate (BER)

BER performance will be observed in QPSK modulation since QPSK has better and low possibility of bit error. The performance of BER will be analyzed and discussed in this section.

Figure 13 displays the BER performance for OFDM signal and PTS-FA in a QPSK modulation over SNR in dB. From the graph above, the proposed technique displays better performance and less BER due to its effectiveness in transmitting data. Whereas OFDM signal shows high BER since it is prone to noise during the transmission of data.

5 Conclusion

OFDM signals and its drawback has been clarified in this project. As previously mentioned, OFDM systems provides significant advantages in wireless communication system but it is often degraded due to the high PAPR. This project is focusing on the implementation of firefly algorithm in PTS to alleviate the larger PAPR and preserve the low complexity as well. As discussed earlier, the higher subblocks can reduce PAPR values after has been conducted in two different modulation schemes. It is because the firefly algorithm itself has less parameters to be adjusted and simple

structure for larger subblocks compared to the existing PTS. In addition, both modulation schemes did not show any differences on the result and did not give negative effect on the reduction of PAPR due to both have higher bandwidth efficacy and power capability to transmit data in communication system. From the simulation results, implementation FA in PTS performs better PAPR reduction by 7–10 dB compared to the existing PTS only reduces PAPR by 4–5 dB. Hence, it proves that PTS-FA technique promises an effective and better trade off amid PAPR reduction and computational complexity.

Acknowledgements This project was supported by “Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with project code FRGS/1/2020/TKO/USM/02/21.

References

1. Ayeswarya R, Amutha Prabha N (2019) Fractional wavelet transform based PAPR Reduction schemes in multicarrier modulation system. *IETE J. Res* 0(0):1–11 <https://doi.org/10.1080/0372063.2019.1621685>.
2. Hosseinzadeh Aghdam M, Sharifi A (2019) PAPR reduction in OFDM systems: An efficient PTS approach based on particle swarm optimization, *ICT Express* 5(3):178–181. <https://doi.org/10.1016/j.icte.2018.10.003>.
3. Thafasal Ijyas VP, Al-Rayif MI (2019) Low complexity joint PAPR reduction and demodulation technique for OFDM systems *IETE J. Res* 0(0):1–11 <https://doi.org/10.1080/03772063.2019.1674194>
4. Sharifi AA, Emami H (2020) PAPR reduction of asymmetrically clipped optical OFDM signals: Optimizing PTS technique using improved flower pollination algorithm. *Opt. Commun* 474(April):126057 <https://doi.org/10.1016/j.optcom.2020.126057>
5. Sandeep Kumar V (2020) Joint iterative filtering and companding parameter optimization for PAPR reduction of OFDM/OQAM signal. *AEU - Int. J. Electron. Commun* 124:153365 <https://doi.org/10.1016/j.aeue.2020.153365>.
6. Singh M, Patra SK (2018) On the PTS optimization using the firefly algorithm for PAPR reduction in OFDM systems. *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)* 35(5): 441–455 <https://doi.org/10.1080/02564602.2018.1505563>.
7. Sharif AA, Hosseinzadeh Aghdam M (2019) A novel hybrid genetic algorithm to reduce the peak-to-average power ratio of OFDM signals. *Comput. Electr. Eng* 80:106498 <https://doi.org/10.1016/j.compeleceng.2019.106498>.

A Comparison of Machine Learning and Deep Learning in Hyperspectral Image Classification



Fady Mohamed Sadek, Mahmud Iwan Solihin, Fahri Heltha, Lim Wei Hong, and M. Rizon

Abstract In recent years, hyperspectral remote sensing has become popular in various applications. This technology can capture hyperspectral images with a large terrestrial data. In this paper, the feasibility of applying various machine learning and deep learning techniques to perform classification on hyperspectral images are investigated and compared. Particularly, a total of three popular machine learning classifiers namely supports vector machine (SVM), K-nearest neighbors (KNN) and artificial neural networks (ANN) are used for hyperspectral image classification, followed by another two deep architectures in convolutional neural networks (CNN). Three benchmarking datasets of hyperspectral images are used to evaluate the classification performances of suggested machine learning and deep learning techniques, namely: Indian Pines (IP) dataset, Salinas dataset, and Pavia University (PU) dataset. Extensive simulation studies reveal the excellent performance of 3D CNN deep learning in solving larger datasets with better classification accuracy despite the longer training time is required. However, it is not really the case when the dataset is not large enough. This is because deep learning is data-hungry architecture. Furthermore, the 3D CNN deep learning models employed in this study have shown more advantageous as compared to other machine learning models for having simplified pre-processing stages such as feature extraction in solving the classification problems of hyperspectral images.

Keywords Hyperspectral image classification · Convolutional neural network (CNN) · Support vector machine (SVM) · k-nearest neighbors (k-NN) · Artificial neural networks (ANN) · Fully convolutional network (FCN)

F. M. Sadek · M. I. Solihin (✉) · F. Heltha · L. W. Hong · M. Rizon
Faculty of Engineering, UCSI University, 56000 Kuala Lumpur, Malaysia
e-mail: mahmudis@ucsiuniversity.edu.my

F. Heltha
USK, 23333 Banda Aceh, Indonesia

1 Introduction

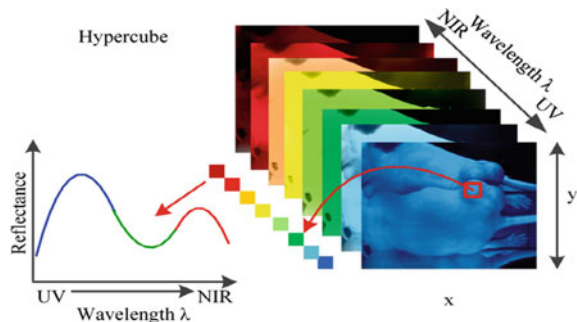
Hyperspectral imaging (HSI) is a spectral detection technique in which an object is observed using various well defined, broad-spectral optical bands. Originally deployed on remote sensation satellite and airborne platforms, HSI has been used in various applications over the past two decades such as for agricultural and water resources management, military protection, art conservation and archeological, medical analyzes of crime scene information, forensics documentation, food quality control, and mini-images [1]. HSI encompasses a wide variety of imaging technologies, such as medical hyperspectral imagery, atmospheric sounds, and hyperspectral imagery at close range. However, it has also been developed for applications in mining and geology, deemed a reliable source of data for a variety of possible applications, such as mineralogy, climate tracking, precision farming, protection and privacy requests, chemical analysis, astronomy and biological sciences, as well as for products quality characterizing in food industry [2].

In the past few decades, hyperspectral image data classification algorithms using classical machine learning such ANN and SVM have been discussed. On the other hand, deep learning has been emerged as a new approach of image classification due to its automated capability in features extraction.

2 Hyperspectral Image

The scene rays are capturing in a layer by layer image and various wavelengths, as shown in a hyperspectral data cube in Fig. 1. For each wavelength, the x–y plane shows the hyperspectral spatial data cube and the spectral contents of the z plane. Every hyperspectral band of images has a dimension that describes a digital number for every pixel that correlates to the radiance value gets by the sensor, where every band fits in a given wavelength. It is commonly depicted as the HSI data cube (3D hypercube) of $n_1 * n_2 * n_b$, where $n = n_1 * n_2$ represents the number of pixels and n_b represents the number of bands.

Fig. 1 Spectral signature in hyperspectral data cube



In a spectral area generated by the number of bands, each pixel is represented as a one-dimensional vector. The similar types of materials are then grouped by spectral, a mutually equivalent clustering algorithms. The famous clustering algorithms used for analysis of hyperspectral images are fuzzy c-clustering, K-means clustering, and unmixed spectral clustering approaches. Due to correlation is strong in a spectral space, the data is reflected in a lower dimensional space, which is less than spectral band count. Data reduction in dimensionality is achieved using other techniques [3] such as principal component analysis (PCA) [4], or independent component analysis (ICA) [5].

3 Convolutional Neural Networks

Hyperspectral imaging contains different image lines. Many of classification methods are based on 2-D CNN. Against the context, the performance of HSI classification is highly dependent upon both spatial and spectral details. The 3-D CNN has been used by very few approaches due to the added computing complexity as shown in Fig. 2. This paper suggests a CNN (HybridSN) dual spectral for classification with HSI. The HybridSN is a typically spectral–spatial 3-D CNN accompanied by 2-D CNN [6].

Considering spectral–spatial data cube of the hyperspectral is given as the following set:

$$I \in R^{M \times N \times D} \tag{1}$$

```

Network :
-----
Layer (type)           Output Shape           Param #
-----
Conv3d-1               [-1, 128, 1, 7, 7]    118,784
Conv3d-2               [-1, 128, 1, 7, 7]    13,312
LocalResponseNorm-3   [-1, 256, 7, 7]      0
Conv2d-4               [-1, 128, 7, 7]      32,896
LocalResponseNorm-5   [-1, 128, 7, 7]      0
Conv2d-6               [-1, 128, 7, 7]      16,512
Conv2d-7               [-1, 128, 7, 7]      16,512
Conv2d-8               [-1, 128, 7, 7]      16,512
Conv2d-9               [-1, 128, 7, 7]      16,512
Conv2d-10              [-1, 128, 7, 7]      16,512
Dropout-11             [-1, 128, 7, 7]      0
Conv2d-12              [-1, 128, 7, 7]      16,512
Dropout-13             [-1, 128, 7, 7]      0
Conv2d-14              [-1, 10, 7, 7]        1,290
-----
    
```

Fig. 2 Architecture of 3D CNN layer

where I is the initial signal, M is the width, N is the height, and D is the spectral bands/depth.

Every HSI pixel in I contains D measured spectral and forms a one-hot label vector expressed as:

$$Y = (y_1, y_2, \dots, y_C) \in R^{1 \times 1 \times C} \tag{2}$$

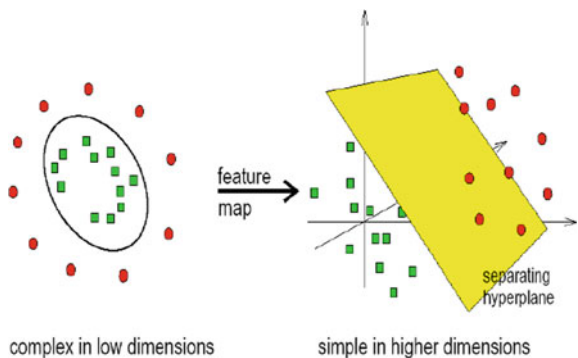
where c stands for types of ground cover [6].

4 Support Vector Machines

Suppose that two class data sets are in the R^2 region, as shown in Fig. 3. It is obviously shown that the data is not linearly separable in the space. Classifying this data into 2D and rendering it in smaller dimensions becomes a complex task, a nonlinear classifier is required. Therefore, the data is mapped to space R^3 . Initial space, in which the data is stored, is referred as the function space called the input space, and the higher dimensional space, in which the data is interpreted.

SVMs are designed to classify a binary class data. However, multiclass data has to be dealt with in the HSI classification. We need to expand SVM effectively to manage multiclass data. Two available methods are currently available: to construct several binary classifiers and combine the results, and to frame an entire data optimization problem. The latter is computationally complex and costly [7].

Fig. 3 Non-Linear data 2D is mapped to 3D



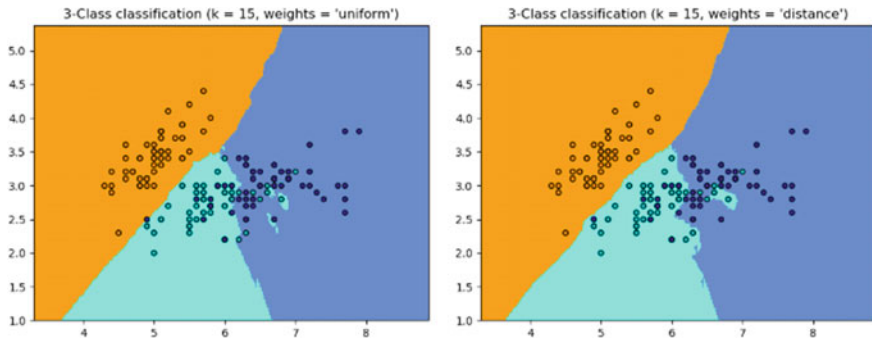


Fig. 4 Illustration of K-NN classification

5 K-Nearest Neighbors

The most widely used machine learning method in classification is K-neighbor grouping, where the best selected value depends heavily on the data. In general, it suppresses the effects of noise, but less distinguishes of the classification borders.

When the data is not distributed equally, the classification of radius-based neighbors may become a better alternative. The user specifies a fixed radius for using fewer closest neighbors to do classification in the sparser neighborhoods. This approach is less successful in high-dimensional parameter spaces due to the so-called ‘dimensionality curse’ [8].

The nearest basic neighbor classification uses uniform weights: i.e., a clear majority vote of the closest neighbors determines the value assigned to a question point. For certain cases, weighting of the neighbors is ideal, so that closer neighbors added to their fitness. This can be achieved by using the keyword weights. The default value, weights = ‘uniform,’ gives each neighbor uniform weights. Weights = ‘distance’ assigns weight equal to the reverse of the question point size. Alternatively, weight can be determined by means of a user-defined distance function [8].

6 Artificial Neural Networks

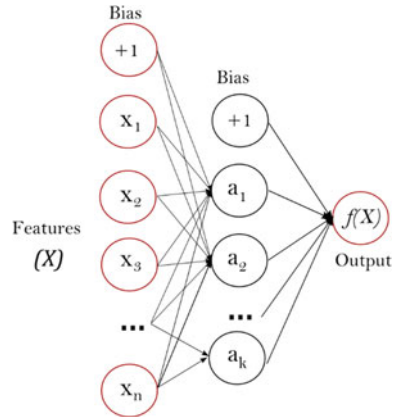
Multi-layer perceptron (MLP) correspond as feedforward of artificial neural network (ANN). MLP is a supervised learning algorithm that learns a function:

$$F(.) : R^m \rightarrow R^o$$

where m is the number of input dimensions and o the number of output dimensions c .

ANN has been used successfully in many applications of different field across discipline from environmental engineering [12, 13], signal processing [14], robotics

Fig. 5 One hidden layer MLP



[15] and particularly image classification [16]. Figure 5 shows the MLP hidden layer with several neurons.

7 Hyperspectral Image Data Collection

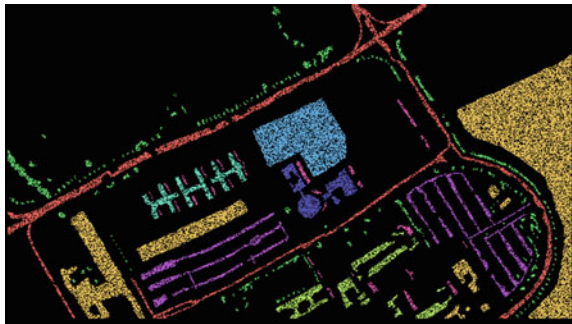
The datasets used in this study has been downloaded from online platform. They have been taken with different spectrometer types such as ROSIS-3 and AVARIS sensors. The study is based on three different types of datasets, Pavia University (PU), Indian Pines (IP) and Salinas. The three different datasets will be used to evaluate the effectiveness of image classifier using the classification methods of machine learning and deep learning. They have respective ground truth labels to classify.

7.1 Pavia University Dataset

The PU datasets consist of 9 ground truth classes and each class has respective sample number of average of 4753 samples shown in Table 1. The object in the same class shows great differences in spatial structure. Figure 6 show the groundtruth of the data sample of PU.

Table 1 Label of classes
(ground truth) for PU dataset

#	Class	Samples
1	Asphalt	6631
2	Meadows	18,649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947
Total Samples	–	42,776

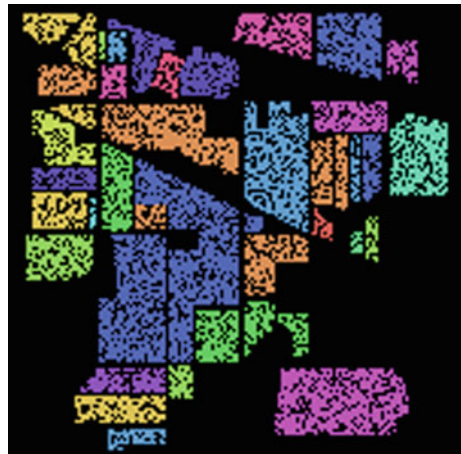
Fig. 6 Ground truth of PU

7.2 Indian Pines Dataset

The IP dataset consists of 16 ground truth unbalanced class and has an average number of samples in each class of 640 samples as shown in Table 2. The average and minimum groups contain 2455 pixels and just 20 pixels, respectively. Figure 7 show the ground truth of the data sample of IP.

Table 2 Label of classes (ground truth) for IP dataset

#	Class	Samples
1	Alfalfa	46
2	Corn no till	1428
3	Corn mantilla	830
4	Corn	237
5	Grass pasture	483
6	Grass Rees	730
7	Grass pasture mowed	28
8	Hay windrowed	478
9	Oats	20
10	Soybean no till	972
11	Soybean mint ill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings Grass Trees Drives	386
16	Stone-Steel-Towers	93
Total samples	–	10,249

Fig. 7 Ground truth for IP

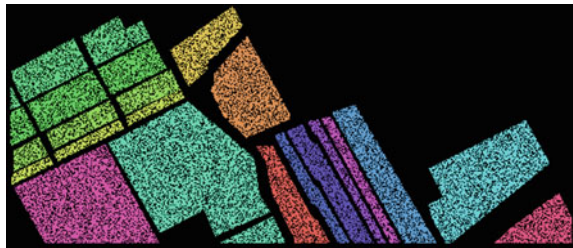
7.3 Salinas Dataset

The data set of Salinas has characteristics of two adjacent spatial classes and a bit of difference in spectral. These are a challenge for the methods of HSI classification. The datasets consist of 16 different classes with total of 54,129 samples as shown in Table 3. Figure 8 shows the ground truth of the data sample of Salinas.

Table 3 Label of classes (ground truth) for Salinas dataset

#	Class	Samples
1	Broccoli green weeds_1	2009
2	Broccoli green weeds_2	3726
3	Fallow	1976
4	Fallow rough plow	1394
5	Fallow smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes untrained	11,271
9	Soil vineyard develop	6203
10	Corn senesced green weeds	3278
11	Lettuce romaine 4wk	1068
12	Lettuce romaine 5wk	1927
13	Lettuce romaine 6wk	916
14	Lettuce romaine 7wk	1070
15	Vineyard untrained	7268
16	Vineyard vertical trellis	1807
Total samples	–	54,129

Fig. 8 Ground truth for Salinas



8 Principal Component Analysis

In this study, the principal component analysis (PCA) is applied on several interrelated variables to minimize the dimensionality of spectral dataset. We represent the PCA reduced data cube by $X \in R^{M \times N \times B}$, where X is the modified input after PCA, M is the width, N is the height, and B is the number of spectral bands after PCA [10].

The steps involved in the PCA is to calculate the mean of all the dimension of dataset, except the labels, then scale the data as shown in Eq. 3.

$$Z = \frac{X - \mu}{\sigma} \tag{3}$$

where Z is the scaled value, x is the initial, and μ and σ are mean and standard deviation, respectively.

To compute the covariance of two variable the Eq. 4 has to be used to find the covariance matrix.

$$Cov(X, Y) = \frac{1}{n - 1} \sum_{i=1}^n (X_i - x)(Y_i - y) \tag{4}$$

Moving on by computing the eigenvectors and corresponding eigenvalue where the eigenvalue is the factor by which eigenvector is scaled. Last, chose the number of dimensions corresponding new datasets from the eigenvector with the largest eigenvalue [11]. The PCA algorithm was applied on the bands of the IP datasets, to show the difference of the bands before and after applying the PCA. The dimension of IP datasets was reduced to 8 dimension as in Fig. 9 and Fig. 10 shows the difference.

After visualizing the band some of the data in IP is missing which will lead to low accuracy, because hyperspectral data are covered by sensor with high spectral resolution which cannot be well described by the second order characteristics. Hence, the PCA is not effective tool for HSI classification since it deals only with second-order statistics [11].

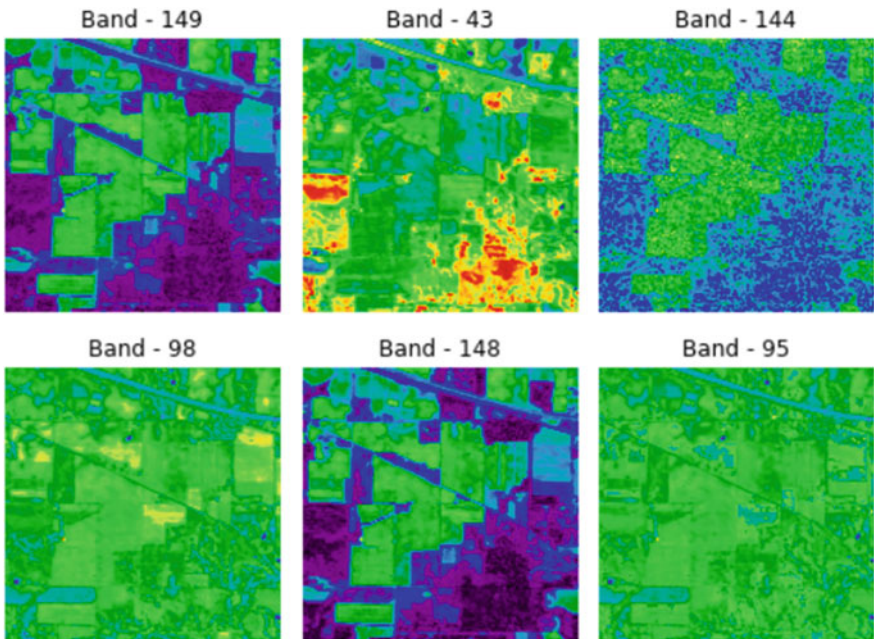


Fig. 9 Visualization of sample bands of IP

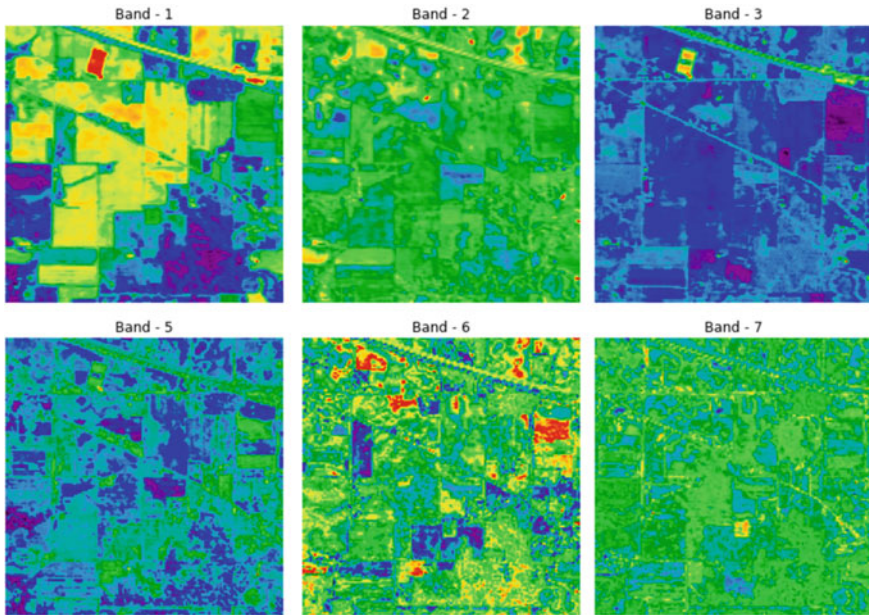


Fig. 10 Visualization of sample bands of IP after applying PCA

9 Classification Results Using Machine Learning

In this study we use three types of machine learning model, SVM, K-NN and ANN, and the three of them give different results in terms of accuracy. The three models have been tested twice where in the beginning, the training data was set to 30% and lastly was set to 90% to explain the difference of the accuracy when we increase the data training samples. Thus, the accuracy was evaluated based on respective 70% and 10% dataset. The data tested was IP, PU and Salinas.

Table 4 shows the performance testing accuracy among the proposed models with respective training data. The models trained with sample of 90% has shown better

Table 4 Classification accuracy for machine learning models

		Accuracy in testing		
Model	Training samples (%)	PU (%)	IP (%)	Salinas (%)
SVM	30	83.90	66.718	87.329
	90	90.603	92.858	90.098
K-NN	30	78.727	72.864	86.379
	90	79.149	77.659	86.200
ANN	30	88.08	70.60	85.326
	90	92.496	92.585	92.444

Table 5 Classification accuracy for deep learning models

		Accuracy in testing		
Model	Training samples (%)	PU (%)	IP (%)	Salinas (%)
3D CNN	30	83.90	66.718	87.329
	90	94.60	92.585	92.296
3D FCN	30	83.629	44.293	88.116
	90	95.185%	64.585	89.835

prediction comparing to 30%, especially for Indian Pines datasets due to it has 16 classes and lack of samples number those generate a very high prediction error when trained using 30% samples. By comparing IP and Salinas, we will notice that the performance of Salinas is higher, even they have the same number of classes. This difference performance is due to Salinas has 54,129 samples and IP has 10,229. Pavia university has good performance in both SVM and ANN model, except for K-NN, even we increased the number of samples it has almost gave the same classification result. It may be due to complexity of the data.

10 Classification Results Using Deep Learning

The 3-D CNN deep learning model has been used throughout this study since the project is dealing with image processing datasets. This model has been trained first by using 30% of the datasets and then increased it to 90% of the datasets.

Table 5 shows the performance testing accuracy for the proposed deep learning models with respective training data. PU and Salinas datasets have good performance comparing to Indian Pines. Using 30% of training sample in the Indian pines datasets, the prediction result shown very low comparing to 90% training samples. This is understandable as normally deep learning requires large dataset, i.e. data-hungry model.

11 Overall Results

The comparisons of the prediction results of both machine learning and deep learning for all discussed models are outlined in here. Table 6 shows the models' performance used in this study together with the training time taken for each executed model. Both machine learning and deep learning are good practical ways in solving image processing system problems. However, each datasets have different performance results in each model. In PU dataset, the model has shown the best performance result using 3D FCN with performance accuracy of 95%, while in other datasets it has not achieved the similar high-performance results. This is perhaps due to a

Table 6 Model performances and computation time

		Accuracy in testing					
Model	Training samples (%)	PU		IP		Salinas	
		Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)
SVM	30	83.90	25	66.718	31	87.329	36
	90	90.603	29	92.858	35	90.098	48
K-NN	30	78.727	18	72.864	24	86.379	27
	90	79.149	21	77.659	28	86.200	32
ANN	30	88.08	240	70.60	283	85.326	311
	90	92.496	308	92.585	347	90.444	384
3D CNN	30	83.90	901	66.718	974	87.329	1054
	90	94.60	1147	92.585	1454	92.296	1753
3D FCN	30	83.629	1098	44.293	1274	88.116	1346
	90	95.185	1941	64.585	2107	89.835	2584

smaller number of classes for PU dataset. In addition, it has taken a longer time of 1941s for training process.

For IP dataset, deep learning has not shown a good result as machine learning. The best performance in IP datasets found for SVM model and has taken 35 s in the training process.

Next, Salinas is a large dataset compared to others, and it has 92% accuracy of performance using 3-D CNN model. It is the highest one, but it has took 1753s time for training process.

We conclude that the 3-D CNN deep learning method generally has higher accuracy comparing to machine learning because of the several convolutional layers that consist in both architecture 3-D CNN and 3-D FCN. Putting the scope on deep learning we will notice that this network has many non-linear components such as local response normalization (LRN), max pooling, linear rectified unit (ReLU) making this model a practical way for pixel classification such for HSI classification as it learns 5×5 pixel centered in individual pixel vector and applied for the entire test picture. That explains why 3-D CNN and 3-D FCN have longer training time comparing to the other models in machine learning.

12 Conclusions

This study aims to compare between machine learning and deep learning models applied for hyperspectral image classification datasets. The datasets used in this study are India Pines, Pavia University and Salinas. Several procedures have been

taken starting with simple preparation of datasets, check their input types per channel and identify the kernel size for deep learning in each dataset.

It is found in the study that applying PCA is not a good option, since it reduces the performance of the prediction, therefore we use the original datasets and apply to the machine learning models namely SVM, K-NN and ANN. This process has been taken twice where we set the training sample to 30% and then we increase it to 90% to check the performance difference for the smaller and bigger training data. The same was applied to the deep learning models with 3D CNN and 3D FCN architecture.

Let's focus on the results of using 90% ratio in training datasets. In general, the classification accuracy of hyperspectral image classification using deep learning (3D CNN and 3D FCN) is better than that of machine learnings (SVM, KNN, ANN) for larger datasets only (Salinas and PU datasets). For smaller dataset, i.e. IP, machine learning accuracy is better than deep learning despite the training time is longer. This confirms the fact that deep learning architecture is data-hungry model as there are huge number of parameters to adapt.

References

1. Thenkabail PS, Gumma MK, Teluguntla P, Irshad A (2014) PE & RS 80(8)
2. ElMasry G, Kamruzzaman M, Sun DW, Allen P (2012) Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Crit Rev Food Sci Nutr* 52(11):999–1023
3. Huang G, Shi H, He Y, Duan, Luo F (2019) Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning. *IEEE Trans Cybern XX(X):1–13*
4. Du Q, Fowler JE (2008) Low-complexity principal component analysis for hyperspectral image compression. *Int J High Perform Comput Appl* 22(4):438–448
5. Wang J, Chang CI (2006) Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans Geosci Remote Sens* 44(6):1586–1600
6. S. K. Roy, S. Member, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN : Exploring 3-D – 2-D CNN Feature Hierarchy for Hyperspectral Image Classification," vol. 17, no. 2, pp. 277–281 (2020).
7. C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," pp. 1–39, 2019.
8. Nowicki RK (2019) Rough Nearest Neighbour Classifier. *Stud. Comput. Intell.* 802:133–159
9. Lokman, G., Çelik, H., & Topuz, V. Hyperspectral Image Classification Based on Multilayer Perceptron Trained with Eigenvalue Decay. *Canadian Journal of Remote Sensing*, 1–19 (2020).
10. Roy, S. K., Member, S., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. HybridSN : Exploring 3-D – 2-D CNN Feature Hierarchy for Hyperspectral Image Classification. 17(2), 277–281 (2020).
11. Wall, M. E., Rechtsteiner, A., & Rocha, L. M. Singular Value Decomposition and Principal Component Analysis BT - A Practical Approach to Microarray Data Analysis (D. P. Berrar, W. Dubitzky, & M. Granzow (eds.); pp. 91–109, Springer US (2003).
12. Hayder G, Solihin MI, Mustafa HM (2020) Modelling of River Flow Using Particle Swarm Optimized Cascade-Forward Neural Networks: A Case Study of Kelantan River in Malaysia. *Appl Sci* 10(23):8670
13. Hayder G, Solihin MI, Bin Kushiar KF (2021) A performance comparison of various artificial intelligence approaches for estimation of sediment of river systems. *J. Ecol. Eng* 22(7): 20–27

14. Astuti W, Tan S, Solihin MI, Vincent RS, Michael B (2021) Automatic voice-based recognition for automotive headlights beam control. *Int J Automot Mech Eng* 18(1):8454–8463–8454–8463
15. William L, Winda A, Satrio D, Sofyan T, Solihin MI (2019) Automotive start-stop engine based on face recognition system. In: in *E3S Web of Conferences*
16. Ang ZH, Ang CK, Lim WH, Yu LJ, Solihin MI (2020) Development of an artificial intelligent approach in adapting the characteristic of polynomial trajectory planning for robot manipulator. *Int. J. Mech. Eng. Robot. Res.* 9(3):408–414

Identify Target Area of Panel for Spraying Using Convolutional Neural Network



Y. K. Chin, S. S. N. Alhady, A. A. A. Wahab, W. A. F. W. Othman, E. A. Bakar, and M. N. Akhtar

Abstract Implementation of Artificial Intelligence in the paint spray process in the automobile industry can increase the efficiency of the paint spray process and reduce waste disposal. This project presents a semantic segmentation with a trained Convolutional Neural Network (CNN) implemented in the paint spray process as it can be used to predict and identify the target area that needs to be sprayed. A dataset contains different types of cars annotated with the car parts. A series CNN trained to classify the types of cars. There are 16 different semantic CNNs with different architecture and data stores trained to compare the result. The last step is to develop a system to identify the spray area. The result showed the accuracy of series CNN is 1, and the best way to train semantic CNN, which trained according to each type of car with the architecture of ResNet-50 with a data store consists of a more class. The validation result of ResNet-50_Type_01, ResNet-50_Type_02 and ResNet-50_Type_03 are 93.8162%, 90.5214% and 91.8023% which all have exceeded 85% and the Mean IoU of ResNet-50_Type_01, ResNet-50_Type_02 and ResNet-50_Type_03 are 0.8456, 0.8392 and 0.8263.

Keywords Semantic segmentation · Car parse · Paint spray

1 Introduction

Paint is used to protect the surface and also increase the surface durability [1]. The paint spray process is one of the essential processes in the automobile industry [2]. BMW specialists are now applying AI algorithms to compare live data from dust particle sensors in the paint booths and dryers with this database [3]. AI will be the revolution of many industries in the twenty-first century. AI can perform tasks

Y. K. Chin · S. S. N. Alhady (✉) · A. A. A. Wahab · W. A. F. W. Othman
School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia
e-mail: sahal@usm.my

E. A. Bakar · M. N. Akhtar
School of Aerospace Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia

including painting processes that require human intelligence to solve, such as helps to solve the complex scientific and engineering workflows through simulating, supplementing, or augmenting human intelligence, and AI will perform a more efficient and precise task [4–7, 9]. According to papers published by GreenPainters [8], paints are one of the significant sources of indoor air pollution, and paint is on the top-five list of environmental hazards according to US Environmental Protection Agency. Paints will release Volatile Organic Compounds (VOCs), which will affect human health. Disposal of the excess paint will cause environmental issues as paints consist of heavy metal, which will cause pollution [6]. Robotic painting provides an effective paint spray process, whereas the implementation of AI in automatic painting will increase the accuracy of identifying the target panel for spraying; implementing AI will help reduce the waste in the paint spraying process in the industry's production line.

Next, machine learning is the sub-field of AI, whereas AI makes machines capable of learning, and this can perform human-like tasks and analyzes more and deeper data using neural networks. Deep learning is a sub-field of machine learning. An artificial neural network is an adaptive system that learns by using interconnected nodes or neurons in a layered structure similar to a human brain. A convolutional neural network (CNN) is a neural network that is beneficial for computer vision applications. It takes images as the input; then, it assigns the importance of the images into weights. The architecture of CNN consists of different layers of an artificial neuron. CNN can train a network that can classify the image for computer vision applications [4, 10, 11].

Classification of car parts can increase the accuracy to identify the target area of the panel. Previous works [12–14] showed how to classify a car's car parts, which researchers identified the car parts of each car at the street view by CNN, which their own CNN developed. The objective of this research is to develop a system to identify the spraying area of a car. In this research, a system of image semantic segmentation with deep learning was developed. A dataset consisting of different views of a different set of cars be an input to the system. This system is used to identify the car parts. The predicted by the network and the testing subset will be compared and analyzed, and a CNN with the best performance will choose to identify the car parts.

2 Methodology

Matlab™ was used in the development of the project. The procedure has four parts: pre-processing datasets that the data store created to train the CNN in the next part. The second part is to prepare the CNN, and two types of CNN are trained. The third part is the measurement parameters which is the parameters to evaluate the result. The last part is the system design process.

2.1 Pre-processing of Datasets

The datasets obtained through [14] contains 20 types of cars and different perspective of view of the cars. Figure 1 shows an example of a photo in the datasets with unwanted elements, such as floor and person in the image. The photos in the datasets are needed to be cropped as the target is the car. The datasets also contained a list of bounding box coordinates in a file. The bounding box coordinates are obtained while the file is extracted into a cell array. Then, the coordinates are converted into a matrix as the photos are cropped with the bounding box coordinates. The photo was resized to the standard specific size of 376×250 . Figure 2 shows the example of the cropped photo in the dataset.

There are two types of data store needed to train the CNN: image data store and pixel label data store. All the cropped photos in the datasets labelled according to their type are store in the image data store. In contrast, a pixel label datastore was created through the application Image Labeler. The Image Labeler is used to label the car parts with regions of interest (ROI) labels. Figure 3 shows the example of the



Fig. 1 A photo in datasets [12]



Fig. 2 The cropped photo in datasets



Fig. 3 A labelled photo

labelled photo. There are two groups of the label: non-spray and spray.-spray refers to the target area that is not to paint, such as the car lamp. Each part is labelled with a different colour. Then all the labelled images will be exported as ground Truth data and stored in a pixel label data store. The data store is split into training, validation, and testing subset with ratios 50:25:25.

2.2 Train Network

There are two types of CNNs trained in this part: series CNN and semantic CNNs. A series CNN is trained to classify the type of car, and this series network is trained with an image data store consisting of three types of car. Figure 4 defines the architecture of the series CNN. The solver used is 'sdgm', while the initial learning rate is set to 0.01.

Figure 5 shows the example of the architecture of semantic CNN used. Two types of layers were replaced: the input and pixel classification layers. The input size must be defined in the input layer, so all the image data stores should have the same size. It is also required to specify the image size and the number of classes that the class is the car's parts. The pixel classification layer is needed to be included in this network. It consists of the number of classes and also class weight. Class weight is the median of pixel count over the total number of pixels.

There are different ResNet-18 which are trained with six labels and also 16 labels. Three different CNN used, ResNet-18, MobileNet-v2 and ResNet-50, are the default CNNs for the DeepLab v3 + function used for semantic segmentation in the software to compare the performance. Some CNNs trained with augmented datastore according to the types of the car. The validation result of the training progress needs to exceed 85%. Table 1 shows the 16 CNNs who have trained in this part.

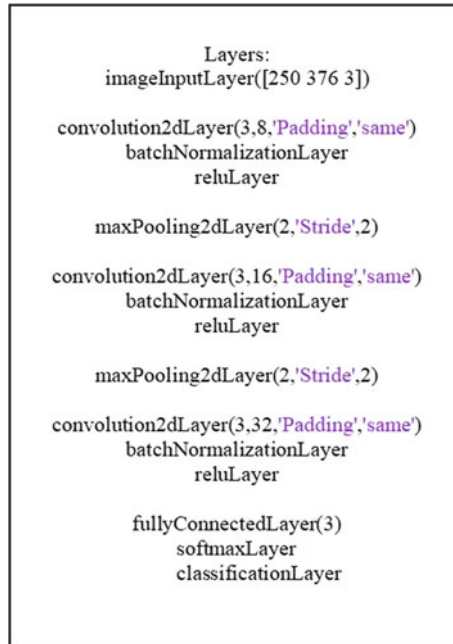


Fig. 4 The architecture of Series CNN

2.3 *Measurement Parameters*

After the network is trained, each network in different CNN trains will be analyzed with Intersection over union (IoU), mean accuracy, global accuracy and Mean BF Score. All of these are evaluated by a built-in function in the software. Next, some test subsets will be assessed with IoU according to the car parts to verify the performance of each car part. The higher the result of these evaluation metrics, the better the prediction of the segmentation of the spray area.

2.4 *System Design Process*

An input number between 1 to 333 is entered as the data store consists of 333 images, and the system will read the image according to the input. Then it will be classified according to its type. The are Type_01, Type_02 or Type_03, with a series network that has been trained to classify the type of car. A suitable network is determined according to the type of car based on ResNet-50. If the type of the car is not stated, then the system will display an error. Next, a variable *i* is entered, which is used to indicate the condition to paint the car. When *i* is equal to 1, then the image of the car will be painted and labelled according to the car parts. When *i* is equal to 2, then

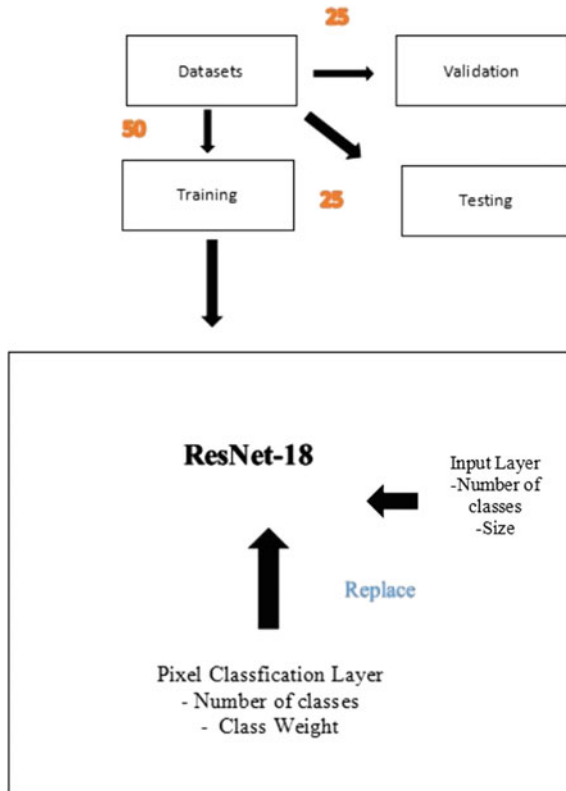


Fig. 5 The architecture of semantic CNN

the image of the car will be painted with a spray area in blue colour. Last, of all, the system will display the result.

3 Result and Discussion

In this research, three parts are to be evaluated. The first part is to assess the series CNN which has been trained to classify the car types. In the second part, the semantic CNNs trained to classify the car parts are evaluated. The last part is evaluating the system to paint the car body and identify the car parts.

Table 1 Different type of CNNs trained

CNN	Type of car	Augmented datastore	Labels
ResNet-18	All	No	16
ResNet-18	Type_01	No	16
ResNet-18	Type_02	No	16
ResNet-18	Type_03	No	16
ResNet-18	Type_01	No	6
ResNet-18	Type_02	No	6
ResNet-18	Type_03	No	6
ResNet-18	Type_01	Yes	16
ResNet-18	Type_02	Yes	16
ResNet-18	Type_03	Yes	16
MobileNet-v2	Type_01	No	16
MobileNet-v2	Type_02	No	16
MobileNet-v2	Type_03	No	16
ResNet-50	Type_01	No	16
ResNet-50	Type_02	No	16
ResNet-50	Type_03	No	16

3.1 Evaluation on Series CNN

The accuracy of the classification of the types of cars is equal to 1. All the labels of the validation subset are the same prediction with the series network, which has been trained to classify the types of cars.

3.2 Evaluation on Semantic CNN

Table 2 shows the validation result of the training progress of different CNNs which have been trained. All the validation results of different CNNs have exceeded 85%, which is required to obtain accurate results. Overall, the validation result of ResNet-18 and ResNet-50 with different types of car and a different number of labels have exceeded 90%.

Table 3 showed that the evaluation metric of Resnet-18, which have been trained. Resnet-18 of Type_01, ResNet-18 of Type_02 and ResNet-18 of Type_03 did not have significant differences, whereas the performance of ResNet-18 with all types was lower than the others. The result is because these three types of cars are different in model and also colour. In the training subset of ResNet-18 with all types, all images will have the same labels. For example, the colour of the body of the three types of cars is different. Different colours of the body of the three types of cars will cause the pixel classification to be confused during the training process. It will cause the

Table 2 Validation result of different CNNs which have been trained

CNN	Type of car	Augmented datastore	Labels	Validation result (%)
ResNet-18	All	No	16	85.0133
ResNet-18	Type_01	No	16	93.2880
ResNet-18	Type_02	No	16	91.3333
ResNet-18	Type_03	No	16	91.3285
ResNet-18	Type_01	No	6	98.0002
ResNet-18	Type_02	No	6	98.0914
ResNet-18	Type_03	No	6	96.5166
ResNet-18	Type_01	Yes	16	94.1393
ResNet-18	Type_02	Yes	16	91.2427
ResNet-18	Type_03	Yes	16	90.6002
MobileNet-v2	Type_01	No	16	85.9541
MobileNet-v2	Type_02	No	16	85.1640
MobileNet-v2	Type_03	No	16	85.8771
ResNet-50	Type_01	No	16	93.8162
ResNet-50	Type_02	No	16	90.5214
ResNet-50	Type_03	No	16	91.8023

Table 3 Evaluation metric for a network trained and the networks which are trained according to the type of the car

	ResNet-18 Type_01	ResNet-18 Type_02	ResNet-18 Type_03	Resnet-18 with all type
Global accuracy	0.9352	0.9430	0.9430	0.7512
Mean accuracy	0.9146	0.9315	0.9315	0.6981
Mean IoU	0.8052	0.8016	0.8016	0.4688
Weighted IoU	0.8863	0.8976	0.8976	0.6207
Mean BF score	0.8918	0.8395	0.8395	0.6190

accuracy to classify the body parts to be lower. Next, the shape of the car body for a different type of car also will be different. Different colours and conditions will cause the boundary of the labels to be inconsistent. There are only three types of cars in the data store of the research. There is less variety in types, whereas there are various car perspectives for every kind of car. It is better to train the network according to the types of cars.

Table 4 showed that the evaluation metric of the different networks which have been trained. The value of all the evaluation metrics of MobileNet-v2 is lower compared to the ResNet-18 and ResNet-50. It is because the number of parameters in ResNet is higher than in MobileNet-v2 and MobileNet-v2 is small, low-latency, low-power models parameterized to meet the resource constraints of a variety of use

Table 4 Evaluation metric of different CNN trained of Type_02

	ResNet-18	MobileNet-v2	ResNet-50
Global accuracy	0.9581	0.8903	0.9640
Mean accuracy	0.9532	0.8728	0.9519
Mean IoU	0.8211	0.6575	0.8392
Weighted IoU	0.9269	0.8143	0.9363
Mean BF score	0.9049	0.7745	0.9310

cases; thus, ResNet will have better performance. The number of the parameters in MobileNet-v2 is 3.5 million, ResNet-18 has 11.7 million, and ResNet-50 has 25.6 million parameters. In 30 epochs, ResNet will have higher accuracy compared to MobileNet-v2 as MobileNet-v2 needs to be trained in the high epoch. The performance of ResNet-50 is better than ResNet-18. The reason is that ResNet-50 has a higher number of parameters and also layers compared to ResNet-18. It causes ResNet-50 will have higher accuracy than ResNet-18.

Table 5 showed that the spray car parts have better IoU compared to non-spray car parts. The car body has slightly affected the lighting effect and occlusion effect compared to other car parts. The wheels have been affected by the occlusion effect obviously as the wheel part has the same colour and boundary as the shadow, so based on Fig. 6, the predicted result has labelled the shadow as the wheel and this cause the IoU is low. Next, the windshield and window are obviously affected by the lighting

Table 5 IoU of each part of different CNN trained

	IoU (ResNet-18)	IoU (ResNet-50)	IoU (Augmentated datastore with ResNet-18)
Body_Side	0	0	0
Body_Front	0	0	0
Body_Back	0.6030	0.6936	0.69317
Window_Right	NaN	0	0
Window_Left	0	0	0
Windshield_Back	0.6077	0.6321	0.5961
Windshield_Front	NaN	0	0
Windshield_SideView	0	0	NaN
Rear_Lamp_BackView	0.5999	0.4764	0.5350
Rear_Lamp_SideView	0	0	0
Front_Lamp_FrontView	NaN	0	NaN
Front_Lamp_SideView	0	NaN	0
Wheel_SideView	0	0	0
Wheel_FrontBackView	0.1686	0.1764	0.1472
Car_Plate_Front	0	0	NaN
Car_Plate_Back	0.6842	0.7167	0.5085



Fig. 6 Predicted result of Type_01 by ResNet-50

effect as they are made of glass, reflecting the spotlight of the surrounding area. The intensity of the light is also inconsistent, which causes the windshield and window to reflect light from a different perspective, which causes the IoU to be inconsistent. The IoU of the front lamp and the rear lamp of the car are high, and the IoU of the side view of the front lamp and rear lamp are lower. The shapes of the side view of the front lamp and rear lamp are inconsistent in different perspectives. Overall, the IoU of the car plate is high, so the predicted result correctly labelled this part.

Next, ResNet-18 is trained by augmented data store; overall, it has a better performance than the predicted result by ResNet-18 for the spray-area part but not the non-spray-area part. It has a better result for the car body parts due to the variety of data stores. The IoU for the car lamp and car plate is lower than ResNet-18 due to non-spray area car parts such as car lamps will disappear caused by the translation.

Table 6 shows the IoU over six car parts according to the ResNet-18, which have been trained. The IoU over six car parts are lower than the IoU over 16 car parts. The car body parts have lower IoU because the predicted result based on Fig. 6 has labelled the car plate as the car body part, whereas there is no car plate label in 6

Table 6 IoU of 6 classes of Type_01 by ResNet-18

	IoU (Resnet-18 with 6 classes)
Body	0.5808
Lower	0.0001
Windshield	0.5635
Window	0
Rear_Lamp	0.3836
Front_Lamp	0

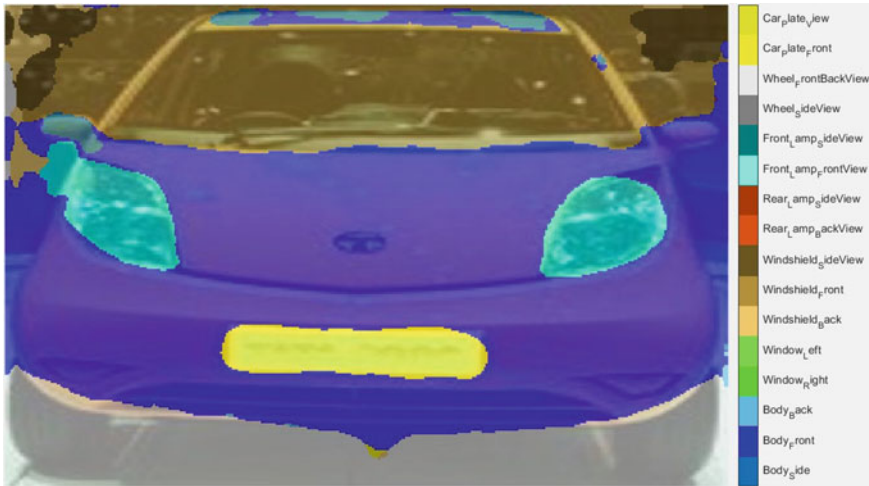


Fig. 7 Car classified with car parts

labels. The rear lamp also has lower IoU because for the six labels, since the shape of the rear lamp will be different in a different perspective, that inconsistent in shape will cause the lower accuracy and IoU.

3.3 Evaluation of System to Identify the Target Area of Car Parts

Figure 7 shows the semantic segmentation by car parts, the output when option 1 is the input. Each car part will be labelled with a different colour. Figure 8 shows the car in which the spray area of the car is painted with blue colour. This type of output will be displayed when option 2 is the input. All the body parts will be labelled with blue colour whereas the other car parts will not be labelled.



Fig. 8 Car painted with a spray area

4 Conclusion

In conclusion, the objective of this final year project has been achieved which a system To identify the spraying area of the car has developed. The car parts identify by semantic segmentation, which CNN's will be trained to classify the car parts. The accuracy of series CNN in this research is 1, which all the cars in the datasets can be labelled accurately. It will be better to train the semantic CNN according to the car types based on ResNet-50 with 16 labels. The validation result of ResNet-50_Type_01, ResNet-50_Type_02 and ResNet-50_Type_03 are 93.8162%, 90.5214% and 91.8023% which all have exceeded 85%. The MeanIoU of ResNet-50_Type_01, ResNet-50_Type_02 and ResNet-50_Type_03 are 0.8456, 0.8392 and 0.8263. Lastly, the system used to identify the spray area of the car has developed, and there are two options: one is used to label the car parts with different colours and the another is to paint the spray area of the car. Identifying the car parts with AI will help the painting process, but it will also help increase the accuracy of robotic painting and reduce pollution.

References

1. Vacher S, Hernandez C, Bärtschi C, Poussereau N (2010) Impact of paint and wallpaper on mould growth on plasterboards and aluminium. *Build Environ* 45(4):916–921

2. Rudzuan MN, Wan Khairunizam I, Zunaidi ZM, Razlan AB, Shahrman AR, Rozman A Shaharizal (2019) Development of automated spray-painting system for anti-static coating process. *IOP Conf Ser Mater Sci Eng* 557(1):012001
3. BMWBLOG homepage. <https://www.bmwblog.com/2020/05/20/bmw-ai-painting-process/>, Accessed 16 Sept 2021
4. Haenlein M, Kaplan A (2019) A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *Calif Manag Rev* 61(4):5–14
5. Alhady SSN, Kai XY (2018) Butterfly species recognition using artificial neural network. In *Intelligent manufacturing & mechatronics*. Springer, Singapore, pp. 449–457
6. Abdubrani R, Alhady SSN (2012) Performance improvement of contactless distance sensors using neural network. In *Proceedings of the 11th WSEAS international conference on Instrumentation, Measurement, Circuits and Systems, and Proceedings of the 12th WSEAS international conference on Robotics, Control and Manufacturing Technology, and Proceedings of the 12th WSEAS international conference on Multimedia Systems & Signal Processing*, pp 146–151
7. Architecture and Design homepage. <https://www.architectureanddesign.com.au/suppliers/greenpainters/paint-industry-impacts-environment-greenpainters>, Accessed 16 Sept 2021
8. Geffen CA, Rothenberg S (2000) Suppliers and environmental innovation: the automotive paint process. *Int J Oper Prod Manag* 20:166–186
9. SAS homepage. https://www.sas.com/en_my/insights/analytics/what-is-artificial-intelligence.html, Accessed 16 Sept 2021
10. Khan S, Rahmani H, Shah SAA, Bennamoun M (2018) A guide to convolutional neural networks for computer vision. *Synth Lect Comput Vision* 8(1):1–207
11. Lu W, Lian X, Yuille A (2014) Parsing semantic parts of cars using graphical models and segment appearance consistency. *arXiv preprint arXiv:1406.2375*
12. Geng Q, Zhang H, Huang X, Wang S, Lu F, Cheng X, Yang R (2018) Part-level car parsing and reconstruction from single street view. *arXiv preprint arXiv:1811.10837*
13. Ozuysal M, Lepetit V, Fua P (2009) Pose estimation for category specific multiview object localization. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE, pp 778–785
14. Lou Y, Bai Y, Liu J, Wang S, Duan L (2019) Veri-wild: a large dataset and a new method for vehicle re-identification in the wild. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 3235–3243

Navigation of Cart Follower Using Artificial Intelligence



C. Y. Ooi, S. S. N. Alhady, A. A. A. Wahab, W. A. F. W. Othman,
E. A. Bakar, and M. N. Akhtar

Abstract Wheelchair users encountered difficulties in carrying their items. A person who uses a wheelchair will get easy to move around with a wheelchair cart follower should be implemented to track and follow the wheelchair automatically while maintaining wireless connect 40 cm with the wheelchair to avoid any unnecessary collision. This project is developed using the camera and ultrasonic sensor inputs as the input to artificial intelligence (AI) to achieve such specifications. In this project, the target object, the wheelchair, is illustrated using a frisbee, while a robot car model illustrates AI. An obstacle avoidance algorithm and an object tracking algorithm using AI are developed. The obstacle avoidance algorithm detects the distance between the target object and the cart follower and controls the cart follower to move backwards if the distance is less than 40 cm. Object tracking algorithm uses deep neural network (DNN) transfer learning in object detection application with pre-trained Single Shot Detector (SSD.) The model was developed to track the target object's position and react accordingly once the target object moves. Based on this proposed system, the obstacle avoidance algorithm achieved 99.66% accuracy in distance ranging. In comparison, the object tracking algorithm provided a complete navigation system with 90% efficiency for the cart follower to track and follow the movement of the target object.

Keywords Wheelchair cart follower · Artificial intelligence · Frisbee · Obstacle avoidance · Object tracking · Transfer learning · Deep neural network

C. Y. Ooi · S. S. N. Alhady (✉) · A. A. A. Wahab · W. A. F. W. Othman
School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal,
Malaysia
e-mail: sahal@usm.my

E. A. Bakar · M. N. Akhtar
School of Aerospace Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia

1 Introduction

In the past years, autonomous robots and vehicles developed based on artificial intelligence [1]. For instance, the electric car company Tesla introduced ‘Tesla Autopilot’ that was created using a DNN [2] to perform semantic segmentation, object detection and monocular depth estimation [3]. The evolution of autonomous robot technologies [4] growing consistently from the first generation of the single-purpose machine to multi-tasking mobile robots today that can be used in industries, military, hospitals and residential. With the application of AI in robot following technologies, the autonomous mobile robots can achieve higher accuracy on object detection and image processing using the transfer learning method. In AI, various DNN algorithms used in object detection, such as Region-Based Convolutional Neural Networks (R-CNNs) [5], Single Shot Detector (SSD.) [6] and You-Only-Look-Once (YOLO) [7].

Recently, wheelchair users encountered difficulties in carrying their items during travelling. Some existing cart followers with colour tracking techniques [8] have some limitations, such as disconnected easily due to the changing of surrounding light intensity that causes the camera to fail to track the exact colour. It is crucial to maintain wireless connect 40 cm between the wheelchair and its cart follower to avoid an unnecessary collision. Existing cart followers using the obstacle avoidance technique on a mobile robot [9] do not equip with a complete navigation system to track the specific object to follow. Besides, a tracking system for FPGA cart followers using Artificial Neural Network (ANN) [10] requires a longer time to train the AI model and more complicated to be developed. Thus, this project designed to solve those limitations. Using transfer learning with a pre-trained AI model and dataset, the system provides a more approachable environment to a developer with higher accuracy in object detection.

This project aims to develop a distance ranging system for follower robots (to a wheelchair) using obstacle avoidance techniques. Implement a navigation system for follower robots using object tracking techniques in AI is proposed. Combining both algorithms allows the object follower robot to follow the movement of a target object from time to time without any collision.

Table 1 compares the significance and limitations between previous proposed systems. First system describes the wheelchair luggage follower using colour tracking technique by using Pixy CMUcam5 [8]. Second system implements the obstacle avoidance robot using ultrasonic sensor and Arduino [9]. Third system develops a tracking system using Artificial Neural Network for FPGA cart follower [10].

Table 1 Significance and limitations of the previously proposed system

Researchers	Significance	Limitations
M. F. Ahmad, H. J. Rong, S. S. N. Alhady, Wan Rahiman and WAFW Othman [8]	Use colour-code as identifier and area comparison for distance estimation	Sensitivity and stability of connection affected by the brightness of environment
Rajesh Mothe, S. Tharun Reddy, G. Sunil, and Chintoju Sidhardha [9]	Assessment on obstacle evading framework is capable of avoiding obstacles and altering the robot’s position	Does not equip with an object tracking function; hence may get interruption easily by another surrounding object
M. F. Ahmad, S. S. N. Alhady, C. C. Moi, A. A. A. Wahab, W. A. F. W. Othman, A. A. M. Zahir and E. A. Bakar [10]	Implementation of ANN produces high accuracy output with slow computation time performance	Require repetition of data collection and training of ANN model for a single object to be tracked

2 Methodology

2.1 Project Implementation Workflow

The overview of project implementation workflow is shown in Fig. 1 and starts with the loading of the model and interpreter, followed by the initialization of GPIO pins and parameters. Next, the program runs a series of looping functions in the system. There are five functions within the looping section: object avoidance algorithm, image capturing, prediction performing by the model, object tracking algorithm and control of motor movement. In the looping section, the system first checks the obstacle’s existence and carries our obstacle avoidance algorithm if obstacles are detected. The system will proceed to the other four functions if there is no obstacle detected. The existence of a looping section is significant as the project involves a real-time processing application. The system is required to track the input from time to time to give the respective output command immediately once the input varies.

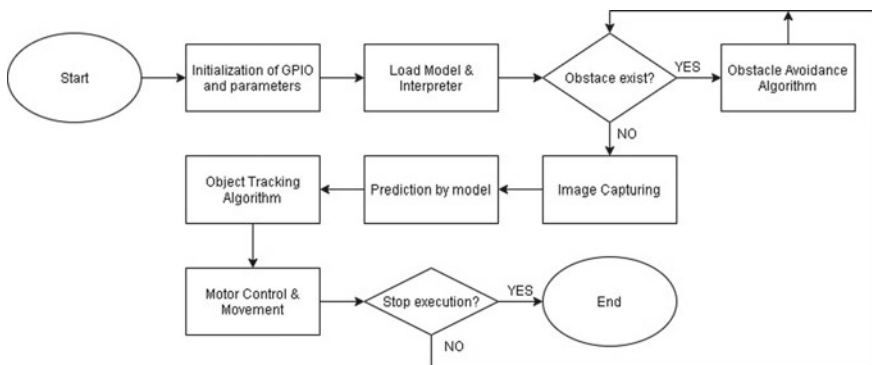


Fig. 1 Overview of project implementation workflow

Table 2 Commands for motor movement and respective direction of rotation in obstacle avoidance algorithm

Commands	The direction of rotation for left side motor	The direction of rotation for right side motor
Moving backward	Rotates backwards	Rotates backwards
Stop moving	Stop rotating	Stop rotating

2.2 Obstacle Avoidance Algorithm

The algorithm sends a signal for the ultrasonic sensor to transmit the ultrasonic sound wave. When an obstacle is detected, the wave is reflected and received by the receiver of the ultrasonic sensor. The system calculates the pulse duration between both transmitted and received waves and convert the pulse duration into the distance through multiplication. The calculation is shown in Eqs. 1 and 2. Condition checking is carried out to check whether the length exceeds 40 cm. If the distance between the wheelchair and the cart follower robot is less than 40 cm, the system sends the ‘move backward’ command to the motors for 0.2 s and follows the ‘stop moving’ command. Next, the system loops back the function and recheck the distance between the obstacle and the cart follower robot until the distance is greater or equal to 40 cm. Throughout this function, the cart follower robot can react accordingly to maintain wireless connections at least 40 cm with the object to avoid the collision. Table 2 shows the commands for motor movement and direction of motor rotates in the obstacle avoidance algorithm.

$$\text{Pulse duration, } \Delta t = \text{pulse end time} - \text{pulse start time} \quad (1)$$

$$\text{Distance, } D = (ct * \Delta t) / 2 \quad (2)$$

2.3 Object Tracking Algorithm

Object tracking algorithm starts with capturing the camera frame from the Raspberry Pi Camera Module. The image captured from the camera frame is converted into an array and to be provided to the interpreter. The system is then interacting with the SSD. Model and COCO datasets to obtain the object lists for prediction purposes. This process can be said as the limitation of the project as it takes up more than half of the processing time.

After the list obtaining process is completed, the system identifies the object’s class name, prediction probability’s score, and the coordinates of the objects in the camera frame. This algorithm compares the information from the model’s prediction with the image captured from the camera. If the target object is not detected, the function

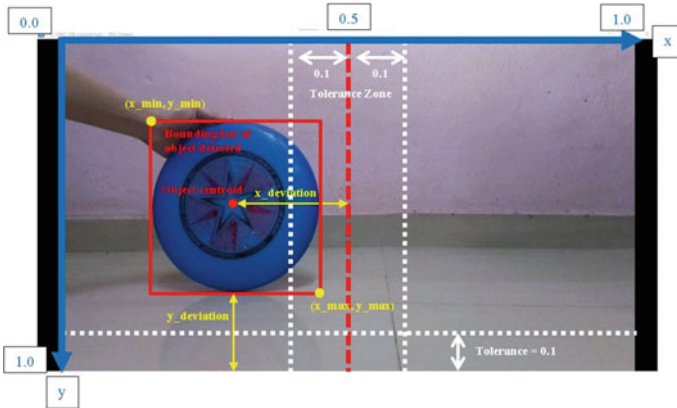


Fig. 2 Declaration of variables in camera frame preview

prints “Target object is not tracked” and returns without any further processing. Once the target object is detected, the calculation is performed to obtain the coordinate of the target object through the box labelling method. The measure is carried out to calculate the object’s x-axis deviation from the centre of the camera frame and the y-axis deviation from the bottom of the camera frame. Once completing execution, the system proceeds to the motor control function.

Figure 2 shows the declaration of variables in the camera frame preview. When the object tracking algorithm detected the target object specified, the system will predict the size of the target object to obtain its respective coordinates. The system pointed out two coordinates (x_min, y_min) and (x_max, y_max) representing the top left corner and bottom right corner of the target object, as shown in Fig. 2. Next, the system calculated object centroid (obj_x_center, obj_y_center) and the x and y deviation of the object using the formula listed from Eqs. 3 to 8 as shown. If the x and y deviation calculated was outside the tolerance zone, the system generated the respective command to control the motors’ movement.

$$x_length = x_max - x_min \tag{3}$$

$$y_length = y_max - y_min \tag{4}$$

$$obj_x_center = x_min + (x_length/2) \tag{5}$$

$$obj_y_center = y_min + (y_length/2) \tag{6}$$

$$x_deviation = 0.5 - obj_x_center \tag{7}$$

$$y_deviation = 1 - y_max \tag{8}$$

2.4 Control of Motor Movement

This function first checks the X and Y coordinates of the object detected and compares its x and y deviation with tolerance values. Upon checking, if the object’s horizontal and vertical deviation is within the tolerance zone, the system sends a ‘stop moving’ command to the motor; hence the robot will not move and display the “reached target object” message. If the object’s horizontal deviation is within the tolerance zone but vertical deviation outside the tolerance zone, the system sends the ‘move forward’ command to the follower robot and follows it by the ‘stop moving’ command after 0.3 s. Suppose the absolute value of the object’s horizontal deviation ($|x_deviation|$) exceeds the tolerance value. In that case, the motor driver will either receive the ‘moving left’ or ‘moving right’ command to rotate the motor accordingly. The duration of executing ‘moving left’ or ‘moving right’ commands can be divided into four intervals which are 0.25, 0.21, 0.18 and 0.15 s, where the decision for selecting the time duration based on the $|x_deviation|$ as shown in Table 3. The commands to control the motor movement and the directions of rotation are shown in Table 4.

Table 3 Durations for the motor to execute commands based on $|x_deviation|$

$ x_deviation $	Duration for executing the command (s)
≥ 0.40	0.25
≥ 0.35 and < 0.40	0.21
≥ 0.20 and < 0.35	0.18
< 0.20	0.15

Table 4 Commands for motor movement and direction of rotation in object tracking algorithm

Commands	The direction of rotation for left side motor	The direction of rotation for right side motor
Moving forward	Rotates forward	Rotates forward
Moving left	Rotates backward	Rotates forward
Moving right	Rotates forward	Rotates backward
Stop moving	Stop rotating	Stop rotating

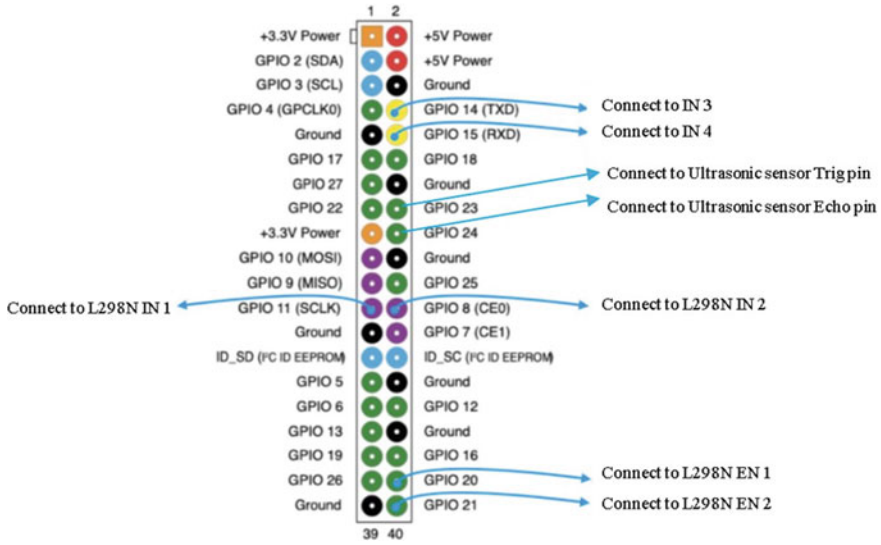


Fig. 3 Pin connection of the hardware to Raspberry Pi

2.5 Hardware Interface

The hardware components used in this project includes Raspberry Pi 3Model B+, Raspberry Pi 8MP camera module V2, HC-SR04P ultrasonic ranging module, L298N motor driver, two-wheels robot car chassis and TT DC dual-axis plastic gearbox motor. The hardware circuit connection is set up as shown in Fig. 3 before building the prototype. The camera module is connected to Raspberry Pi by inserting the cable of the camera module into the Raspberry Pi’s camera port. Figure 4 shows the circuit connection of the L298N Motor Driver to DC motors, and Fig. 5 shows the circuit connection ultrasonic sensor to Raspberry Pi GPIO pins.

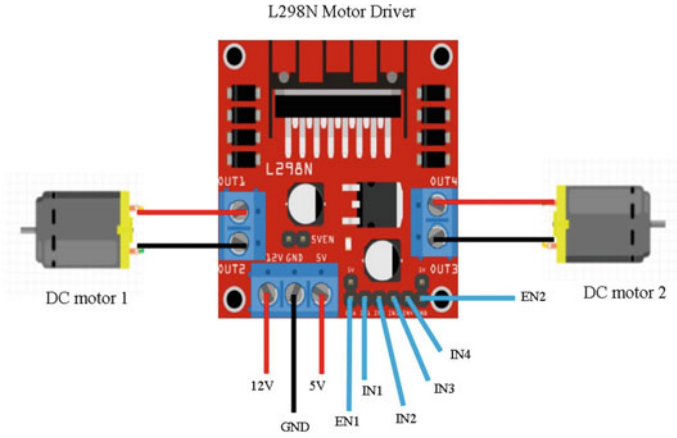


Fig. 4 Connection of L298N motor driver to raspberry Pi and DC motor

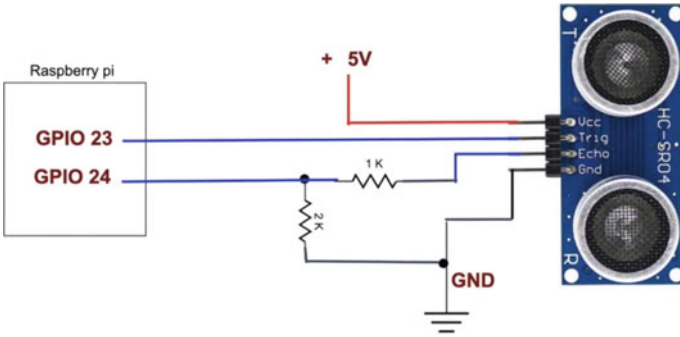


Fig. 5 Connection of ultrasonic sensor to raspberry Pi

2.6 Project's Prototype

The project prototype in different views such as top view, front view, and side view, were shown in Figs. 6(a), (b), and (c).

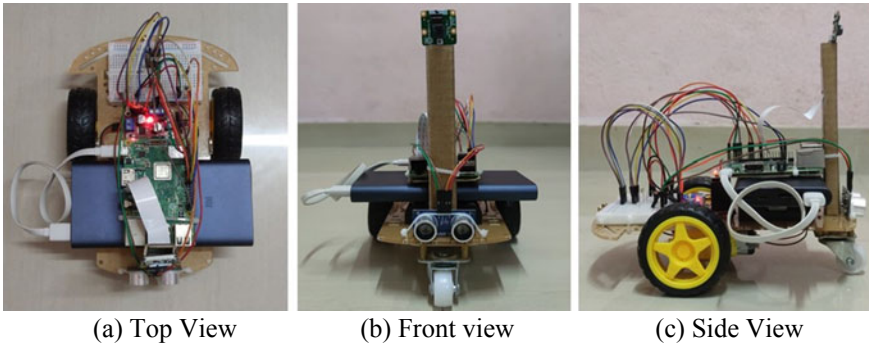


Fig. 6 Project’s prototype in different views

3 Results and Discussions

3.1 Obstacle Avoidance Algorithm

The results of the obstacle avoidance algorithm were compared and shown in Table 5.

Based on Table 5, the results taken from the system’s calculation and from exact measurement for obstacles placed at 10 cm in front of the object follower robot were compared. The system recorded the pulse start time when the pulse was sent to the ultrasonic sensor’s trigger pin, and the pulse ended when the pulse was received

Table 5 Results are taken from the system’s calculation, and exact measurement for obstacle placed 10 cm in front of object follower

Result from system’s calculations	Result from exact measurement
<pre> pulse_start = 1624938179.49955 pulse_end = 1624938179.41814 pulse_duration = 0.00062 obstacle distance = 39.83 cm, (TOO NEAR) Moving Backward !!! (Obstacle Avoidance Algorithm) : pulse_start = 1624938179.81735 pulse_end = 1624938179.81837 pulse_duration = 0.00102 obstacle distance = 37.48 cm, (TOO NEAR) Moving Backward !!! (Obstacle Avoidance Algorithm) : pulse_start = 1624938171.22481 pulse_end = 1624938171.22631 pulse_duration = 0.0015 obstacle distance = 25.71 cm, (TOO NEAR) Moving Backward !!! (Obstacle Avoidance Algorithm) : pulse_start = 1624938171.63191 pulse_end = 1624938171.63239 pulse_duration = 0.00049 obstacle distance = 34.99 cm, (TOO NEAR) Moving Backward !!! (Obstacle Avoidance Algorithm) : pulse_start = 1624938179.6894 pulse_end = 1624938179.69193 pulse_duration = 0.00254 obstacle distance = 43.53 cm, (SECURE RANGE) (Obstacle Avoidance Algorithm) : (Obstacle Tracking Algorithm) : Target object is not tracked. FPS = 0.9</pre>	

Distance of obstacle after object follower stopped moving = 43.53cm

Distance of obstacle after object follower stopped moving = 43.50cm

Table 6 Overview of results for obstacle avoidance algorithm and its accuracy

Distance of obstacle placed in front of object follower (cm)	10.00	20.00	30.00	40.00	50.00
Initial distance detected by the system (cm)	10.83	20.75	32.72	40.12	50.08
Distance calculated between obstacle to object follower after it stops moving, D_{calc} (cm)	43.53	42.51	42.32	40.12	50.08
Exact distance measured from obstacle to object follower after it stops moving, D_{meas} (cm)	43.50	43.00	42.30	40.00	50.00
Difference between distance calculated and exact distance measured, ΔD (cm)	0.03	0.49	0.02	0.12	0.08
Accuracy of distance ranging (%)	99.93	98.86	99.95	99.70	99.84

from the echo pin. Initially, the distance between obstacle and object follower robot calculated by the system was 10.83 cm in the first loop. The system kept running this algorithm in a loop if the distance of the obstacle detected was less than 40 cm. Meanwhile, the obstacle is ‘too near’ to the robot. Therefore, the robot received the ‘Move Backward’ command followed by the ‘Stop Moving’ command after 0.2 s and responded to the commands accordingly. At the fifth loop of the function, the distance of the obstacle in front of the object follower robot detected by the system was 43.53 cm, which indicated a ‘secure range’ to prevent a collision. Thus, the system stopped executing the obstacle avoidance algorithm and proceeded to the object tracking algorithm. Throughout exact measurement, the real distance between the obstacle and the follower robot was 43.50 cm. There was a trivial difference of 0.03 cm between the results calculated by the system and taken from exact measurements. Hence, the obstacle avoidance algorithm achieved high accuracy in distance ranging.

The same procedures were repeated to obtain the results for an obstacle placed at 20, 30, 40 and 50 cm in front of the object follower robot. The results obtained are shown in Table 6. When the obstacle placed initially was more or equal to 40 cm, the system displayed the obstacle was in a ‘secure range’. Thus, the object follower robot was not moving, and the system proceeded to object tracking algorithm. To review the accuracy of the obstacle avoidance algorithm, the difference between distance detected by the system and distance from exact measurement, ΔD , must be known. The calculation formulas for ΔD and accuracy percentage were shown in Eqs. 9 and 10. Table 6 shows the overview results for the obstacle avoidance algorithm and its accuracy in distance ranging.

$$\Delta D = |D_{calc} - D_{meas}| \quad (9)$$

$$Accuracy (\%) = [(|D_{meas} - \Delta D|) / D_{meas}] \times 100\% \quad (10)$$

$$Accuracy\ mean = \left(\sum Accuracy \right) / 5 \quad (11)$$






By referring to Table 6, the object follower robot maintained a wireless connection for at least 40 cm in all different conditions. This result proved that the obstacle avoidance algorithm was implemented successfully. The algorithm had accuracy exceeded 98% in the distance ranging in all conditions based on the calculation. Hence, it achieved high performance in avoiding collision from the obstacle. The mean of accuracy for all five taken results could be calculated using the formula in Eq. 11. The accuracy mean calculated for the obstacle avoidance algorithm was 99.66%.

3.2 Object Tracking Algorithm

Table 7 shows the camera preview and the system’s results in different conditions.

When the target object was not present in the image captured by the camera frame, as shown in Condition 1 in Table 7, the system displayed the message “Target object

Table 7 Results displayed by the system for object tracking algorithm in different conditions

Camera Preview	Messages displayed by the system
	<pre> (Obstacle Avoidance Algorithm) : pulse_start = 3824964687.43357 pulse_end = 3824964687.43942 pulse_duration = 0.00050 obstacle distance = 100.35 cm, (SECURE RANGE) (Object Tracking Algorithm) : Target object is not tracked. EPS = 3.7</pre>
	<pre> (Obstacle Avoidance Algorithm) : pulse_start = 3824966609.87505 pulse_end = 3824966609.87728 pulse_duration = 0.00052 obstacle distance = 45.13 cm, (SECURE RANGE) (Object Tracking Algorithm) : Target object is tracked. x_min= 0.468 , y_min= 0.182 , x_max= 0.948 , y_max= 0.792 object center (f 0.707 , 0.487) (x_deviation, y_deviation) = (-0.207 , 0.208) Start Moving !!!!! Moving Right !!! EPS = 3.3</pre>
	<pre> (Obstacle Avoidance Algorithm) : pulse_start = 3824966620.25886 pulse_end = 3824966620.26066 pulse_duration = 0.00072 obstacle distance = 42.42 cm, (SECURE RANGE) (Object Tracking Algorithm) : Target object is tracked. x_min= 0.583 , y_min= 0.183 , x_max= 0.583 , y_max= 0.787 object center (f 0.333 , 0.476) (x_deviation, y_deviation) = (0.187 , 0.213) Start Moving !!!!! Moving Left !!! EPS = 3.4</pre>
	<pre> (Obstacle Avoidance Algorithm) : pulse_start = 3824966609.88214 pulse_end = 3824966609.88217 pulse_duration = 0.00032 obstacle distance = 47.4 cm, (SECURE RANGE) (Object Tracking Algorithm) : Target object is tracked. x_min= 0.37 , y_min= 0.224 , x_max= 0.672 , y_max= 0.622 object center (f 0.525 , 0.423) (x_deviation, y_deviation) = (-0.021 , 0.189) Start Moving !!!!! Moving Forward !!! EPS = 3.2</pre>
	<pre> (Obstacle Avoidance Algorithm) : pulse_start = 3824966609.84283 pulse_end = 3824966609.84877 pulse_duration = 0.00218 obstacle distance = 48.15 cm, (SECURE RANGE) (Object Tracking Algorithm) : Target object is tracked. x_min= 0.253 , y_min= 0.358 , x_max= 0.728 , y_max= 0.971 object center (f 0.49 , 0.665) (x_deviation, y_deviation) = (0.01 , 0.029) Start Moving !!!!! Reached target object. EPS = 3.0</pre>

is not tracked” due to the object tracking algorithm. The object follower robot was not moving until the target object was tracked in the camera frame or an obstacle was too near to the robot. For Condition 2 and Condition 3 in Table 7, when the target object was detected and existed at the right side or left side of the image captured by the camera frame, the object tracking algorithm started to acquire the size and position of the object that predicted from model and dataset, then performed calculations on the values of x_{min} , y_{min} , x_{max} and y_{max} as shown in the results of Table 7. The system subsequently calculated the object centroid, horizontal deviation, and vertical deviation of the object to the camera frame. When both absolute values of x and y deviation exceeded the tolerance zone of 0.1, the system would react according to x deviation. Hence, the system sent the ‘Moving Right’ command to the robot and followed by the ‘Stop Moving’ command after 0.18 s for Condition 2 while sending the ‘Moving Left’ command to the robot and followed by ‘Stop Moving’ command 0.15 s for Condition 3. The system continued to run the program in a loop until the x deviation value was within the tolerance zone.

For Condition 4 in Table 7, the target object existed in the image captured by the camera frame but far away from the frame’s bottom. As a result, its x deviation is inside the tolerance zone, but its y deviation exceeded the tolerance value of 0.1, which was 0.369 as calculated. Therefore, the system sent the ‘Moving Forward’ command to the motor driver and the ‘Stop Moving’ command after 0.3 s so that the motor of the follower robot could react accordingly. The system proceeded to run the program’s main function in a new loop and took the same action if the target object tracked was outside the tolerance zone during the object tracking algorithm. For Condition 5, the target object existed in the image captured by the camera and was located at the x -axis centre and close to the bottom of the frame. Its x deviation and y deviation values obtained were 0.01 and 0.029, which was below the tolerance value. Therefore, the robot was not moving, and the system displayed a message showing it reached the target object. The system proceeded to run the program’s primary function in a new loop and took the same action during the object tracking algorithm if the target object tracked was the same as in Condition 5. The object tracking algorithm achieved 90% efficiency in the target object following the task based on the results obtained. There was a 10% tolerance zone for the object to be tracked in different positions during little movement.

4 Conclusions

In conclusion, both objectives of this project were achieved as the obstacle avoidance algorithm achieved high performance with 99.66% accuracy in distance ranging. In comparison, the object tracking algorithm provided a complete navigation system with 90% efficiency for the cart follower to track and follow the movement of the target object.

References

1. Tian Y, Pei K, Jana S, Ray B (2018) Deeptest: automated testing of deep-neural-network-driven autonomous cars. In: Proceedings of the 40th international conference on software engineering, pp 303–314
2. Kurdthongmee W (2020) A comparative study of the effectiveness of using popular DNN object detection algorithms for pith detection in cross-sectional images of parawood. *Heliyon* 6(2):e03480
3. Dikmen M, Burns CM (2016) Autonomous driving in the real world: Experiences with tesla autopilot and summon. In: Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications, pp 225–228
4. Matthews G, Hancock PA, Lin J, Pangniban AR, Reinerman-Jones LE, Szalma JL, Wohleber RW (2021) Evolution and revolution: personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Pers Individ Diff* 169:109969
5. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
6. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, Cham, pp 21–37
7. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
8. Ahmad MF, Rong HJ, Alhady SSN, Rahiman W, Othman WAFW (2017) Colour tracking technique by using pixy CMUcam5 for wheelchair luggage follower. In: 2017 7th IEEE international conference on control system, computing and engineering (ICCSCE). IEEE, pp 186–191
9. Mothe R, Reddy ST, Sunil G, Sidhardha C (2020) An IoT based obstacle avoidance robot using ultrasonic sensor and arduino. In: IOP conference series: materials science and engineering, vol 981, no 4. IOP Publishing, p 042002
10. Ahmad MF, Alhady SSN, Moi CC, Wahab AAA, Othman WAFW, Zahir AAM, Bakar EA (2021) Tracking system using artificial neural network for FPGA cart follower. *J Phys Conf Ser* 1874(1):012008

Object Tracking for Autonomous Vehicle Using YOLO V3



William Chin Wei Hung, Muhammad Aizzat Zakaria, M. I. Ishak, and P. M. Heerwan

Abstract Accuracy and performance of an object detection model have always been the main requirements for an object tracking system. In this project, the performance of machine learning based object detection using YOLO v3 technique will be investigated. Two models were provided where one model is trained using online Common Objects in Contact (COCO) dataset only, and the other model is trained with additional images from Universiti Malaysia Pahang (UMP) with several different locations dataset. The performance of the trained models were evaluated using mean Average Precision (mAP), and precision techniques. The model with highest precision was selected to be implemented on actual road test. The results show that the model 2 has the highest precision and was able to detect every class of objects. Each output box had displayed the class and the distance to the objects from the RGBD camera of the vehicle. It is observed that the first model that was trained to perform the mAP value of 90.2% and a performance of 0.484 precision. For the second model, it can be seen that the accuracy of the detections are higher than that of model 1. Therefore, model 2 has a better performance with a value of 0.596 precision.

Keywords Object tracking · Autonomous vehicle · mAP

1 Introduction

Object tracking system is one of the most important safety systems in an autonomous vehicles. Object detection is the first stage of this system, which is first to determine whether any instance of objects is in an image frame. The location of the instances found can be identified by drawing a bounding box around them which the process called object localization. Image classification involves in assigning a class label

W. C. W. Hung · M. A. Zakaria (✉)
IMAMS Laboratory, Faculty of Manufacturing and Mechatronics Engineering Technology,
Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: maizzat@ump.edu.my

W. C. W. Hung · M. A. Zakaria · M. I. Ishak · P. M. Heerwan
Autonomous Vehicle Laboratory, Centre for Automotive Engineering, Universiti Malaysia
Pahang, 26600 Pekan, Pahang, Malaysia

to the objects detected. After the object detection stage, the tracking operation is used to determine the positions of the detected objects in each frame [1]. To achieve the tracking operation, the concept of perspective transformation is applied in this project. It is a feature that transform the image in a straight manner after Perspective Transformation is applied to it [2].

In this project, RGBD camera is used to detect the existence of the objects and to determine the distance between the objects exist with the vehicle. The performance of the object tracking system using RGBD camera and YOLO v3 technique will be investigated [4]. The function of this system is mainly to detect the objects that exist in front of the vehicle and its perimeter like cars, human, animals, pumpers and other obstacles that come into the lane of the it [5]. The Autonomous Vehicle Laboratory team from Universiti Malaysia Pahang is also conducting research on the development of autonomous vehicles [6]. Figure 1 displays the autonomous prototype of the compact shuttle. The camera embedded on the autonomous shuttle will be used to collecting data for research like object tracking and lane classification for autonomous vehicle [7].

Modern architecture based on transfer learning, such as Faster RCNN (Region Based Convolutional Neural Networks), SSD (Single Shot MultiBox Detector) mobile network and YOLO (You Only Look Once) are the most common neural networks that can be used for this kind of object detection project. The procedure to carry out this project is basically distributed into four steps which are data acquisition, data pre-processing, model training and evaluation of model.



Fig. 1 Autonomous prototype of compact shuttle with camera and lidar sensor mounted on top

2 Methodology

The project was started by evaluating the object detection algorithms based on literature review. The algorithms were evaluated in terms of accuracy and detection speed. YOLO v3 was chosen as it has high detection speed compare the others such as Faster RCNN and SSD [8]. The project was then proceed to acquiring data and pre-processing data, which COCO dataset and UMP dataset were acquired and labelled for model training. After the training process, the performance of the models trained were evaluated. The flowchart of the project is shown in Fig. 2 below:

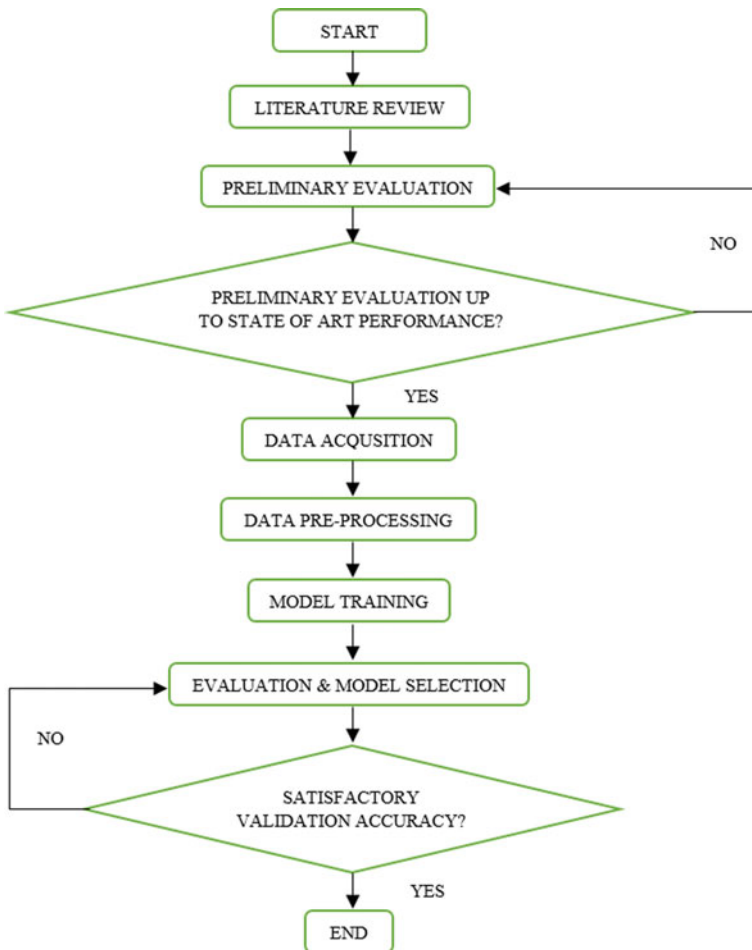


Fig. 2 Flowchart of project

2.1 Machine Learning Algorithm: YOLO (You Only Look Once) V3

The architecture of YOLO v3 is shown in Fig. 3 below:

YOLO v3 algorithm is a regression based algorithm, which anticipates bounding boxes in the image detected in one run of the algorithm. The image is split into number of cells, commonly 19×19 grid. The function of each cell is to predict number of bounding boxes. The probability of specific classes appeared in the cell is determined by YOLO v3 in the one pass of forwards propagation [3].

2.2 Data Acquisition and Data Pre-processing

COCO dataset contains 80 classes of objects such as car, person, motorcycle, bicycle etc. Each object displayed in the images are annotated and labelled by drawing a bounding box surrounding the object. The ratio on the data split for training, testing and validating is 70:20:10. The dataset is mixed with few set of UMP road images at Faculty of Manufacturing and Mechatronic Engineering Technology (FTKPM) and Residential College 5 (KK5) of UMP for second model training purpose. By doing this, the accuracy of the performance of the model at FTKPM and KK5 of UMP can be improved.

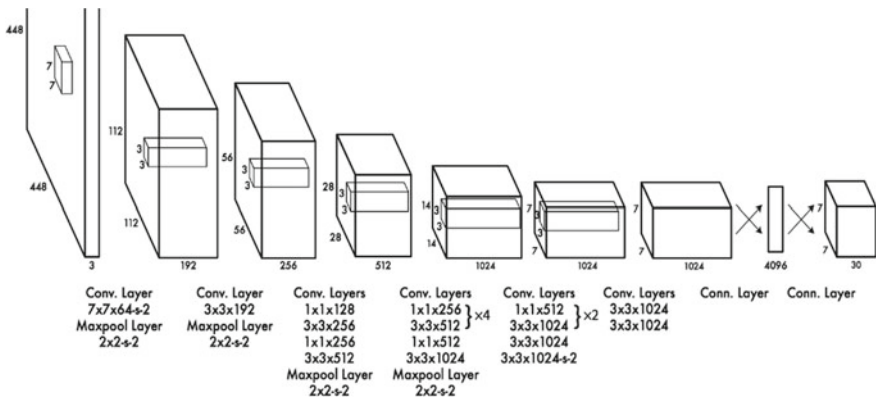


Fig. 3 YOLO v3 architecture

2.3 Model Training

In this project, two models were provided for object tracking training. Model 1 was trained using COCO data set only, while Model 2 was trained using COCO and UMP data set. The training process was set to 100 epoches which is approximately 40,000 steps. Each model will also have different training duration to determine the effect on the accuracy of the model, training time for model 1 is 24 h and training time for model 2 is 36 h. The performance of the trained models will be evaluated in term of the mean average precision (mAP) and precision graph using Tensorboard. The mAP is a measure that compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detections. The model with highest mAP will be implemented in the object tracking system.

3 Results and Discussion

The precision is calculated based on the formula of $(TP)/(TP + FP)$ where TP and FP represent the number of true positives and false positives respectively. Figures 4 and 5 shows the results for mean average precision for model 1 and model 2 trained with YOLO v3 algorithm. Based on the graphs, it can be observed that the model 2 had better average than model 1 as the mAP of the former is approximately 0.991 while the latter is 0.902. In terms of the precision of the models, Fig. 6 shows that model 2 had a higher value of 0.596 while model 1 had only 0.484 as shown in Fig. 7.

Table 1 mAP and precision of the trained models

Model	mAP	Precision
Model 1	0.902	0.484
Model 2	0.991	0.596

Fig. 4 Graph of mAP value against number of epoch of the trained model 1

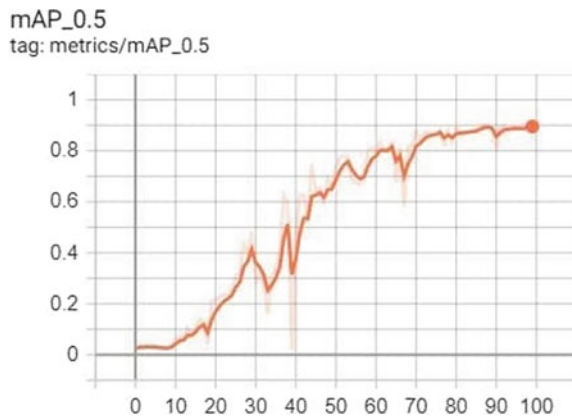


Fig. 5 Graph of mAP value against number of epoch of the trained model 2

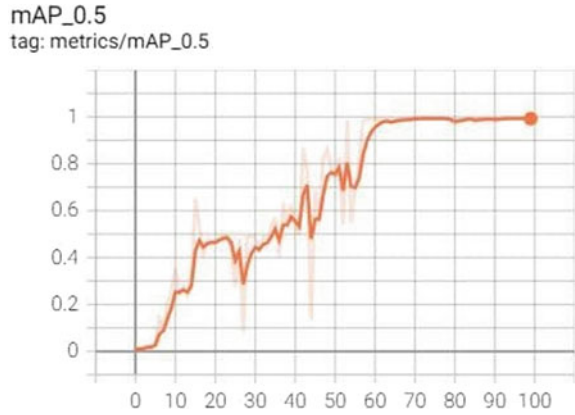


Fig. 6 Graph of precision against number of epoch of the trained model 1

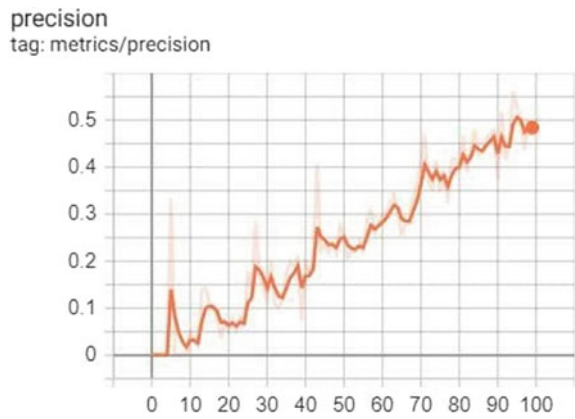


Fig. 7 Graph of precision against number of epoch of the trained model 2

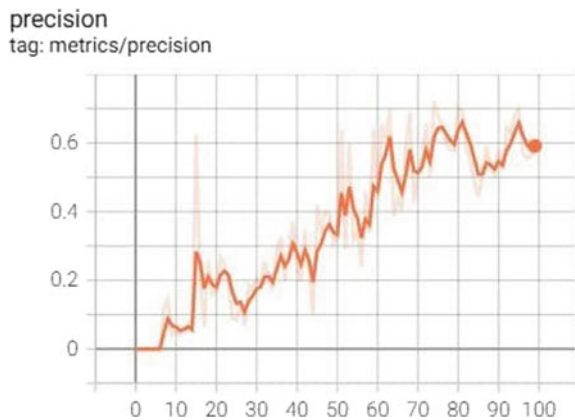




Fig. 8 Objects detected in image taken at FTKPM by model 2



Fig. 9 Person detection in image taken at road to KK5 by model 2

These results are due to longer training duration for model 2 in comparison to model 1. Therefore, model 2 had been selected to be applied on the object tracking system and tested on the actual roads (Table 1).

Model 2 was selected for the object tracking system and tested on road around Faculty of Manufacturing and Mechatronics Engineering Technology (FTKPM) and Residential College 5 (KK5) in UMP. Figures 8, 9 and 10 are some of the image dataset taken during the testing. From the selected image of the object detection results, all classes of object was detected and appeared by the system in the output box respectively. Each output box had displayed the class and the distance to the objects from the RGBD camera of the vehicle. For each image, one color of output box indicates one object. In Fig. 8, there are 3 cars detected with the distances of 9.37, 14.85, 30.45 m from the vehicle respectively. A motorbike is detected with distance of 29.71 m. In Fig. 9, two persons are detected with the distances of 11.28 and 12.00 m from the vehicle. Lastly in Fig. 10, a truck is detected with distance of 12.08 m from the vehicle.



Fig. 10 Vehicle detection in image taken at road to KK5 by model 2

4 Conclusion

This research was conducted to detect the class of objects appear in front of the autonomous vehicle with the estimation of the distance. Then, an object detection system with high accuracy had been the goal of this research. Then, 4 major steps which are data acquisition, data pre-processing, data training and data testing are needed to build an object detection system. The first 3 steps are equally important for a good object detection system to be built because the accuracy of the final model will be affected by them if our data weren't handled well during and before training process. Therefore, YOLO v3 object detection system performed the best result when the second model was trained as the mAP value of it is 99.1% and 0.596 of precision. The accuracy of model is lower, this is probably because the dataset used for model 1 training process did not contain our own dataset, which the images in COCO dataset contain images of foreign countries, the resolution of the features extracted might be different with the ones taken in Malaysia. Because of this factor, model 2 which include the image dataset taken in UMP in training process, performs higher accuracy detection.

Acknowledgements The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2018/TK08/UMP/02/1 (University reference RDU190104) and Universiti Malaysia Pahang for laboratory facilities as well as additional financial support under Internal Research grant RDU1903139.

References

1. Rella S, Veeravalli S, Sai MRS (2019) Object detection and identification
2. Shaikh R (2020) OpenCv perspective transformation. <https://medium.com/analytics-vidhya/open-cv-perspective-transformation-9edffefb2143>
3. Manishgupta (2020) YOLO—you only look once, a state of the art algorithm for real-time object detection system. <https://towardsdatascience.com/yolo-you-only-look-once-3dbdbb608ec4>

4. Kamal A YOLO, YOLOv2 and YOLOv3: all you want to know
5. Nayak S (2019) Training YOLOv3 : deep learning based custom object detector. Learn OpenCV
6. Habeeb Mohamed AH, et al. (2022) Rain classification for autonomous vehicle navigation using machine learning. In: Ab. Nasir AF, Ibrahim AN, Ishak I, Mat Yahya N, Zakaria MA, P. P. Abdul Majeed A (eds) Recent Trends in Mechatronics Towards Industry 4.0. Lecture Notes in Electrical Engineering, vol 730. Springer, Singapore. https://doi.org/10.1007/978-981-33-4597-3_80
7. Čorović A, Ilić V, Đurić S, Marijan M, Pavković B (2018)The real-time detection of traffic participants using YOLO algorithm. In: 26th telecommunications forum, TELFOR 2018
8. Cao C, Wang B, Zhang W, Zeng X, Yan X, Feng Z, Wu Z (2019)An improved faster R-CNN for small object detection. IEEE

A Brief Review on Motion Sickness for Autonomous Vehicle



Sarah 'Atifah Saruchi, Nor Aziyatul Izni, Mohd Hatta Mohammed Ariff, and Nurbaiti Wahid

Abstract Motion sickness (MS) is an unpleasant sensation such as headache and nausea which occurs during travelling by vehicle. Extensive studies had been carried out regarding the factor and the mitigation methods of MS, especially for the vehicle's passengers. Nowadays, a revolution from the automotive industry resulting from the development of the autonomous vehicles. One of the key concerns in autonomous vehicle research is that its possibility to have a higher chance to contribute to MS among the occupants compared to the conventional vehicle. Hence, this paper presents reviews on the MS reduction methods, focusing on the application towards the autonomous vehicle. Considering the importance of MS reduction in improving the occupant's comfort level, it is concluded that this issue requires more attention among autonomous vehicle researchers.

Keywords Motion sickness · Autonomous vehicle · Comfort · User acceptance

1 Introduction

Autonomous vehicle is a driverless vehicle or self-driving vehicle which can operate without human driver. In recent years, autonomous vehicle becoming one of the components that achieved progressive growth in automotive technology. The rapidly

S. 'Atifah Saruchi (✉)

Department of Mechanical and Mechatronic Engineering, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia
e-mail: atifah@ucsiuniversity.edu.my

N. A. Izni

Institute of Actuarial Science and Data Analytics, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia

M. H. M. Ariff

Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

N. Wahid

School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia

evolving technology succeeded in gaining extensive attention among researchers from automotive manufacturers, software corporations and universities [1].

The major benefit of autonomous vehicle technology is that it promises an improvement in terms of safety [2]. In an autonomous vehicle, a safer mode of transportation can be improved by integrating high technologies systems such as advanced sensing, algorithm and collision avoidance [3, 4]. Most vehicle crashes incidents occur due to human error. The transfer from human to vehicle control could avoid the causes of human error such as distraction and inattention.

The autonomous vehicle offers the increment of personal activities to the occupants [5]. The automation enables the driver to do other non-driving activities such as texting, reading and relaxing [6]. The conceptional design of the vehicle also contributes to the engagement between occupants. The rotatable and adjustable occupants' seats allow more interactions among them [7].

The autonomous vehicle provides positive contributions environmentally. It can help to reduce consumption and gas emissions [8]. Another benefit of an autonomous vehicle is the improvement in mobility. It offers more mobility freedom for those who are partially excluded from the individual mobility such as senior citizens, children and disabled persons [9, 10]. In other words, the self-driving capability of the autonomous vehicle gives opportunity to the people who cannot drive themselves.

Besides the advantages, autonomous vehicle researchers face an unavoidable challenge which is the user acceptance. User acceptance is important because it determine whether the vehicle will be used by the public or not [11]. The critical issue that need to be overcome to build the trust from the public is regarding the safety [12]. Faulty sensors, system hacking and crashes caused by other non-autonomous vehicle or pedestrians are some of the harmful possibilities [13–15]. Thus, extensive research had been carried out to solve these problems. For example, development of fault tolerant control [16], pedestrian or lane detection system [17] and collision avoidance control system [18]. It is also reported that researchers started to examine potential security issues [19].

The next challenge in the middle of autonomous vehicle development is the passenger comfort. There is correlation existed between comfort and user acceptance in highly automated driving [20]. Therefore, it is crucial to provide a comfortable driving experience to ensure the acceptance of the autonomous vehicle [21, 22]. The transfer of control from a human driver to an autonomous system brings negative effects toward the path naturalness which resembling human generated paths [12]. The reduction of path naturalness produces unfamiliar manoeuvres, hence cause interruption towards the occupant's comfort level. It has been reported that many researches had been done to execute familiar movement such as modelling human driver control behaviour and application of machine learning to mimic human control [12].

Another issue that affects comfort level among the occupants of an autonomous vehicle is motion sickness (MS). MS is an uncomfortable sensation caused by the movement during travel [23]. Generally, the signs and symptoms of MS include headache, sweating, salivation, dizziness, nausea, vomiting and other physical discomforts [24, 25]. Researchers used various ways to quantify MS such as questionnaires [26, 27], mathematical equations [28, 29] and sensors [30, 31]. Questionnaire

is a simple and cheap method as it is not involving any devices. On the other hand, mathematical equations analyse the MS by using real data of vehicle's movement or occupants. Evaluation of MS can also be done by assessing the response from brain area and abdominal and back muscles.

The loss of controllability of a human driver triggers the MS [24]. MS is considered as a human factor issue that require a better understanding to encourage user acceptance, hence ensure a successful introduction of vehicle automation [20]. It had been reported that the negative effects of MS resulted from autonomous driving is an issue that has gone unnoticed [32]. There were a very minimal number of studies conducted regarding the MS factors and solutions for road vehicle [33]. Thus, this paper focuses on the MS related issues for an autonomous vehicle. Here, the mini review of MS for the autonomous vehicle is presented. Firstly, in the next section, the review of the MS factor is outlined. Then, the reviews on MS minimisation strategies are discussed.

2 Factors of Motion Sickness

One of the MS factors is the lowered capacity of the occupants to estimate the course of movement [34]. Autonomous driving reduces the ability to predict vehicle movement [35]. The shift from the driver's control to autonomous driving reduced the anticipation of future trajectory. Thus, the occupants including the driver will have higher chances to experienced MS, when the driving mode is turned to full automated.

The susceptibility of MS is higher when travelling on a curvy path than a straight path. MS occurs in a low frequency of lateral acceleration environment [36]. The primary factor of MS is lateral acceleration resulted from the driver's steering skill. High lateral acceleration is produced by the excessive driver's turning skill causes a higher MS level. There is a correlation between lateral acceleration, head tilt movement and MS level [37]. Based on the correlation, when a high lateral acceleration produced during cornering, the occupants will angle their heads more towards the lateral acceleration direction. In this situation, their MS level increased. For a conventional vehicle, excessive wheel's turning is preventable by improving the drivers' skills of driving. For an autonomous vehicle where the driving is operated automatically, the main objective for its motion planning system is to get the vehicle to reach its designated destination. Most of the researchers did not consider MS elements in their motion planner [12].

From an interior design perspective, an autonomous vehicle is expected to enhance social interactions among the occupants. Researchers proposed interior flexibility such as rotatable seats. The rotatability allows the occupants at the front seats to face backwards. However, facing backwards will cause a failure of anticipating the future movement [24]. Also, the occupants do not require a window to view the outside environment because they don't have to control the vehicle [26]. The limitation of the visual input affects the ability to predict the vehicle movement sufficiently [38].

The autonomous vehicle allows the occupants to engage with non-driving activities such as chatting, reading, relaxing and playing games. The engagement in those activities contributes to the sensory conflict which then contribute to the increment of MS level [39]. Based on the sensory conflict theory, MS is developed when there is a mismatch between sensory input information from the visual and vestibular system [40]. The vestibular system consists of semi-circular canals and otolith organs in which each of them distinguishes neck rotational and translational motion [2]. These non-driving activities also reduce the anticipation of future vehicle direction among the occupants.

Another concern of autonomous driving is its path naturality. Lack of control from the human can cause abnormality in the vehicle movement. The abnormality or non-smooth movement can cause an unstable posture [23]. Postural instability contributes to develop MS [41]. Occupants who had degraded control of their bodies such as stance and locomotion will experience MS symptoms [42]. It is important for the vestibular system to minimise swaying and maintaining a stable posture to avoid MS. It has been proven that during travels, the occupants who were consistent in their movement were less likely to experience MS than those who exhibited changes in their movement [43].

Figure 1 illustrates the summarisation of the MS factors in autonomous vehicles. It shows the link between the autonomous vehicle features with the factor of MS occurrence.

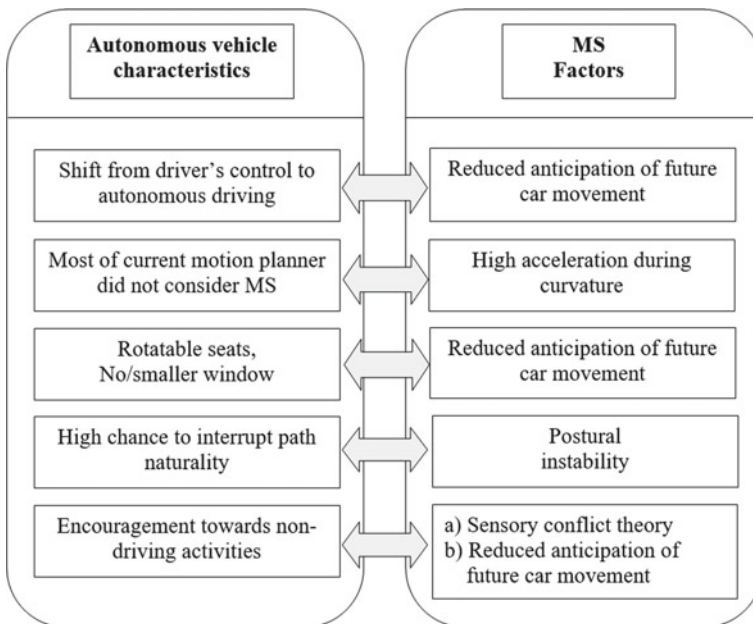


Fig. 1 MS factors in autonomous vehicle

3 Strategies to Minimise Motion Sickness in Autonomous Vehicle

This section presents the review from the previous research on the MS minimisation methods in an autonomous vehicle. The literatures are selected based on the factors mentioned in the previous section.

Based on Fig. 1 in Sect. 2, the most repeated MS factor is the reduction of the ability to predict the future vehicle movement. Engaging with non-driving activities and the interior design of the vehicle are among the reasons on why the anticipation is reduced. Aiming to increase awareness of the vehicle movement among autonomous vehicle occupants, Md. Yusof et al. proposed a device that can alert the occupants about the direction of the autonomous vehicle [44]. The device which is called as a haptic feedback device is attached to the passenger's forearms during travel. During the experiment, the author created an in-vehicle situation where opaque windows were used. The device invention managed to reduce the MS symptoms.

Non-driving activities such as watching television in autonomous vehicles reduce the ability of the passenger to anticipate the future vehicle direction. Karjanto et al. overcome this situation by proposing a peripheral visual feedforward system (PVFS) to provide information of the future vehicle direction [45]. The PVFS device is installed at the right and left sides of the television display inside an autonomous vehicle. Results show that the symptoms of MS were lessened after the installation of the device.

Kuiper et al. [46] proposed the application of auditory cues to reduce MS in the autonomous vehicle. The auditory cues increased the anticipation of the vehicle motion, hence managed to reduce motion sickness. The auditory cues gave information of the timing and the future vehicle direction to the occupants.

Postural instability can be avoided if the occupant of the autonomous vehicle can react accordingly on time when the vehicle receives unexpected forces. The on-time reactions help them to balance their body posture in such situations. The reactions are produced by the human senses with sufficient situation awareness. Md Yusof et al. proposed a vibrotactile display to increase situation awareness in autonomous driving [47]. Results show that the increment of situation awareness due to the implementation of the device can mitigate MS.

Reading in a vehicle is also a non-driving activity that can lead to MS. MS occurs due to sensory conflict and postural instability. DiZio et al. introduced an active suspension system to mitigate the MS but at the same time enable the reading activity in the vehicle [48]. The active suspension system reduced the vibration. Based on the results, the reduction of the vibration can minimise the MS level among the occupant, even though they were tasked to read texts. Ekchian et al. also proposed a high-bandwidth active suspension system to mitigate MS [49]. It was reported that by applying the proposed system, the severity of MS symptoms decreased.

Autonomous vehicles need to install a motion planning system that considers the MS element. In terms of vehicle dynamics, the high lateral acceleration produced during cornering is one of the primary factors of MS. Elbanhawi et al. stated that

the MS issue can be resolved by designing a smooth lateral control system [12]. The author integrated the vehicle's path tracking system with a continuous curvature path planning algorithm. Based on the results, it had been proved that the combination of Pure Pursuit controller and continuous path planner could reduce lateral acceleration, hence improving the comfort level [50].

Integration of a lateral control system with a longitudinal control system can also produce a smooth vehicle's cornering. Technically, a longitudinal control system or speed control system is used to control the acceleration and the deceleration of the vehicle. Wada et al. presented a velocity profile based on the expert driver's profile to be installed in the autonomous vehicle to reduce MS [51]. During curvature, the vehicle's speed will be lowered to achieve low lateral acceleration.

Saruchi et al. proposed the integration of a path tracking system with an active front steering system (AFS) to minimise the lateral acceleration in slalom paths [47, 48, 52, 53]. The author applied a Stanley controller in the path tracking control system and a Fuzzy-Proportional Derivative Integral (PID) controller in the AFS. The combination of these two systems managed to mitigate the MS index in both simulation and experiment platforms [54].

Htike et al. implemented an optimal control formulation in the motion planner of an autonomous vehicle [55]. The application of optimum trajectory and velocity profile succeed in mitigating the MS. The performance is verified through experiments using the curvy path. By using the proposed system, the autonomous vehicle travelled close to the road boundary in a peak cornering, thus produced minimum lateral acceleration.

Table 1 tabulates the summarisation of MS minimisation methods. Based on the table, it can be found that there are various kinds of ways to mitigate MS. The MS

Table 1 Summary of the proposed strategies to reduce MS

How to minimize MS?	The proposed strategies
Increase anticipation of future vehicle direction	<ul style="list-style-type: none"> • Haptic feedback device • Peripheral visual feedforward system • Auditory cues
Increase postural stability	<ul style="list-style-type: none"> • Vibrotactile display • Active suspension system • High-bandwidth suspension system
Reduce sensory conflict	<ul style="list-style-type: none"> • Active suspension system • High-bandwidth suspension system
Consider MS in motion planning system	<ul style="list-style-type: none"> • Integration of path tracking system with continuous curvature path planning algorithm • Integration of lateral and longitudinal systems • Integration of path tracking system with active front steering system • Apply optimal control formulation in motion planner

can be reduced by inventing devices, modifying the vehicle and applying control systems.

4 Conclusion

MS is an issue that can lead to a negative impact on the user acceptance towards the autonomous vehicle because it is closely related to the occupant's comfort. Concerns had been raised by several researchers due to the lack of MS studies in autonomous driving. Hence this paper focuses on reviewing the MS factors and mitigation strategies in an autonomous vehicle. Based on the review, it can be concluded that MS can be minimised using different ways from different perspectives. Due to the importance of tackling the MS issue, further investigations are necessary. The attention towards autonomous vehicle technology keeps increasing day by day. In the future, it is expected that more strategies will be proposed by the researchers.

Acknowledgements This study was supported by Ministry of Higher Education Malaysia (MoHE) through Universiti Teknologi Malaysia (UTM) Fundamental Research Grant Scheme (FRGS/1/2019/TK08/UTM/02/10) Vot No.: R.K130000.7843.5F204 and Malaysian Institute of Road Safety Research (MIROS) through Asean NCAP Colaborative Holistic Research Grant (ANCHOR- PHASE 3) Vot No.: R.K130000.7343.4B582.

References

1. Bellem H, Klüver M, Schrauf M, Schöner HP, Hecht H, Krems JF (2017) Can we study autonomous driving comfort in moving-base driving simulators? A validation study. *Human Fact* 59(3):442–456
2. Iskander J, Attia M, Saleh K, Nahavandi D et al (2019) From vehicle sickness to autonomous vehicle sickness: a review. *Transp Res F Traffic Psychol Behav* 62:716–726
3. Tian D, Wu G, Boriboonsomsin K, Barth MJ (2017) A co-benefit and tradeoff evaluation framework for connected and automated vehicle applications. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, USA, pp 953–958
4. Seppelt BD, Victor TW (2016) Potential solutions to human factors challenges in road vehicle automation. In: Meyer G, Beiker S (eds.) *Road vehicle automation 3. lecture notes in mobility*. Springer, Cham, pp 131–148
5. Brenner W, Herrmann A (2018) An overview of technology, benefits and impact of automated and autonomous driving on the automotive industry. In: *Digital marketplaces unleashed*. Springer-Verlag GmbH Germany, pp 427–442
6. Offer GJ (2015) Automated vehicles and electrification of transport. *Energy Environ Sci* 8(1):26–30
7. Jorlöv S, Brohman K, Larsson A (2017) Seating positions and activities in highly automated vehicles – a qualitative study of future automated driving scenarios. In: *Ircobi Conference*, Antwerp, Belgium, pp 13–22
8. Sarkar R, Ward J (2016) DOE smart mobility: systems and modeling for accelerated re-search in transportation. In: Meyer G, Beiker S (eds.) *Road vehicle automation 3. Lecture Notes in Mobility*. Springer, Cham, pp 39–52

9. Chan CY (2017) Advancements, prospects, and impacts of automated driving systems. *Int J Transp Sci Technol* 6(3):208–216
10. Paden B, Cap M, Yong SZ, Yershov D, Frazzoli E (2016) A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans Intell Veh* 1(1):33–55
11. Cunningham M, Regan MA (2015) Autonomous vehicle: human factors issues and future research. In: *Proceedings of the 2015 Australasian road safety conference*, Gold Coast, pp 1–11
12. Elbanhawi M, Simic M, Jazar R (2015) In the passenger seat: investigating ride comfort measures in autonomous vehicles. *IEEE Intell Transp Syst Mag* 7(3):4–17
13. Diels C, et al (2017) Designing for comfort in shared and automated vehicles (SAV): a conceptual framework. In: *1st International Comfort Congress*, Salerno, pp 1–8
14. Van Brummelen J, O'Brien M, Gruyer D, Najjaran H (2018) Autonomous vehicle perception: the technology of today and tomorrow. *Transp Res Part C Emerg Technol* 89:384–406
15. Bagloee SA, Tavana M, Asadi M, Oliver T (2016) Autonomous vehicles: challenges, opportunities and future implications for transportation policies. *J Mod Transp* 24(4):284–303
16. Boukhari MR, Glaser S, et al (2017) Fault tolerant design for autonomous vehicle. In: *4th international conference on control, decision and information technologies (CoDIT)*, Barcelona, pp 0721–0728
17. Zakaria NJ, Zamzuri H, Ariff MH et al (2018) Fully convolutional neural network for Malaysian road lane detection. *Int J Eng Technol* 7:152–155
18. Wahid N, Zamzuri H, Amer NH et al (2020) Collision avoidance motion planning strategy using artificial potential field with adaptive multi-speed scheduler. *IET Intell Transp Syst* 14(10):1200–1209
19. De La Torre G, Rad P, Choo KKR (2018) Driverless vehicle security: challenges and future research opportunities. *Fut Gener Comput Syst* 108:1–6
20. Bellem H, Thiel B, Schrauf M, Kreams JF (2018) Comfort in automated driving: an analysis of preferences for different automated driving styles and their dependence on personality traits. *Transp Res Part F Traffic Psychol Behav* 55:90–100
21. Hartwich F, Beggiato M, Kreams JF (2018) Driving comfort, enjoyment and acceptance of automated driving—effects of drivers' age and driving style familiarity. *Ergonomics* 61(8):1017–1032
22. Eriksson J, Svensson L (2015) Tuning for ride quality in autonomous vehicle application to linear quadratic path planning algorithm. *Uppsala Universitet*
23. Green P (2016) Motion sickness and concerns for self-driving vehicles: A Literature Review Technical Report UMTRI-2016. University of Michigan Transportation Research Institute, USA, pp 1–83
24. Diels C (2014) Will autonomous vehicles make us sick? In: *Contemporary Ergonomics and Human Factors 2014*. Taylor and Francis, Southampton, United Kingdom, pp 301–307
25. Kennedy RS, Drexler J, Kennedy RC (2010) Research in visually induced motion sickness. *Appl Ergon* 41(4):494–503
26. Kuiper OX, Bos JE, Diels C (2018) Looking forward: In-vehicle auxiliary display positioning affects vehiclesickness. *Appl Ergon* 6:169–175
27. Treleaven J et al (2015) Simulator sickness incidence and susceptibility during neck motion-controlled virtual reality tasks. *Virtual Real* 19(3–4):267–275
28. Aykent B, Merienne F, et al (2014) Motion sickness evaluation and comparison for a static driving simulator and a dynamic driving simulator. *Proc Inst Mech Eng Part D J Automob Eng* 228:1–12
29. Wada T, Fujisawa S, et al (2010) Effect of driver's head tilt strategy on motion sickness incidence. In: *The 11th IFAC/IFIP/IFORS/IEA symposium on analysis, design and evaluation of human-machine system*, France, vol 11, pp 1–6
30. Lin CT, Tsai SF, Ko LW (2013) EEG-based learning system for online motion sickness level estimation in a dynamic vehicle environment. *IEEE Trans Neural Netw Learn Syst* 24(10):1689–1700

31. Shafeie M, Zolfaghari N, Valter McConville KM (2013) Abdominal muscle behavior and motion sickness during paired visual input with roll motion. In: 2013 6th international conference on human system interactions, Poland, pp. 247–251
32. Smyth J, Jennings P, et al (2018) Too sick to drive: how motion sickness severity impacts human performance. In: 21st IEEE conference on intelligent transportation systems, proceedings, ITSC, USA, pp. 1787–1793
33. Jones MLH (2019) Queasy passengers: a testbed for motion sickness in driverless vehicles. Technical Report 9. University of Michigan Transportation Research Institute, USA, pp 1–9
34. Golding JF, Gresty MA (2005) Motion sickness. *Curr Opin Neurol* 18(1):29–34
35. Diels C, Bos JE (2015) User interface considerations to prevent self-driving vehiclesickness. In: Adjunct proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications. Association for Computing Machinery, New York, pp 14–19
36. Turner M, Griffin MJ (1999) Motion sickness in public road transport: the effect of driver, route and vehicle. *Ergonomics* 42(12):1646–1664
37. Wada T, Fujisawa S, Doi S (2018) Analysis of driver's head tilt using a mathematical model of motion sickness. *Int J Ind Ergon* 63:89–97
38. Sivak M (2015) Motion sickness in self-driving. Technical Report UMTRI-2015-12. University of Michigan Transportation Research Institute, pp 1–15
39. Diels C, Bos JE (2016) Self-driving vehiclesickness. *Appl Ergon* 53(B):374–382
40. Zhang LL, Wang JQ, Qi RR, Pan LL, Li M, Cai YL (2015) Motion sickness: current knowledge and recent advance. *CNS Neurosci Ther* 22(1):15–24
41. Riccio GE, Stoffregen TA (1991) An ecological theory of motion sickness and postural instability. *Ecol Psychol* 3:195–240
42. Koslucher F, Haaland E, Stoffregen TA (2016) Sex differences in visual performance and postural sway precede sex differences in visually induced motion sickness. *Exp Brain Res* 234(1):313–322
43. Dong X, Stoffregen TA (2010) Postural activity and motion sickness among drivers and passengers in a console video game. In: Proceedings of the human factors and ergonomics society annual meeting, vol 54, no 18. SAGE Publishing, pp 1340–1344
44. Yusof NM, Karjanto J, Kapoor S, Terken J, Delbressine F, Rauterberg M (2014) Experimental setup of motion sickness and situation awareness in automated vehicle riding experience. In: Adjunct proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications (AutomotiveUI 2017). Association for Computing Machinery, New York, pp 104–109
45. Karjanto J, Yusof NM, Waeg C, Terken J, Delbressine F, Rauterberg M (2018) The effect of peripheral visual feedforward system in enhancing situation awareness and mitigating motion sickness in fully automated driving. *Transp Res F Traffic Psychol Behav* 58:678–692
46. Kuiper OX, Bos JE, Diels C, Schmidt EA (2020) Knowing what's coming: anticipatory audio cues can mitigate motion sickness. *Appl Ergon* 85:103068
47. Yusof NM, Karjanto J, Terken JMB, Delbressine FLM, Rauterberg GWM (2020) Gaining situation awareness through a vibrotactile display to mitigate motion sickness in fully-automated driving vehicles. *Int J Autom Mech Eng* 17(1):7771–7783
48. Dizio P, Ekchian J, Kaplan J, Ventura J et al (2018) An active suspension system for mitigating motion sickness and enabling reading in a vehicle. *Aeros Med Human Perf* 89(9):822–829
49. Ekchian J, Graves W, Anderson Z, Giovanardi M, et al (2016) A high-bandwidth active suspension for motion sickness mitigation in autonomous vehicles. SAE Technical paper 2016-01-1555. Brandeis University, pp 1–10
50. Elbanhawi M, Simic M, Jazar R (2015) The role of path continuity in lateral vehicle control. *Procedia Comput Sci* 60:1289–1298
51. Wada T (2016) Motion sickness in automated vehicles: the elephant in the room. In: Proceedings of the 13th International Symposium on Advanced Vehicle Control (AVEC 2016), Germany, pp 1–6
52. Saruchi SA, Ariff MHM, Zamzuri H, Amer NH, Wahid N, Hassan N, Kadir ZA (2020) Lateral control strategy based on head movement responses for motion sickness mitigation in autonomous vehicle. *J Braz Soc Mech Sci Eng* 42(223):1–14

53. Saruchi SA, Ariff MHM, Zamzuri H, Amer NH, Wahid N, Hassan N, Izni NAA (2021) Lateral control with neural network head roll prediction model for motion sickness incidence minimisation in autonomous vehicle. *J Eng Sci Technol* 16(2):1630–1643
54. Saruchi SA, Ariff MHM, Zamzuri H, Amer NH, Wahid N, Hassan N, Kassim KAA (2020) Novel motion sickness minimization control via fuzzy-PID controller for autonomous vehicle. *Appl Sci* 10(14):4769
55. Htike Z, Papaioannou G, Siampis E, Velenis E, Longo S (2020) Minimisation of motion sickness in autonomous vehicles. In: 2020 IEEE intelligent vehicles symposium (IV). IEEE, pp 1135–1140

Investigation of the Combination of Kinematic Path Planning and Artificial Potential Field Path Planning with PI Controller for Autonomous Emergency Braking Pedestrian (AEB-P) System



Z. Abdullah, P. M. Heerwan, M. A. Zakaria, and M. I. Ishak

Abstract Autonomous Emergency Braking Pedestrian (AEB-P) is a fundamental capacity of the advanced driver assistance system (ADAS) to maintain a distance and avoid a collision. The fatality of pedestrian in traffic accident as well as near-miss accidents are the most frequent type of accidents in Malaysia as the improvisation of AEB-P system are obligatory. To generate optimum vehicle deceleration from the path planner in the presence of a pedestrian in front of the vehicle, an Artificial Potential Field (APF) path planner with a kinematic path planner is proposed in this research. The kinematic path planner will produce maximum deceleration for the vehicle, 8 m/s^2 , as the vehicle violates the threshold. The value is combining with the APF value to fetch to the PI controller. Thus, the AEB-P system was designed considering the pedestrian walked in front of the vehicle at 4.32 km/h and vehicle travelled at 60 km/h, dry and wet road surface condition, time for Front Collision Warning (FCW), and full braking was included for the limit APF is developed. The PI controller will tune the deceleration using its variable on dry road surface ($P = 0.003$, $I = 5$) and on wet road surface ($P = 0.003$, $I = 8500$). The host vehicle starts to give warning signal at 37.29 m and experience full braking at 21.3 m when the vehicle travel on both types of surfaces. The vehicle manages to stop from hitting the pedestrian at 2.21 and 1.5 m on the dry and wet road surface. The proposed AEB-P architecture can avoid the collision with pedestrian as the vehicle manage to stop from hitting the obstacle at a comfortable distance.

Keywords Autonomous Emergency Braking (AEB-P) · Artificial Potential Field (APF) path planner · Kinematic path planner · PI controller

Z. Abdullah · P. M. Heerwan (✉)

Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia
e-mail: mheerwan@ump.edu.my

M. A. Zakaria

Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia

M. I. Ishak

College of Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

1 Introduction

Advanced Driver Assistance Systems (ADAS) is an active safety system to prevent collision. Adaptive Cruise Control (ACC), Autonomous Emergency Braking (AEB), Anti-lock Braking System (ABS) are examples of an active safety system in ADAS [1]. For collision avoidance involving uncertain obstacle, ADAS offers Autonomous Emergency Braking Pedestrian (AEB-P) to maintain a distance or avoid a collision.

Pedestrian-vehicle collisions are viewed as the most deliberate type of accidents since they yield a high fatality rate. In Malaysia, pedestrian fatalities registered roughly 500 deaths each year [2]. According to the Malaysian Institute of Road Safety Research (MIROS) finding, frontal collisions are the most frequent type of accident in Malaysia [3]. However, practically the road conditions are complex, whereby the brake distances would be affected by numerous factors, such as tyre properties, road friction, brake pad exhaustion, etc. Furthermore, the most extreme deceleration would be changed with conditions as well as tyre and brake system conditions [4].

The AEB-P system's ability can be improved by integrating the system with artificial potential field (APF) risk assessment [5–7]. Another pivotal aspect of the collision avoidance system of an autonomous vehicle is its ability to provide a possible replanned path in a hazardous situation. The concept of the Artificial Potential Field (APF) method is the robot move by the results of the attractive force and repulsive force summation in its surrounding [8]. The attractive force is the force that attracts the robot to move to its destination, while repulsive force is the force that pushes the robot away from the obstacle [9]. The favour of an APF replanned path usually has a limited amount of arithmetic [10]. A simple kinematic approach is enough to accommodate the collision avoidance (CA) objective for some uncomplicated avoidance scenarios [11]. The future replanned trajectory has been formulated based on a vehicle kinematic model as it is enabling an immediate replanning process such as braking. It is not dependable in a complicated scenario due to its simplicity as it is not considered many vehicles restraint nor its algorithm.

The near-miss accident still happens, where the distance of the host vehicle is too close to the obstacle [3]. So, to avoid this phenomenon, enhancing the AEB-P product, which acknowledges and manages the safe distance towards the collision point, is developed. Subsequently, conventional Autonomous Emergency Braking Pedestrian systems (AEB-P) have limitations in avoiding a collision when the vehicle moves at medium and high speed [3]. This paper analyses the combination of kinematic path planning with Artificial Potential Field path planning to get a feasible minimum safety distance of 1.0–2.5 m from the obstacle. The PI controller is used to track the difference value between both path planners to convert into feasible deceleration values.

2 Numerical Analysis on Autonomous Emergency Braking Pedestrian (AEB-P) System

2.1 The Architecture of the AEB-P System

AEB-P is a system that helps the vehicle to avoid collision by executing an emergency braking during the appearance of the pedestrian. Figure 1 shows the AEB-P system is operated when the pedestrian occurs in front of a vehicle lane [12]. This system is linked to TTC that enforced to a time for a vehicle passing a pedestrian and to TTC_v where the time of a pedestrian meets the road border and to TTC_e whereby the time for the pedestrian escaped the vehicle by surpass the road border. TTC_p stand for the value of time when pedestrian surpassed the point of TTC_v . The collision will occur when the algorithm of the system; $TTC_v \leq TTC \leq TTC_e$ & $TTC_p \leq TTC_e$ [12]. The distance, $\times 1$ is a safety margins as the pedestrian enter the road lane, while $\times 2$ is a safety margin as the pedestrian exit the road lane.

Figure 2 shows the block diagram of the AEB-P system that consists of time-to-collision, combination of kinematic path planner with APF path planner, brake actuator, vehicle dynamic modelling as well as PI controller.

Fig. 1 Driving scenario during presence of pedestrian

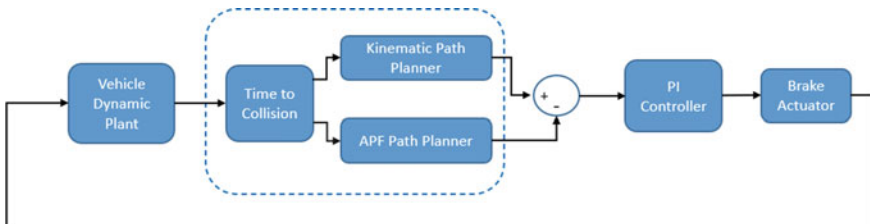
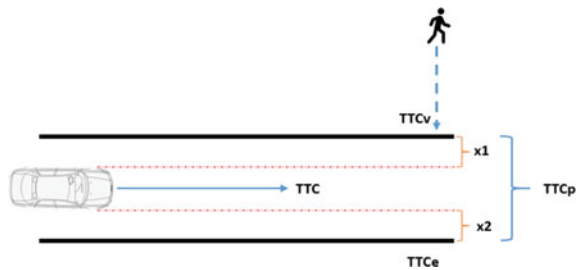


Fig. 2 Block diagram of the AEB-P system

Fig. 3 Free body diagram of the vehicle model

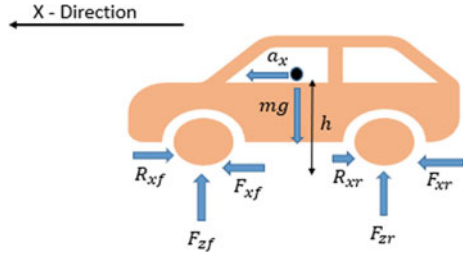


Table 1 Optimized Vehicle Parameters

Details	Symbol	Value	SI unit
Mass (Kerb Weight)	m	1330	kg
Center of gravity (c.g) length towards frontal part	L_f	1.107	m
Center of gravity (c.g) length towards rear part	L_r	1.643	m
Height of center gravity	h	0.479	m
Effective radius of the tyre	r_{eff}	0.393	m

2.2 Vehicle Dynamic Modelling

Longitudinal Dynamic Model

A simplified vehicle dynamic design is modelling in this paper to analyze the dynamic of the vehicle behaviour. Figure 3 shows a simplified model comprises longitudinal model.

In analytical analysis, the vehicle’s force is expected to drive on a level and straight road. The bicycle model is used in the analysis by assuming the left and right tyres are the same. The linear longitudinal forces are expressed as follows:

$$ma_x = F_{xf} + F_{xr} - R_x \tag{1}$$

where m is the mass of the vehicle, a_x is the longitudinal acceleration, F_{xf} and F_{xr} is the longitudinal force for front and rear tire, respectively, R_x is the rolling resistance. In the simulation, the UMP Test Car is used as the vehicle model. The specifications of the vehicle are shown in Table 1 [13].

2.3 AEB-P Path Planner

2.3.1 Time to Collision (TTC)

TTC is the time for the driver to reduce the speed of a vehicle by braking to prevent collision with the front obstacle. TTC is calculated by deriving the kinematic model

equation as shown in Eq. 2.

$$|p| = \begin{cases} -vt, & a = 0 \\ -vt + \frac{1}{2}at^2, & a \neq 0 \end{cases} \tag{2}$$

The TTC can be taken by rearranging the kinematic model equation, as shown in Eq. 3. The velocity of the host vehicle is symbolized as v while the acceleration of the host vehicle is symbolized as α . The time taken for the host vehicle to reach at certain distance is symbolized as t while the distance travelled by the host vehicle is represented as p .

$$|TTC| = \begin{cases} \frac{-v \pm \sqrt{v^2 + 2pa}}{a}, & v \geq 0 \text{ and } a < 0 \end{cases} \tag{3}$$

2.3.2 Kinematic Path Planner

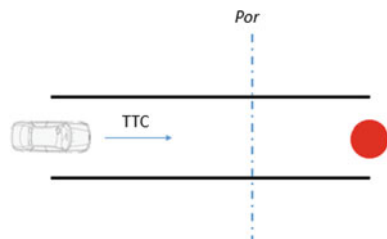
Kinematic path planner is an algorithm derived from the kinematic equation used in the AEB-P path planner. The fundamental kinematic algorithm is suitable for an uncomplicated scenario of collision avoidance such as the appearance of obstacle in front of the vehicle as the vehicle is approaching it. Figure 4 shows the threshold activation for kinematic path planner, whereas TTC is a time-to-collision while ρ_{or} is the maximum safety distance for braking.

Each variable in kinematic path planner threshold, ρ_{or} such as d_0 , v_c and a_{max} stands for critical safety distance, which is 2 m, current velocity, and maximum deceleration of the vehicle as shown in Eq. 4.

$$\rho_{or} = d_0 + (v_c \times TTC) + \frac{v_c^2}{2a_{max}} \tag{4}$$

Equation 5 shows that the kinematic path planner threshold violation as the value of spacing headway of the vehicle, ρ_r cross the maximum safety distance produce by kinematic path planner, ρ_{or} . The maximum deceleration is set to 8 m/s^2 and provided to the braking actuator when the $\rho_r \leq \rho_{or}$.

Fig. 4 Threshold activation for kinematic path planner



$$a_k = \begin{cases} 8, & \rho_r \leq \rho_{or} \\ 0, & \rho_r > \rho_{or} \end{cases} \tag{5}$$

2.3.3 Artificial Potential Field (APF) Path Planner

The Vehicle Conditional Artificial Potential Field (VC-APF) is different from the traditional APF as the vehicle condition influences the design. The manipulated range of the repulsion force is ρ_{or} and is influenced by the speed, maximum deceleration as well as time headway of the vehicle while n and m are represented as a repulsive gain factor and mass of the vehicle. Equation 6, 7 and 8 show the full braking, warning signal, and no signal while Eq. 9 shows the deceleration after PI controller tuning.

$$F_{rep}(X) = \frac{1}{2}n\left(\frac{1}{\rho_r} - \frac{1}{\rho_{or}}\right)\frac{1}{\rho_r^2} \quad \text{if, } \rho_r \leq \rho_{or} \tag{6}$$

$$F_{rep}(X) = \text{warning signal} \quad \text{if, } \rho_{or} \leq \rho_r \leq \rho_{or} + 1.5 \tag{7}$$

$$F_{rep}(X) = \text{no signal} \quad \text{if, } \rho_r \geq \rho_{or} + 1.5 \tag{8}$$

$$\alpha_{pi} = \frac{F_{rep}(X)}{m} \tag{9}$$

2.3.4 Proportional Integrative (PI) Controller

The PI controller is a simple control formulation that is widely used in various industries. Each of the PI formulation variables stands for proportional and integrative as it has its abilities. The deceleration error is the difference of the deceleration between the kinematic path planner and APF path planner as shown in Fig. 5.

Table 2 and 3 shows the PI controller parameter where the AEB-P system is simulated on the dry and wet road surface conditions. The value for PI controller is chosen regarding the minimum safety distance of the vehicle from an obstacle. From

Fig. 5 Block diagram of AEB-P controller

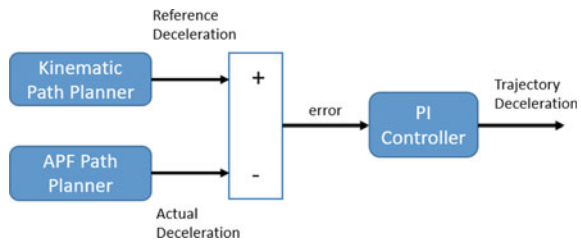


Table 2 PI controller parameter on dry road surface condition

Road surface	P	I	Minimum safety distance (m)
Dry	0.003	5	2.21
	0.003	2.5	2.42
	0.003	7.5	2.21
	3	5	2.72
	7.5	5	2.72

Table 3 PI controller parameter on wet road surface condition

Road surface	P	I	Minimum safety distance (5)
Wet	0.003	8500	1.5
	0.003	4500	1.22
	0.003	12,500	1.11
	50	8500	1.5
	5000	8500	1.5

the result in Table 2 and 3, the chosen values for proportional variable, P on dry and wet road surface is 0.003 while the integrative variable, I on dry and wet road surface are 5 and 8500.

2.3.5 Brake Actuator System

The disc brake system operated as the wheel slowed down and finally stopped due to output of braking torque by actuation force at the wheel. The braking torque, T_{br} produced by the multiplication of the actuating force, F_a and effective radius, r_{eff} while the F_{clamp} stands for the force that clamped the disc brake. $F_{friction}$ stands for frictional force occurred at the disc brake after the actuation force acting on it.

$$F_{clamp} = 2F_a \quad (10)$$

$$F_{friction} = \mu F_{clamp} \quad (11)$$

$$T_{br} = F_{friction} r_{eff} \quad (12)$$

The braking torque on disc brake can be written as in Eq. 13 where μ is stand for friction coefficient of the brake pad.

$$T_{br} = 2\mu F_a r_{eff} \quad (13)$$

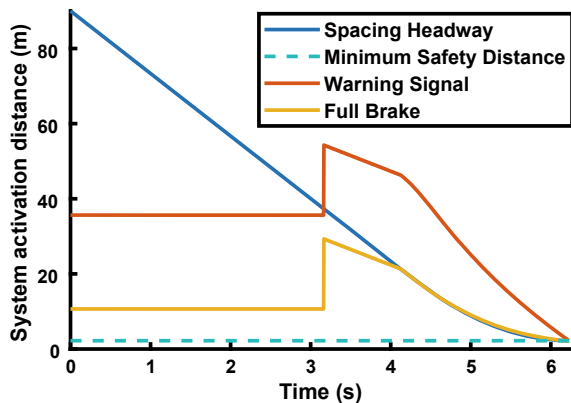
3 Results and Discussion

3.1 Dry Road Surface Condition

Figure 6 shows the system activation when the vehicle is approaching the obstacle. It can be seen that the warning signal starts to be activated at time 3.166 s. In this condition, the spacing headway is crossing the warning signal. The emergency brake starts to be activated at time 4.1248 s when the spacing headway crosses the full brake. As a result, the path planner starts to give deceleration for the emergency braking phase, as shown in Fig. 7(a), (b) and (c). Figure 6 shows that the vehicle stops at 2.21 m, as shown by the sky-blue dotted line.

Figure 7(a), (b) and (c) indicates the behaviour of deceleration for the host vehicle as activation of AEB-P system occur at 4.1248 s. Figure 7(a) states that the behaviour produced by APF path planning as it increases smoothly until it achieves the maximum value, which is 0.0015 m/s^2 at time 6.05 s. Figure 7(b) shows the behavior of the reference deceleration produced by the kinematic path planner as the maximum value is set to be 8 m/s^2 . The deceleration starts to oscillate heavily as the value of spacing headway is abruptly enter and exit full brake phase as it is influencing the braking actuator from 4.1248 to 4.5646 s as shown in Fig. 8. Figure 7(c) shows the effect of proportional and integrative variables on deceleration behaviour after the value from APF are tuning by the PI controller based on the reference deceleration from kinematic path planner. The combination of deceleration from APF and kinematic path planner is highlighted in bold red colour which is the area of effected by PI variable in Fig. 7(c). The proportional variable act as a gain to increase the step response of the deceleration while the integrative variable is to reduce the oscillation. The deceleration starts to increase when the brake was applied at 4.1248 s and reaches the maximum value, which is 8 m/s^2 at 4.5646 s and stays constant until the vehicle stops colliding with obstacles.

Fig. 6 System activation distance against time on dry road surface



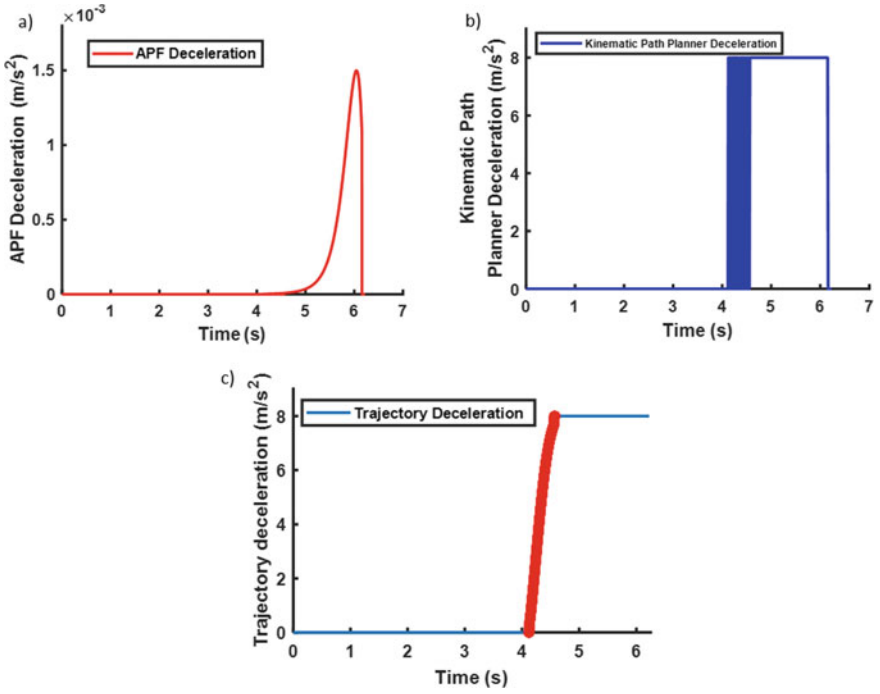
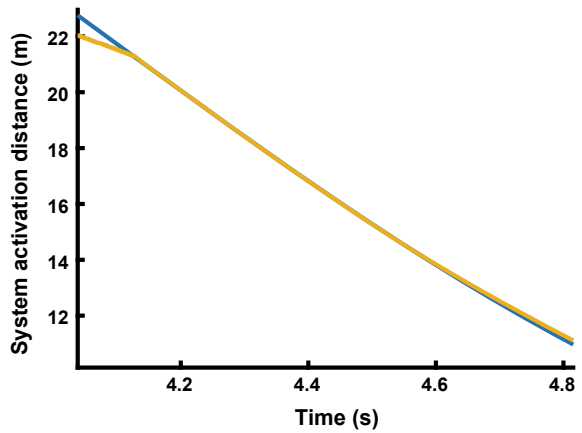


Fig. 7 The behavior of the deceleration on dry road surface **a)** APF deceleration, **b)** Deceleration from kinematic path planner **c)** Trajectory deceleration

Fig. 8 Oscillation behavior for system activation distance against time on dry road surface

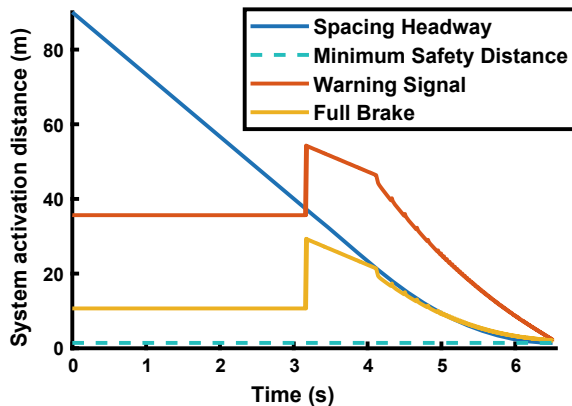


3.2 Wet road surface condition

Figure 9 shows the system activation distance when a vehicle approaches an obstacle on a wet road surface. The warning signal starts to be triggered at time 3.166 s when it crosses the spacing headway. Meanwhile, the emergency braking starts to activate at time 4.1248 s. The vehicle stops at 1.5 m, as shown by the sky blue dotted line in Fig. 9.

Figure 10(a), (b) and (c) shows the behaviour of deceleration for the host vehicle as activation of the AEB-P system occur at 4.1248 s. Figure 10(a) shows the behaviour of deceleration produced by APF path planning as it is increasing in light oscillation pattern until it achieves the maximum value, which is 0.00672 m/s^2 at time 6.3953 s. The value pattern decrease as a final value for deceleration is 0.0047 m/s^2 at 6.506 s when the vehicle is in a stationary position. Figure 10(b) illustrated the behavior of the reference deceleration produced by the kinematic path planner as the maximum value for the reference deceleration is set to be 8 m/s^2 . The deceleration is initiated at time 4.1248 s and oscillate until at the time of 4.125 s. Then, the value of the deceleration stays constant until the vehicle reaches at time 4.9533 s. However, the deceleration starts to oscillate again as the oscillation phenomenon is due to the spacing headway's value abruptly entering and exit the maximum safety distance region as it directly influences the braking actuator. Finally, the deceleration stays constant at a maximum deceleration of 8 m/s^2 at time 5.18 s until the vehicle stops. The oscillation of deceleration is due to the full braking and spacing headway indicator, as shown in Fig. 11. After PI controller is tuning, the vehicle's actual trajectory deceleration during the brake actuation is smooth, as in Fig. 10(c). Figure 10(c) shows the effect of proportional and integrative variables in the PI controller for trajectory deceleration. The proportional variable act as a gain to increase the step response of the deceleration as highlighted in red colour at the vertical line, which is from 4.1248 to 4.125 s and from 4.9531 to 4.9533 s. The integrative variable is to increase the zero value from 4.125 to 4.9531 s and reduce the oscillation from

Fig. 9 System activation distance against time on wet road surface



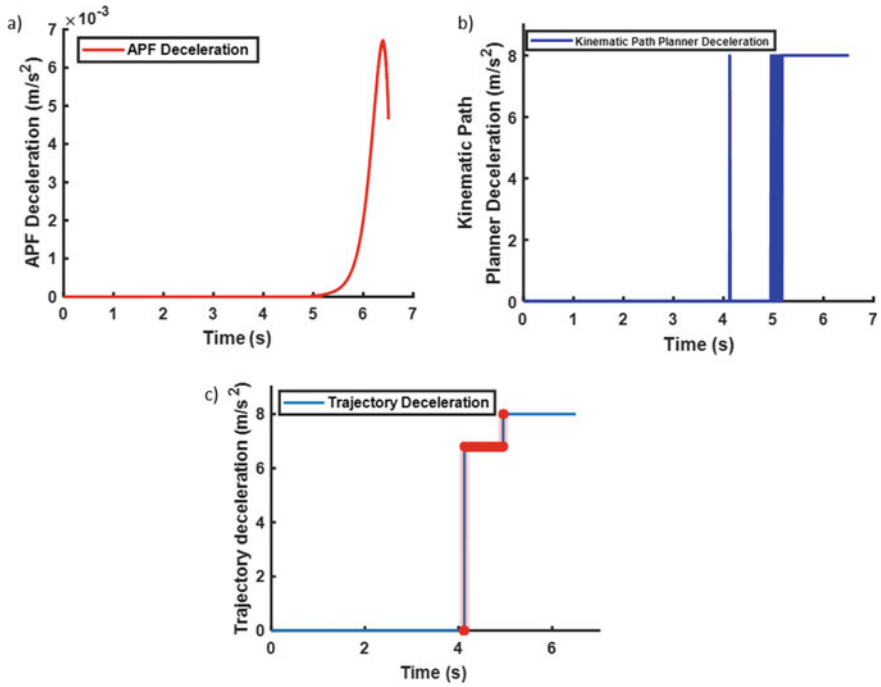
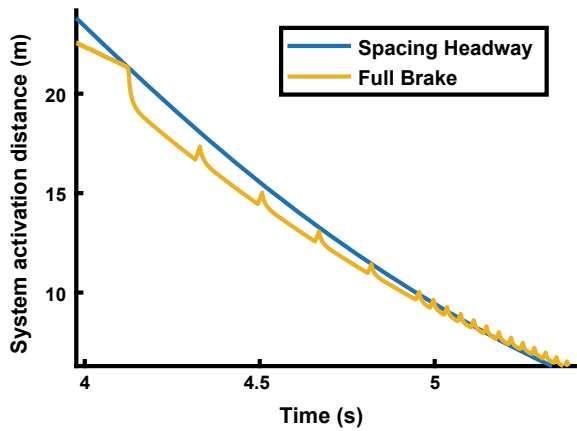


Fig. 10 The behavior of the deceleration on wet road surface a) APF deceleration, b) Deceleration from kinematic path planner c) Trajectory deceleration

Fig. 11 Oscillation behavior for system activation distance against time on wet road surface



4.9531 to 5.1828 s after the combination of deceleration from APF and kinematic path planner as indicated as bold red horizontal line in Fig. 10(c).

4 Conclusion

This paper implements the AEB-P architecture on the dry and wet road surface when the vehicle travels at a speed of 60 km/h and pedestrian walks in front of the vehicle at a speed of 4.32 km/h. The combination of APF and kinematic path planner is proposed in this paper to yield optimal deceleration performance of a vehicle to ensure the vehicle manages to stop at a comfortable distance from an obstacle. Compared to conventional APF path planners, the proposed path planner is feasible. The algorithm formulation includes the vehicle velocity, maximum deceleration, and the minimum safety distance after the vehicle stop. The proposed AEB-P system also can avoid a near-miss accident as the simulation result shows that the deceleration produce is enough to stop the vehicle in the dry and wet road surface condition. Next, the higher speed of the host vehicle should be considered during the simulation or will be used for the validation of the system. The various braking scenario should be included for the researcher to create a robust AEB-P architecture system.

Acknowledgements The task given in this researched is funded by University Malaysia Pahang through financial support via the UMP internal grant RDU192219. The author would like to thank the Automotive Engineering Center, University Malaysia Pahang (www.ump.edu.my) for giving test car and technical assistance.

References

1. Hojjati-Emami K, Dhillon B, Jenab K (2012) Reliability prediction for the vehicles equipped with advanced driver assistance systems (ADAS) and passive safety systems (PSS). *Int J Ind Eng Comput* 3(5):731–742
2. Syazwan SM, Baba MD, Zarifah HN, Hafeez AA, Faradila PN (2017) Prevalence of distracted pedestrians while crossing: a study of Malaysia's situation. *MATEC Web Conf* 90:01031
3. Hamid UZA, et al (2017) Autonomous emergency braking system with potential field risk assessment for frontal collision mitigation. In: 2017 IEEE conference on systems, process and control, ICSPC 2017, vol 2018-January. Institute of Electrical and Electronics Engineers Inc, pp 71–76
4. Chen YL, Shen KY, Wang SC (2013) Forward collision warning system considering both time-to-collision and safety braking distance. In: 2013 IEEE 8th conference on industrial electronics and applications (ICIEA). IEEE, pp 972–977
5. Ruan Y, Chen H, Li J (2017) Longitudinal planning and control method for autonomous vehicles based on a new potential field model. *SAE Technical Paper* 0148–7191
6. Wang H, Huang Y, Khajepour A, Zhang Y, Rasekhipour Y, Cao D (2019) Crash mitigation in motion planning for autonomous vehicles. *IEEE Trans Intell Transp Syst* 20(9):3313–3323. Art no 8617711

7. Yang W, Zhang X, Lei Q, Cheng X (2019) Research on longitudinal active collision avoidance of autonomous emergency braking pedestrian system (AEB-P). *Sensors (Switzerland)* 19(21). Art no 4671
8. Hamid UZA, et al (2017) Autonomous emergency braking system with potential field risk assessment for frontal collision mitigation. In: Presented at the 2017 IEEE conference on systems, process and control (ICSPC)
9. Tang J, Sun J, Lu C, Lao S (2019) Optimized artificial potential field algorithm to multi-unmanned aerial vehicle coordinated trajectory planning and collision avoidance in three-dimensional environment. *Proc Inst Mech Engineers Part G J Aerosp Eng*
10. Feng Y, Wu Y, Cao H, Sun J (2018) UAV formation and obstacle avoidance based on improved APF. In: 10th international conference on modelling, identification and control, ICMIC 2018. Institute of Electrical and Electronics Engineers Inc
11. Brezak M, Petrović I (2013) Real-time approximation of clothoids with bounded error for path planning applications. *IEEE Trans Rob* 30(2):507–515
12. Lee HK, Shin SG, Kwon DS (2017) Design of emergency braking algorithm for pedestrian protection based on multi-sensor fusion. *Int J Autom Technol* 18(6):1067–1076
13. Saito Y, Raksincharoensak P (2016) Shared control in risk predictive braking maneuver for preventing collisions with pedestrians. *IEEE Trans Intell Vehicles* 1(4):314–324
14. Zuhlilmi I, Peeie M, Asyraf S, Sollehudin I, Ishak I (2020) Experimental study on the effect of emergency braking without anti-lock braking system to vehicle dynamics behaviour. *Int J Autom Mech Eng* 17(2):7832–7841

Simulation Studies on Bimorph and Unimorph PZT Piezoelectric Transducer for Energy Harvesting Application



Sanjeef Singh A/L Sarbjeet Singh, Sew Sun Tiang, Wei Hong Lim, Kah Hou Teng, and Chin Hong Wong

Abstract Various studies have shown that energy harvesting using piezoelectric material is a promising path opened to more opportunities for future applications. This study is mainly focused on the comparison between bimorph and unimorph cantilever transducer with three parameters taken into consideration, which are the weight of proof mass, types of piezoelectric material used and the dimension of both the cantilever transducer. For the bimorph transducer, various investigated results show that it is able to generate a maximum output voltage of 14.66 V and 1.07 mW electrical power at natural frequency 114 Hz. On the other hand, the unimorph transducer is able to generate a maximum output voltage of 12.81 V and 0.82 mW electrical power at natural frequency 139 Hz. Hence, both designs are proposed to use as an energy harvester device.

Keywords Piezoelectric · PZT · Cantilever · Bimorph · Unimorph

1 Introduction

The world is currently facing an issue of depletion of natural resources. There are many sources for renewable energy harvesting such as solar, wave, wind and biomass [1]. Renewable energy is a topic that is being researched on different levels. The main idea of harvesting renewable energy is to utilize piezoelectric materials to capture the vibration energy from the ambient surroundings and convert it into electrical energy [2]. Some of the common piezoelectric materials used for energy harvesting

S. S. A/L Sarbjeet Singh · S. S. Tiang (✉) · W. H. Lim · K. H. Teng
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, Malaysia
e-mail: tiangss@ucsiuniversity.edu.my

C. H. Wong
Faculty of Science and Engineering, Maynooth University, Mariavilla, Maynooth, Co. Kildare, Ireland

K. H. Teng
Built Environment and Sustainable Institution (BEST), Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool L3 3A, UK

purposes are PZT (Lead Zirconate Titanate) and PVDF (Polyvinylidene Fluoride) [3]. For this paper, three types of PZT materials will be investigated in order to perform several simulation analyses to review the performances of both bimorph and unimorph cantilever transducer. The PZT materials used for this experiment are PZT-4, PZT-5A and PZT-5H. A cantilever-based transducer is an option for energy harvesting as it has a lower natural frequency and simpler in operation [4]. The frequency domain general study is chosen in order to obtain the graphical result of output voltage and electrical power at the same time to determine the natural frequency of the transducer with the step rate of 1 Hz. The purpose of this paper is to perform a comparison analysis to determine a more efficient energy harvester device via simulation works. It is so that the proposed design can serve as a reference for future research to be held.

2 Literature Review

This section aims to review and evaluate similar research journals which can be used as reference for this project. Table 1 shows the comparison on various research works on piezoelectric energy harvesting.

Bhuvana et al. [4] have conducted a simulation experiment using COMSOL in designing a cantilever-based piezoelectric vibration sensor. The design of the cantilever transducer has a dimension of $25 \times 13 \times 0.2$ mm. The piezoelectric material used for this experiment are PVF2, PZT-5H and ZnO. Based on their findings, it is deduced that the natural frequency of the cantilever transducer will be affected by the length, width and thickness dimension parameter. Jamain et al. [5] have performed a simulation analysis using COMSOL to review the performance of a MEMS piezoelectric energy harvester. The dimension of the bimorph cantilever transducer is $150 \times 50 \times 1$ μm with ZnO as the piezoelectric material. Based on the results, the highest generated output voltage was 9.91 V at the cantilever fixed end with a natural frequency of 0.171 MHz. Uddin et al. [6] has carried out a simulation experiment in COMSOL by designing a bimorph cantilever transducer for energy harvesting using mechanical ambient vibration. The dimension of the transducer was $12 \times 2.5 \times 0.03$ mm. PZT-5H was chosen as the piezoelectric material in this design.

Table 1 Overview of the literature review

Author	Year	Piezoelectric material	Natural frequency	Voltage Output	Electrical power output
[5]	2014	ZnO	0.17 MHz	9.91 V	–
[6]	2016	PZT-5H	345.75 Hz	595.5 mV	14.85 μW
[7]	2018	ZnO	101 Hz	8.5 V	3.61 μW
[8]	2017	PZT-5A	75.5 Hz	5.39 V	1 mW
Proposed work	2021	PZT-5H	114 Hz	14.66 V	%1.%2W

The final results indicate that the transducer was able to generate voltage output of 595.5 mV and 14.85 μ W of electrical power when the load resistance was 12.6 k Ω under 1 g acceleration. The value of natural frequency was 345.75 Hz. Chaudhuri et al. [7] have performed a simulation experiment using COMSOL by designing various cantilever structures for comparison purposes. ZnO is chosen as the piezoelectric material for this experiment. The results showed that a tapered perforated structure is the best as it produces an output voltage of 8.5 V with electrical power of 3.61 μ W where the natural frequency was recorded at 101 Hz. Dixit & Rangra [8] have used COMSOL to design a piezoelectric bimorph cantilever for electrical energy harvesting via ambient vibration. The chosen piezoelectric material was PZT-5A. When connected to a 12 k Ω load resistance, the transducer was able to generate the output voltage of 5.39 V and 1 mW of electrical power at a natural frequency of 75.5 Hz.

It was discovered that using PZT-5H or PZT-5A as the piezoelectric material is not as efficient as shown in Table 1. Therefore, the objective is to propose a better energy harvester design in generating a higher output voltage and electrical power by investigating on various parameters that would affect the output via simulation works.

3 Methodology

COMSOL Multiphysics 5.5 was used as the simulation software to run this experiment where Fig. 1 shows the geometry model for the cantilever transducer design with proof mass.

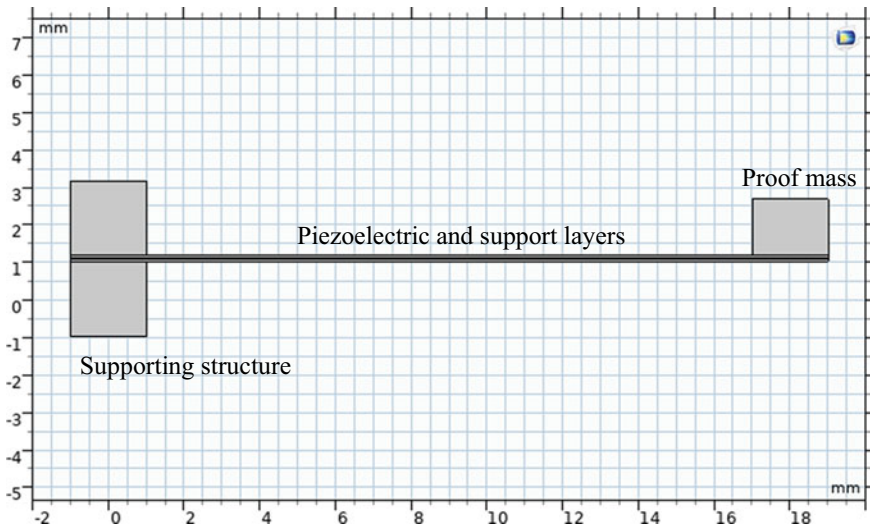


Fig. 1 Cantilever transducer design with proof mass

The designs are divided into domain entities in COMSOL. The bimorph transducer has six domains while the unimorph has five because the former has another extra piezoelectric layer beneath the supporting layer. For both bimorph and unimorph transducer, the domain entities can be classified into the piezoelectric layer, supporting layer, proof mass as well as the supporting structure as shown in Fig. 1. The proof mass was added in the design as it is able to increase the deflection at the free end of the cantilever beam, thus directly increasing the obtained output values [9]. Next, Fig. 2 shows the structural difference between bimorph and unimorph designed using COMSOL, where the blue layer is the piezoelectric PZT layer while the grey layer is the supporting layer. Prior to this step, several parameters have to be declared in order to run this simulation. The three parameters that have been declared are the acceleration of 1g, the load resistance of 100 kΩ and transducer width of 8 mm.

Both the piezoelectric transducer has a default dimension of length 20 mm, width 8 mm and thickness of 0.16 mm. The next step is to assign the materials to their respective domains. The materials used for this simulation are PZT-4, PZT-5A and PZT-5H for the piezoelectric layer and structural steel for the other domains. The material properties are as in Table 2 whereas the investigated parameters are as in Table 3. The purpose of altering the parameters as shown in Table 3 is to provide a more thorough analysis by investigating the transducer parameters in order to determine the better cantilever transducer with respect to the total harvestable voltage and electrical power output. Lastly, the solid mechanics, electrostatics, electrical circuit, piezoelectric effect and meshing physical interface were taken into account for the simulation process.

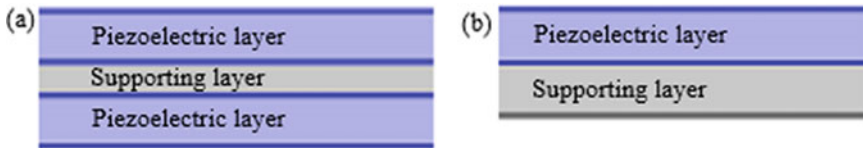


Fig. 2 Piezoelectric layers: **a** bimorph, **b** unimorph

Table 2 Material properties

Materials	Values			
	PZT-4	PZT-5A	PZT-5H	Structural steel
Density (kg/m ³)	7500	7750	7500	7850
Young’s modulus (GPa)	78	66	64	20
Poisson’s ratio	0.31	0.31	0.31	0.30
Piezoelectric coefficient (C/N)	289	374	593	–

Table 3 Investigated transducer parameters

Parameters	Bimorph	Unimorph
PZT material	PZT-4, PZT-5A, PZT-5H	PZT-4, PZT-5A, PZT-5H
Supporting material	Structural steel	Structural steel
Proof mass (g)	1, 3, 5	1, 3, 5
PZT length (mm)	20, 22, 24	20, 22, 24
PZT width (mm)	8, 10, 12	8, 10, 12
PZT thickness (mm)	0.16, 0.18, 0.20	0.16, 0.18, 0.20

4 Results and Discussion

4.1 Parametric Analysis on Various Proof Mass and Piezoelectric Material

Tables 4 and 5 shows the comparison results when the weight of proof mass and material parameters were used as a variable while the dimension of the transducer is constant. Based on the results for both bimorph and unimorph, when the material and weight of proof mass is set as PZT-5H and 1 g respectively, it is able to generate the highest output voltage and electrical power compared to the others. Figures 3 and 4 shows the graph obtained that was plotted in Microsoft Excel after the extraction of plot data from COMSOL Multiphysics 5.5.

Table 4 Results for bimorph when proof mass and material are varied

Material	Proof mass (g)	Natural frequency (Hz)	Voltage output (V)	Electrical power output (mW)
PZT-4	1	128	8.86	0.39
	3	124	8.60	0.37
	5	120	8.29	0.34
PZT-5A	1	114	9.07	0.41
	3	111	8.64	0.37
	5	107	8.48	0.36
PZT-5H	1	114	10.66	0.56
	3	111	10.59	0.56
	5	108	10.00	0.50

Table 5 Results for unimorph when proof mass and material are varied

Material	Proof mass (g)	Natural frequency (Hz)	Voltage output (V)	Electrical power output (mW)
PZT-4	1	148	8.15	0.33
	3	143	7.88	0.31
	5	139	7.78	0.30
PZT-5A	1	138	7.88	0.31
	3	134	7.66	0.29
	5	130	7.48	0.28
PZT-5H	1	139	9.00	0.40
	3	135	8.65	0.37
	5	131	8.43	0.35

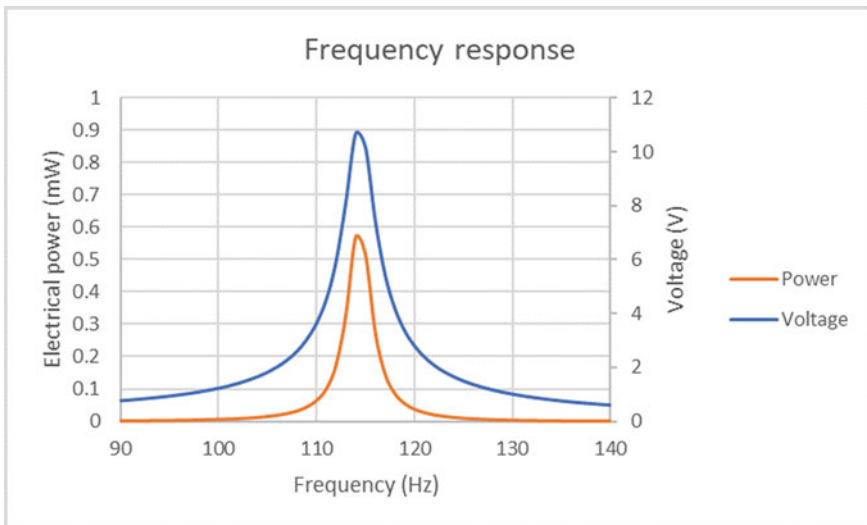


Fig. 3 Graph of bimorph when proof mass and material are varied

4.2 Parametric Analysis on Various Length, Width and Thickness Dimension

Tables 6 and 7 shows the comparison results when length, width and thickness dimensions were used as variable while the piezoelectric material and weight of proof mass of the transducer is constant. The piezoelectric material used is PZT-5H and the weight of proof mass is 1 g. Similarly, when dimensions of both transducers are set to length of 20 mm, width 12 mm and thickness 0.16 mm, the output values are a cut above the rest. Figures 5 and 6 shows the graph obtained that was plotted in Microsoft Excel after the extraction of plot data from COMSOL Multiphysics 5.5.

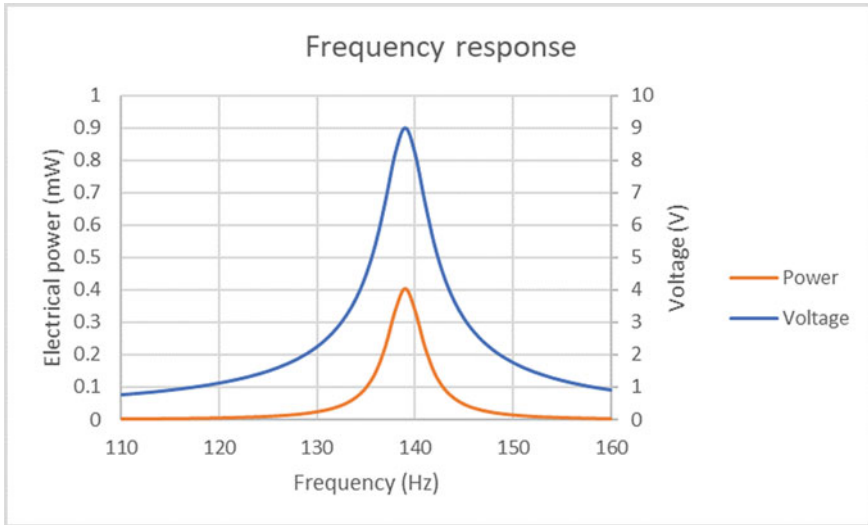


Fig. 4 Graph of unimorph when proof mass and material are varied

Table 6 Results for bimorph when thickness, length and width are varied

Length (mm)	Width (mm)	Thickness (mm)	Natural frequency (Hz)	Voltage output (V)	Electrical power output (mW)
20	8	0.16	114	10.66	0.56
		0.18	135	10.19	0.52
		0.20	155	9.96	0.49
20	8	0.16	114	10.66	0.56
22			96	11.74	0.69
24			82	12.54	0.78
20	8	0.16	114	10.66	0.56
	10		114	12.79	0.81
	12		114	14.66	1.07

The vertical axis on the right for all the graph output are plotted with respect to the obtained output voltage, meanwhile the vertical axis on left is for the electrical power output.

The main difference between both the transducer is that bimorph has two active piezoelectric layers sandwiching the support layer whereas a unimorph has only one active piezoelectric layer. As for Results 4.1, based on Table 4 and Table 5, for the bimorph transducer, the maximum voltage output and electrical power output was 10.66 V and 0.56 mW respectively as per Fig. 3, at a natural frequency of 114 Hz. Similarly, for unimorph, at a natural frequency of 139 Hz the maximum voltage output and electrical power output were 9.00 V and 0.40 mW, respectively as shown in Fig. 4. It can be observed that the natural frequency and the output

Table 7 Results for unimorph when thickness, length and width are varied

Length (mm)	Width (mm)	Thickness (mm)	Natural frequency (Hz)	Voltage output (V)	Electrical power output (mW)
20	8	0.16	139	9.00	0.40
		0.18	164	8.31	0.34
		0.20	189	8.04	0.32
20	8	0.16	139	9.00	0.40
22			117	9.55	0.45
24			100	10.05	0.50
20	8	0.16	139	9.00	0.40
	10		139	10.97	0.60
	12		139	12.81	0.82

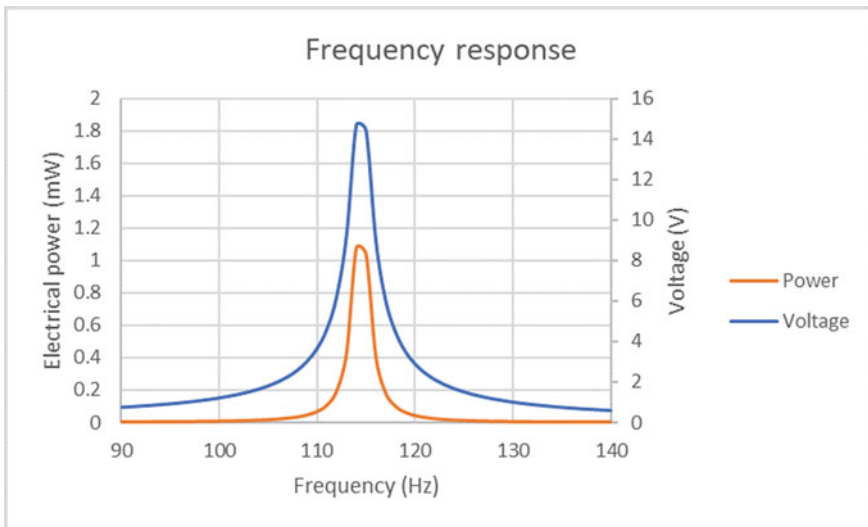


Fig. 5 Graph of bimorph when thickness, length and width are varied

values are inversely proportional to the weight of proof mass, regardless the piezo-electric material used. The possible explanation is that the damping effect due to the increasing weight of proof mass reduces the overall voltage and electrical power output. [10]. The obtained results from Sect. 4.1 indicates that PZT-5H is the better material for energy harvesting compared to PZT-4 and PZT-5A. This is because PZT-5H has a higher d_{33} piezoelectric coefficient as shown in Table 2, which holds the better advantage in generating a higher output voltage and electrical power [11]. Another factor that might affect the results is that the electrical power generated has an inversely proportional relationship to the damping ratio [12]. Thus, it is important to optimize the energy harvester design with suitable parameters.

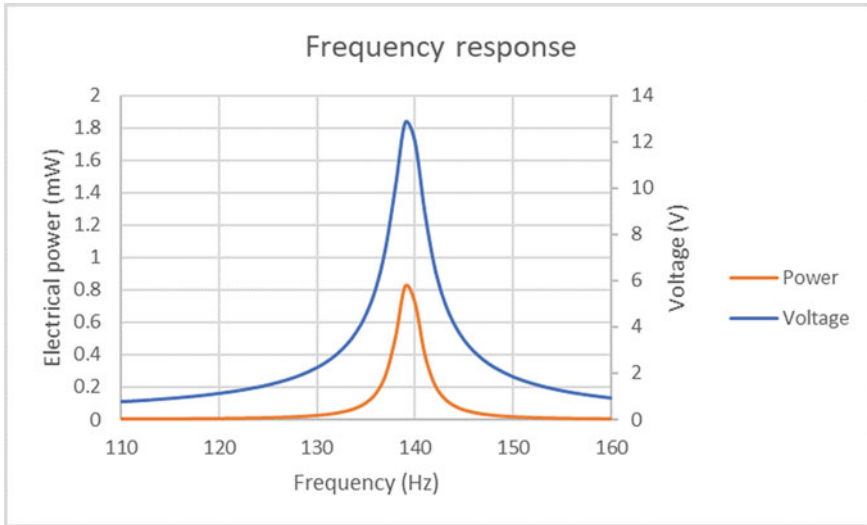


Fig. 6 Graph of unimorph when thickness, length and width are varied

Based on Tables 6 and 7, for the bimorph transducer, the maximum voltage output and electrical power output was 14.66 V and 1.07 mW respectively as per Fig. 5, at a natural frequency of 114 Hz. Similarly, for unimorph, at a natural frequency of 139 Hz the maximum voltage output and electrical power output were 12.81 V and 0.82 mW, respectively as shown in Fig. 6. The piezoelectric material PZT-5H and the 1 g proof mass is set to constant due to both the bimorph and unimorph transducer were able to generate the best results at default transducer dimension. It can be seen that the width of the transducer has lesser impact on the natural frequency. On the other hand, the natural frequency shows a directly proportional relationship to the thickness and an inversely proportional relationship to the length. As for the voltage and electrical power output, it is inversely proportional to the thickness and directly proportional to the length and width. The results obtained in terms of natural frequency is analogous to the research carried out by Bhuvana et al. [4].

5 Conclusion

In this paper, the performance of a bimorph and unimorph cantilever transducer was designed and carried out using COMSOL Multiphysics. By optimizing the type of piezoelectric material used, weight of proof mass as well as length, width and thickness, the obtained results were analyzed and discussed with support from various work of researchers. Hence, the bimorph cantilever transducer with a dimension of $22 \times 12 \times 0.16$ mm with PZT-5H and proof mass of 1 g at a natural frequency of 114 Hz shows the best result amongst all the other parameters that were investigated.

Lastly, it can be said that a bimorph cantilever transducer is more efficient for energy harvesting compared to unimorph. This paper can be served as a reference for future studies to be carried out. Future work will also be carried out in experimental format to verify the results obtained from this paper.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme under Proj-FRGS/1/2019/TK04/UCSI/02/1.

References

1. Lund H (2007) Renewable energy strategies for sustainable development. *Energy* 32(6):912–919
2. Wong C-H, Dahari Z, Abd Manaf A, Miskam MA (2014) Harvesting raindrop energy with piezoelectrics: a review. *J Electron Mater* 44(1):13–21
3. Anton SR, Sodano HA (2007) A review of power harvesting using piezoelectric materials (2003–2006). *Smart Mater Struct* 16(3):R1–R21
4. Bhuvana S, Prathiksha H, Sindhu V, Vasudha H (2018) Design and analysis of piezoelectric cantilever based vibration sensor. In: *IEEE international conference on system, computation, automation and networking, ICSCA, Pondicherry, India*
5. Jamain UM., Ibrahim NH, Ab Rahim R (2014) Performance analysis of zinc oxide piezoelectric MEMS energy harvester. In: *IEEE international conference on semiconductor electronics, proceedings, ICSE2014, Kuala Lumpur, Malaysia*
6. Uddin MN, Islam MS, Sampe J, Ali SHM, Bhuyan MS (2016) Design and simulation of piezoelectric cantilever beam based on mechanical vibration for energy harvesting application. In: *IEEE international conference on innovations in science, engineering and technology, ICISSET, Dhaka, Bangladesh*
7. Chaudhuri D, Kundu S, Chattoraj N (2019) Design and analysis of MEMS based piezoelectric energy harvester for machine monitoring application. *Microsyst Technol* 25(4):1437–1446
8. Dixit NK, Rangra KJ (2017) Vibration sensor: Piezoelectric power generator. In: *IEEE international conference on information, communication, instrumentation and control, ICICIC, Indore, India*
9. Wang L, Zhao L, Jiang Z, Luo G, Yang P, Han X, Li X (2019) Maeda R High accuracy comsol simulation method of bimorph cantilever for piezoelectric vibration energy harvesting. *AIP Adv* 9(9):095067
10. Ali W, Ibrahim S (2012) Power analysis for piezoelectric energy harvester. *Energy Power Eng* 4(6):496–505
11. Swallow LM, Luo JK, Siores E, Patel I, Dodds D (2008) A piezoelectric fibre composite based energy harvesting device for potential wearable applications. *Smart Mater Struct* 17(2):025017
12. Priya S (2007) Advances in energy harvesting using low profile piezoelectric transducers. *J Electroceram* 19(1):167–184

Simulated Kalman Filter with Modified Measurement, Substitution Mutation and Hamming Distance Calculation for Solving Traveling Salesman Problem



Suhazri Amrin Rahmad , Zuwairie Ibrahim , and Zulkifli Md Yusof 

Abstract Simulated Kalman filter (SKF) is a population-based optimization algorithm based on the Kalman filter framework. Each agent in SKF is treated as a Kalman filter. To find the global optimum, the SKF employs a Kalman filter mechanism that includes prediction, measurement, and estimate. However, the SKF can only operate in numerical search space. In literature, various techniques and modifications have been applied to numerical meta-heuristic algorithms in order to operate in a discrete search space. This paper presents modifications to measurement and estimation in SKF to fit the discrete search space. The modified algorithm is called Discrete Simulated Kalman Filter Optimizer (DSKFO). In solving the Travelling Salesman Problem (TSP), the DSKFO has been compared to four other SKF algorithms and it outperforms all of them.

Keywords Combinatorial · Simulated kalman filter · Travelling salesman problem

1 Introduction

Combinatorial optimization is a branch of optimization problems that has applications in a variety of fields. It is frequently utilized in a wide range of areas, including applied mathematics, artificial intelligence, computer science, and electronic engineering. An example of a combinatorial optimization problem is a Travelling Salesman Problem (TSP).

The travelling salesman problem (TSP) is the commonly known routing problem in combinatorial optimization. It has caught the interest of researchers since it is both simple to comprehend and challenging to solve. The TSP can be stated as follows: A salesman begins his or her journey in one city before moving on to the next set of

S. A. Rahmad (✉) · Z. Ibrahim
College of Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang,
Kuantan, Pahang, Malaysia
e-mail: mek18007@stdmail.ump.edu.my

Z. Md Yusof
Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia

cities. The objective of TSP is to determine the shortest and most cost-effective path travelled by the travelling salesman.

Over the past decades, many metaheuristic algorithms have been proposed originally to solve combinatorial problems, such as Genetic algorithm (GA) [1], Ant colony optimization (ACO) [2], and Simulated Annealing (SA) [3]. A numerical optimization algorithm is designed to operate in numerical search space. To operate in discrete search space, the algorithm must be modified or perform additional computations.

A numerical search space is a space of all feasible solution where the variables include a set of real numbers, whereas a discrete search space is a space of all feasible solution where the variables include a set of integers. For example, an objective function, $f(x_1, x_2)$, contains two variables, x_1 and x_2 , with an interval of $[0, 4]$. Thus, a feasible solution of the objective function in numerical search space is illustrated in Fig. 1(a), where the variables can be any set of real numbers. On the other hand, a feasible solution in discrete search space is shown in Fig. 1(b), where the variables consist of integers only.

The simulated Kalman filter (SKF) [4, 5] is an optimization algorithm originally introduced for numerical optimization problems. The SKF has been enhanced by adding a computation in order to solve combinatorial problems. Few existing combinatorial algorithm that are developed based on the SKF are binary SKF (BSKF) [6], distance evaluated SKF (DESKF) [7], angle modulated SKF (AMSKF) [8], and state encoded distance evaluated SKF (SEDESKF) [9].

In the proposed algorithm, the mutation and Hamming distance techniques, which are suitable for discrete search space, are employed. Mutation is a genetic operator that is used to create a new solution by making a random change in the string. In this paper, a substitution mutation utilized by [10] is used in the proposed algorithm. The mutation operates by replacing the value of the selected gene in the string with a value from another string. Figure 2 depicts how the mutation operates. In this figure,

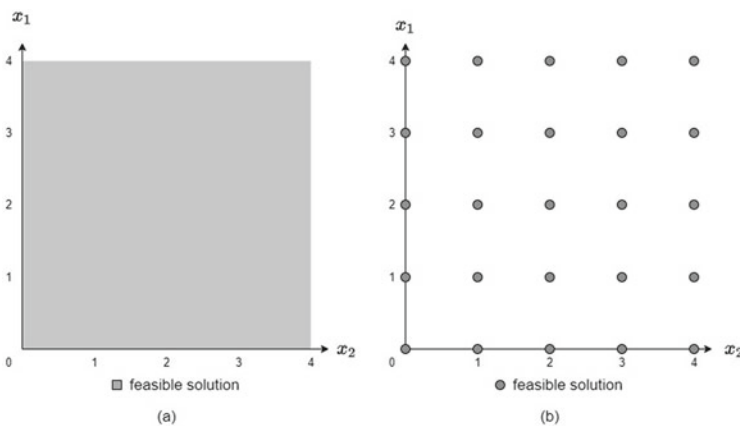


Fig. 1 Examples of numerical search space **a** and discrete search space **b**

Fig. 2 Illustration of the substitution mutation in the DSKFO algorithm

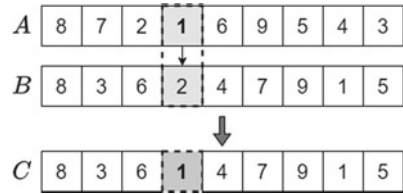
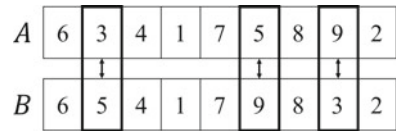


Fig. 3 Hamming Distance between string A and string B is 3



the chosen gene which is the fourth dimension of string B is replaced with a value of the same dimension from string A to create a new string C. The use of mutation in this algorithm encourages exploration at the beginning of the search process and promotes exploitation as the search process reaches its end.

The Hamming distance is a length metric used to calculate the difference between two comparable strings of equal length. Figure 3 depicts an illustration of measuring the Hamming distance between strings A and B. In this figure, three positions in these two strings are distinct. Hence, the Hamming distance between A and B, $H(A, B)$, is 3.

In this paper, the computations in the SKF are modified to allow the SKF algorithm to fit within a discrete search space. The performance of the modified algorithm is then compared with the existing combinatorial SKFs. The performance of the proposed algorithm is analyzed and discussed.

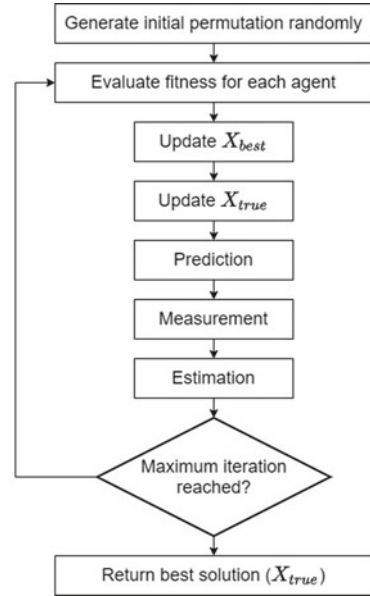
2 Methodology

2.1 Discrete Simulated Kalman Filter Optimizer (DSKFO)

The discrete simulated Kalman filter optimizer (DSKFO) algorithm is illustrated in Fig. 4. The algorithm begins by assigning a random sequence to N agents. The number of iterations is represented by t and the stopping condition for the algorithm is set at the maximum number of iterations, t_{max} . The initial value of the error covariance estimate, $P(0)$, the process noise, Q , and the measurement noise, R , all of which required for Kalman gain calculation are also set, where $\{P, Q, R\} \in \mathbb{R}$.

Each solution is comprised of D -dimensional vector. The first dimension, as well as all other dimensions in a solution are in the form of state, in which each state is subset of all states. Each state may consist of integer number ranging from 1 until D . In other word, a state vector, X , can be expressed as $X = \{x_1, x_2, \dots, x_D\}, \{x, D\} \in$

Fig. 4 Discrete simulated Kalman filter optimizer (DSKFO) algorithm



$\mathbb{Z}, x \in [1, D]$. The state vector, X , of the i^{th} agent at time t in the DSKFO is shown as

$$X_i(t) = \{x_i^1(t), x_i^2(t), \dots, x_i^d(t), \dots, x_i^D(t)\} \tag{1}$$

After that, each agent is put through an evaluation in order to determine their fitness value. The fitness values are compared, and the agent with the best fitness value at each iteration is set as $X_{best}(t)$. For minimization problem,

$$X_{best}(t) = \min_{i \in \{1, \dots, n\}} fit_i(X(t)) \tag{2}$$

and for maximization problem,

$$X_{best}(t) = \max_{i \in \{1, \dots, n\}} fit_i(X(t)) \tag{3}$$

The best solution obtained so far is called X_{true} . If a better fitness value is discovered, it will be taken as the value of X_{true} . The X_{true} is updated if the $X_{best}(t)$ is better than the X_{true} depending on the type of problem being evaluated ($X_{best}(t) < X_{true}$ for minimization problem, or $X_{best}(t) > X_{true}$ for maximization problem).

In the prediction stage, the time-update equations are computed as follows:

$$X_i(t|t + 1) = X_i(t) \tag{4}$$

$$P(t|t + 1) = P(t) + Q \tag{5}$$

where $X_i(t|t + 1)$ and $X_i(t)$ represent the predicted state and the current state, respectively, and $P(t|t + 1)$ and $P(t)$ are the predicted error covariant estimate and the current error covariant estimate, respectively.

The next step is measurement. In DSKFO, a substitution mutation utilized by [10] is used in the measurement step. The procedure for the measurement step is shown in Pseudocode 1.

Pseudocode 1. Procedure for Mutation in the Measurement Step

```

01: for each agent, i
02:     for each dimension, d
03:         generate rand with range of [0,1]
04:         if rand > 0.5
05:              $Z_i^d(t) = X_i^d(t|t + 1)$ 
06:         else
07:              $Z_i^d(t) = X_{true}^d$ 
08:         end
09:     end
10: end

```

The final step is estimation. A substitution mutation mechanism from [10] is also used in this step. The Kalman gain, $K(t)$, is computed as follows:

$$K(t) = \frac{P(t|t + 1)}{P(t|t + 1) + R} \tag{6}$$

The Kalman gain is a weight assigned to the measurements and the current state estimation. High gain places more weight towards the measurement and lower gain follows more closely to the prediction. The measurement residual is another element that influences the mutation during the estimate step. The measurement residual is the difference between the measurement, $Z_i(t)$, and the predicted state, $X_i(t|t + 1)$. In DSKFO, the Hamming distance is used to compute the difference between these two variables. The Hamming distance are then converted into a range of [0, 1] using following equation.

$$y_i(t) = \frac{H(Z_i(t), X_i(t|t + 1))}{D} \tag{7}$$

where $y_i(t)$ is the measurement residual as $y_i(t) \in [0, 1]$, and the Hamming distance between the measurement and predicted state is represented as $H(Z_i(t), X_i(t|t + 1))$. Next, the measurement residual, $y_i(t)$, is multiplied by the Kalman gain, $K(t)$, to produce a correction, $\hat{K}(t)$, for the predicted state. The $\hat{K}(t)$ value will determine the probability of mutation in each dimension. High $\hat{K}(t)$ value

leads more dimensions to take a value from measurement, $Z_i(t)$, whereas low $\hat{K}(t)$ value allows more dimension takes a value from predicted state, $X_i(t|t+1)$. A mutations are then occurred based on correction, $\hat{K}(t)$ as shown in Pseudocode 2 to produce the estimated states for following iteration, $X_i(t+1)$.

Pseudocode 2. Procedure for Mutation in the Estimation Step

```

01: for each agent,  $i$ 
02:   for each dimension,  $d$ 
03:     generate  $rand$  with range of  $[0, 1]$ 
04:     if  $rand > \hat{K}(t)$ 
05:        $X_i^d(t+1) = X_i^d(t|t+1)$ 
06:     else
07:        $X_i^d(t+1) = Z_i^d(t)$ 
08:     end
09:   end
10: end

```

In early iteration, the algorithm promotes exploration process as the mutation in estimation step occurs in many dimensions. As the iteration progresses, the Hamming distance between the measurement value and the predicted state decreases, lowering the chance of mutation in each dimension in the estimation step. The reduction of chosen dimensions for mutation in estimation step causes the algorithm to proceed to the exploitation process. After that, the estimated error covariance for the following iteration, $P(t+1)$ is computed as follows:

$$P(t+1) = (1 - K(t))P(t|t+1) \quad (8)$$

Finally, the procedures are performed for the following iteration until the maximum number of iterations is achieved.

3 Discrete Simulated Kalman Filter Optimizer (DSKFO) for Solving Travelling Salesman Problem (TSP)

A solution of the TSP is represented by a sequence of integer numbers. For example, there are ten cities in the TSP. In the DSKFO, a dimension represents a city, and a set of dimensions represents the travel sequence of the cities. A travel sequence of 1-2-3-4-5-6-7-8-9-10 can be represented by the estimated state as $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Since there are only ten cities involved, the number of dimensions is set at 10. More dimensions are required for a larger number of cities in the TSP.

After the sequence is generated for each agent, the solutions are evaluated by calculating the fitness value. In TSP instances, each city is given two coordinates to indicate its position. These coordinates are used to calculate the distance between

each city by using the Euclidean distance. Then, the fitness value, which is the total distance of the path taken, is calculated using the following equation.

$$fit(x_i(t)) = \sum_{d=1}^{D-1} \text{dist}(x_i^d, x_i^{d+1}) + \text{dist}(x_i^D, x_i^1) \quad (9)$$

where $\text{dist}(x_i^d, x_i^{d+1})$ represents the distance between estimated state in d^{th} dimension and $d + 1^{\text{th}}$ dimension.

4 Experiment, Result, and Discussion

A TSP benchmark set consists of 47 TSP instances that are used to evaluate the performance of the DSKFO. A list of the TSP instances as well as their ideal fitness are provided in Table 1. The TSP benchmark is divided into three instance sizes defined by [11], which are small-sized TSP, medium-sized TSP, and large-sized TSP.

The performance of the DSKFO is compared against four existing combinatorial simulated Kalman filter (SKF) algorithms, which are binary SKF (BSKF), distance evaluated SKF (DESKF), angle modulated SKF (AMSKF), and DESKF with state encoded (SEDESKF). Table 2 shows the parameters and values utilized for DSKFO, BSKF, DESKF, AMSKF, and SEDESKF. The parameters in the DSKFO are the same as those in other combinatorial SKFs. The quality of DSKFO findings is then assessed based on the overall fitness values of 50 trials in minimizing the total distance.

The comparison between the performance of DSKFO against the BSKF, DESKF, AMSKF, and SEDESKF is shown in Table 3, 4, and 5. Based on the findings, the DSKFO performs better than the BSKF, DESKF, AMSKF, and SEDESKF for solving small-sized, medium-sized, and large-sized instances in minimizing the distance of the TSP. The findings have proven that the DSKFO outperforms other combinatorial SKFs.

Table 1 List of TSP benchmark instances

Instance size	No	TSP instance	Ideal fitness
Small	1	berlin52	7542
	2	eil51	426
	3	st70	675
	4	eil76	538
	5	pr76	259,045
	6	rat99	8806
	7	kroA100	21,282
	8	kroB100	22,141
	9	kroC100	20,749
	10	kroD100	21,294
	11	kroE100	22,068
	12	rd100	7910
Medium	13	eil101	629
	14	lin105	14,379
	15	pr107	44,303
	16	pr124	59,030
	17	bier127	118,282
	18	ch130	6110
	19	pr136	96,772
	20	pr144	58,537
	21	ch150	6528
	22	kroA150	26,524
	23	pr152	73,682
	24	rat195	1211
	25	d198	35,002
	26	kroA200	29,368
	27	kroB200	29,437
Large	28	ts225	126,643
	29	pr226	80,369
	30	gil262	2378
	31	pr264	49,135
	32	pr299	48,191
	33	lin318	42,029
	34	pr439	107,217
	35	pcb442	56,892
	36	d493	48,912
	37	rat575	2323

(continued)

Table 1 (continued)

Instance size	No	TSP instance	Ideal fitness
	38	p654	34,643
	39	d657	50,801
	40	rat783	6773
	41	dsj1000	18,659,688
	42	pr1002	108,159
	43	pcb1173	50,778
	44	d1291	15,780
	45	r11304	252,948
	46	r11323	270,199
	47	r11889	316,536

Table 2 Experimental parameter settings

Parameter	DSKFO	BSKF	AMSKF	DESKF	SEDESKF
Iteration	1000	1000	1000	1000	1000
Number of agents	30	30	30	30	30
Number of trials	50	50	50	50	50
Initial error covariance estimate	1000	1000	1000	1000	1000
Process noise	0.5	0.5	0.5	0.5	0.5
Measurement noise	0.5	0.5	0.5	0.5	0.5

Table 3 Average performance comparison of DSKFO with combinatorial SKFs for solving small-sized TSP

Instance	SEDESKF	AMSKF	DESKF	BSKF	DSKFO
berlin52	22,406.98	22,874.86	22,932.2	22,847.64	19,033.89
eil51	1267.804	1266.809	1268.42	2127.613	1102.025
st70	2887.399	2902.092	2882.996	2890.875	2401.752
eil76	2043.157	2039.967	2052.86	23,782.28	1822.935
pr76	459,240.7	461,176	461,023.3	461,949.7	406,375.4
rat99	6632.39	6718.733	6696.176	6732.771	5939.013
kroA100	135,675.3	136,954.9	137,043	137,188.7	120,715.6
kroB100	133,147.7	134,818.2	134,923.4	134,786.5	128,634
kroC100	133,605.2	135,858.8	135,469.5	135,539.3	119,472
kroD100	130,181.9	131,561.2	131,622.3	131,396.8	118,427
kroE100	136,381.9	137,716.4	138,503.9	138,610.7	120,180.4
rd100	45,651.72	45,664.31	46,096.3	45,944.33	40,210.75

Table 4 Average performance comparison of DSKFO with combinatorial SKFs for solving medium-sized TSP

Instance	SEDESKF	AMSKF	DESKF	BSKF	DSKFO
eil101	2840.658	2856.43	2845.66	2853.754	2555,159
lin105	98,295.8	98,766.64	99,036.19	99,045.13	85,650.8
pr107	438,474.9	446,571.5	446,386.8	449,263.3	399,598.7
pr124	572,756.4	573,148.5	580,257.8	579,691.2	493,549.5
bier127	536,858.9	544,059.5	544,106.7	542,440	489,010.7
ch130	39,426.15	39,357.7	39,254.37	39,267	35,321.95
pr136	689,707.8	689,959.7	690,108.3	689,880.4	623,854.7
pr144	679,453.9	686,191.1	682,605.3	682,410.8	591,661.7
ch150	46,136.63	46,168.05	46,270.79	46,174.03	41,839.67
kroA150	214,278.5	215,813.7	216,442.1	215,796.9	194,157
pr152	880,010	886,369	886,217.2	886,457.3	777,614.8
rat195	19,329.74	19,441.65	19,422.96	19,461.39	17,846.2
d198	143,879.8	158,018.6	157,618.5	158,476.9	136,081.7
kroA200	289,059.9	291,098.8	291,940.4	291,063.8	272,885.1
kroB200	283,920	285,558.9	285,802.7	286,095.5	248,305.7

The DSKFO differs from the four existing combinatorial SKF algorithms in that the four algorithms do additional computation for the SKF to solve the TSP, whereas the DSKFO does not. Although no additional computation is required in the DSKFO, the SKF is modified by providing a new interpretation for solving the TSP via mutation and Hamming distance. The findings of the algorithms indicate that the use of techniques appropriate for discrete search spaces, such as mutation and Hamming distance, may be one of the reasons for the DSKFO's superior performance.

Table 5 Average performance comparison of DSKFO with combinatorial SKFs for solving large-sized TSP

Instance	SEDESKF	AMSKF	DESKF	BSKF	DSKFO
ts225	1,409,269	1,410,333	1,411,955	1,409,169	1,285,080
pr226	1,472,488	1,477,167	1,479,082	1,482,490	1,319,486
gil262	23,718.7	23,851.59	23,846.46	23,853.9	21,832.33
pr264	945,954.9	954,069.8	954,199	958,776.8	881,561.7
pr299	663,592.6	667,263.2	664,537	666,494.6	616,103.6
lin318	527,431.1	528,817.1	527,049.5	529,112.7	497,228.2
pr439	1,714,496	1,732,577	1,737,005	1,731,523	1,623,659
pcb442	706,661.8	707,728.3	708,486.4	708,016.9	671,691.3
d493	405,901.5	411,931.2	411,998.9	411,621	380,277.4
rat575	103,555.9	104,311.9	103,909.3	104,248	99,569.35
p654	1,835,950	1,848,103	1,845,492	1,849,637	1,722,673
d657	794,117	796,174.9	796,175.3	796,929.4	757,082
rat783	165,840	167,018.8	166,512.9	166,983	158,673.6
dsj1000	5.20E + 08	5.23E + 08	5.24E + 08	5.24E + 08	5.02E + 08
pr1002	6,070,399	6,078,577	6,085,013	6,079,543	5,854,900
pcb1173	1,332,669	1,335,124	1,333,055	1,335,923	1,309,015
d1291	1,640,368	1,646,428	1,645,013	1,648,227	1,597,967
rl1304	8,880,536	8,908,134	8,917,743	8,916,299	8,625,869
rl1323	9,275,698	9,303,447	9,303,794	9,302,486	9,050,985
rl1889	14,114,830	14,159,546	14,171,974	14,157,634	13,702,898

5 Conclusion

This paper presents a modified Simulated Kalman Filter (SKF) algorithm for solving combinatorial optimization problem. Since the search space is in discrete, mutation and Hamming distance are applied in proposed algorithm. The proposed DSKFO algorithm is then employed to solve the TSP, which is a combinatorial optimization problem. After that, the proposed algorithm is compared with four combinatorial SKFs. According to the findings, the proposed DSKFO outperforms the BSKF, DESKF, AMSKF, and SEDESKF for solving the TSP. Further research of the proposed DSKFO can be considered for future studies.

Acknowledgements The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2018/TK04/UMP/02/9 (University reference RDU190176).

References

1. Holland JH (1992) Genetic algorithms. *Sci Am* 267:66–72
2. Colormi A, Dorigo M, Maniezzo V (1992) An investigation of some properties of an ‘Ant algorithm.’ *PPSN* 92:509–520
3. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680
4. Ibrahim Z, Abdul Aziz NHA, Nor NA, Razali S, Mohamad MS (2016) Simulated kalman filter: a novel estimation-based metaheuristic optimization algorithm. *Adv Sci Lett* 22(10):2941–2946. <https://doi.org/10.1166/asl.2016.7083>
5. Ibrahim Z et al (2015) A Kalman filter approach for solving unimodal optimization problems. *ICIC Express Lett* 9(12):3415–3422
6. Yusof ZM, Ibrahim I, Satiman SN, Ibrahim Z, Aziz NHA, Aziz NAA (2016) BSKF: simulated kalman filter. In: *Proceedings - AIMS 2015 3rd international conference on artificial intelligence modelling simulation*, pp 77–81. <https://doi.org/10.1109/AIMS.2015.23>.
7. Yusof ZM et al (2016) Distance evaluated simulated kalman filter for combinatorial optimization problems. *ARNP J Eng Appl Sci* 11(7):4911–4916
8. Yusof ZM et al (2016) Angle modulated simulated kalman filter algorithm for combinatorial optimization problems. *ARNP J Eng Appl Sci* 11(7):4854–4859
9. Yusof ZM et al (2018) Distance evaluated simulated kalman filter with state encoding for combinatorial optimization problems. *Int J Eng Technol* 7(4):22–29. <https://doi.org/10.14419/ijet.v7i4.27.22431>
10. Ab Rahman T, Ibrahim Z, Ab Aziz NA, Zhao S, Abdul Aziz NH (2018) Single-agent finite impulse response optimizer for numerical optimization problems. *IEEE Access* 6(c):9358–9374. <https://doi.org/10.1109/ACCESS.2017.2777894>
11. Halim AH, Ismail I (2019) Combinatorial optimization: comparison of heuristic algorithms in travelling salesman problem. *Arch Comput Methods Eng* 26(2):367–380. <https://doi.org/10.1007/s11831-017-9247-y>

Development of Heart Rate Sensor Warning System to Estimate driver's Cognitive Distraction Level



Akmal Shahmi Bin Azhar and Ahmad Khushairy Bin Makhtar

Abstract Distraction is a task that mentally demands and prevents focusing on the primary task, such as talking to the phone. This kind of distraction was measured using heart rate because the mental condition of the driver had initiated the autonomic nervous system through the sympathetic and parasympathetic nervous system in the activation of the increase or decrease of heart rate. The percentage of road accidents in Malaysia contributed by drivers to be more caused explicitly by cognitive distraction, such as talking to the passengers and talking to the phone while driving, which is the cause of death in Malaysia. The purpose of this study is to evaluate the relationship between heart rate and the cognitive distraction level of drivers that had influenced their driving performance. The research started by developing the heart rate sensor warning system that provided warning alarms to the respondent once they were distracted. Eight participants (Mean = 24.25, SD = 0.463) participated in the experiment and set up a pulse sensor on a driving simulator. The respondent has been given the N-Back Task while driving and evaluated by NASA Task Load Index. The result showed that the heart rate and driving performance affected when the driver was cognitively distracted. The research also revealed that drivers need to put more effort towards driving to maintain good driving performance when they were cognitively distracted. Lastly, distraction while driving has been contributed to accidents. Thus, this device was successfully alerted the driver and increased their driving performance when they cognitively distracted. This developed device could be used to reduce the number of traffic accidents in Malaysia due to distraction while driving.

Keywords Heart rate · Secondary tasks · Warning system

A. S. B. Azhar · A. K. B. Makhtar (✉)

School of Mechanical Engineering, Universiti Teknologi Mara, Shah Alam, Malaysia

e-mail: ahmadkhushairy@uitm.edu.my

1 Introduction

Malaysia Road Safety Department released the report that stated that from 2017 until June 2019, road accidents recorded in 2017 were 533,875 cases and gradually increased in 2018 with 548,598 cases. The recent report from Malaysia Road safety Department (2019) shows that until June 2019, road accidents occurred where 281,527 have been recorded. In contrast, Ministry of Transport Malaysia stated that through the report from Malaysian Transport Statistics (2018), from 2009 road accidents recorded were 397,330 with 6,745 total deaths. Road accidents cases drastically increase in 2018 with 548,598 cases with 6,284 total deaths [1].

After reviewing a large number of previous studies on the causes of road accidents, it was discovered that distraction caused by a secondary task was accounting for roughly 23% of all crashes and near-crashes (task other than driving) [2]. In general, driver's perception response time increases when the drivers are cognitively distracted because of inadequate cognitive control resources. For example, when drivers had conversations using mobile phones, their perception response time increased in vehicle braking scenarios [3]. Furthermore, the study proved that driver heart rate increases under high cognitive demands by tested it using driver monitoring applications during driving sessions simultaneously to secondary cognitive tasks [4].

Most vehicle manufacturers, especially car manufacturers, depend only on such safety measures like airbags, breaking mechanisms, and Intelligent Speed Assist (ISA), but few are concerned about driver's condition as one of the vital safety features of their car. A challenging part of current research is detecting a driver's cognitive distraction, but it has become increasingly important to develop a driver-precautionary warning system. Different methods and devices have been used in the studies to estimate a driver's cognitive distraction. It is however, unclear how accurate the device measurements will be going forward.

Several mental conditions compose the human mind. These mental conditions are generally widely referred to as emotion (anxiety, pain, love). Cognitions and perceptions also included in the mental condition [5]. The driver's heart rate may determine a driver's cognitive distraction. The autonomic nervous system (ANS) is the component of the peripheral nervous system that acts as a control mechanism for maintaining the body under stable conditions. ANS has two main components: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). These two components will innervate the sinoatrial node to the heart, which will activate the increasing heart rate while PNS activation will decrease the heart rate [6]. Observing the changes in any of these systems could be used as a guide to determine the level of cognitive distraction experienced by the individual. This process represents the conceptualization of how physiological measurements actually work. Examples of physiological measurements include electrocardiography (ECG), electromyography (EMG), electro-dermal activity (EDA), respiratory (RSP), skin temperature (SKT), pupil diameter (PD), eye activity, and speech recordings [7, 11, 12].

Subjective assessments of mental workload could be conducted by asking participants to rate their perceived workload while performing a task. Some subjective

measurement techniques can be performed, such as the NASA-TLX (NASA-Task Load Index) and the Subjective Workload SWAT-Assessment Technique. NASA-TLX is a widely used for subjective measurement instrument that assesses six factors: mental demands, physical demands, temporal demands, performance, and effort and frustration levels. The primary reason for choosing the NASA TLX method is its ease of implementation, and its effectiveness is primarily related to workload evaluation over extended periods. However, Hoedemaeker affirmed that SWAT is not as sensitive or valid as NASA-TLX [8]. Lastly, this research aims to develop a heart rate sensor warning system that can detect a driver's heart rate while driving. The IBM SPSS statistical software was used to analyze physiological, performance, subjective measurement and device performance to meet the research objective.

2 Methodology

2.1 Participants

Eight respondents have participated in this experiment consists of 4 males and 4 females. Their age was between 23–25 years old (Mean = 24.25, SD = 0.463). Every respondent had a valid driving's license and they drove almost daily. All the respondents had driving experience between 1–8 years. With the balance number of respondents, the results were enough to provide convincing results.

2.2 Heart Rate Warning System Device

The main component for this heart rate warning system is the pulse sensor. The pulse sensor detecting heart rate based on the principle of infrared light travelling through the blood to determine the blood pressure and analyze the heart rate. To begin, the sensor is attached to the human fingertip. Blood is circulated to the fingertip when the infrared sensor light passes through the blood to the photodiode, which measures the blood pressure. This measured value is then sent to the Arduino controller. The controller compares the detected value to the threshold value to determine whether there is a difference in the output. Finally, the controller sends the signal that triggers the buzzer to produce an alarm and the Arduino controller displays the sensor output value on the LCD display. The heart rate sensor comprises a photodiode and an infrared sensor. The IR sensor is directed toward the finger on one side in Fig. 1, while the photodiode receives the signal and measures the pulse and blood count.

The blood intensity decreases and increases in proportion to the heart rate, making it easy to determine if the heart is normal or abnormal. Figure 2 shows the flowchart that summarizes the development of the heart rate sensor warning system. The sensor

Fig. 1 Components of the pulse sensor [13]

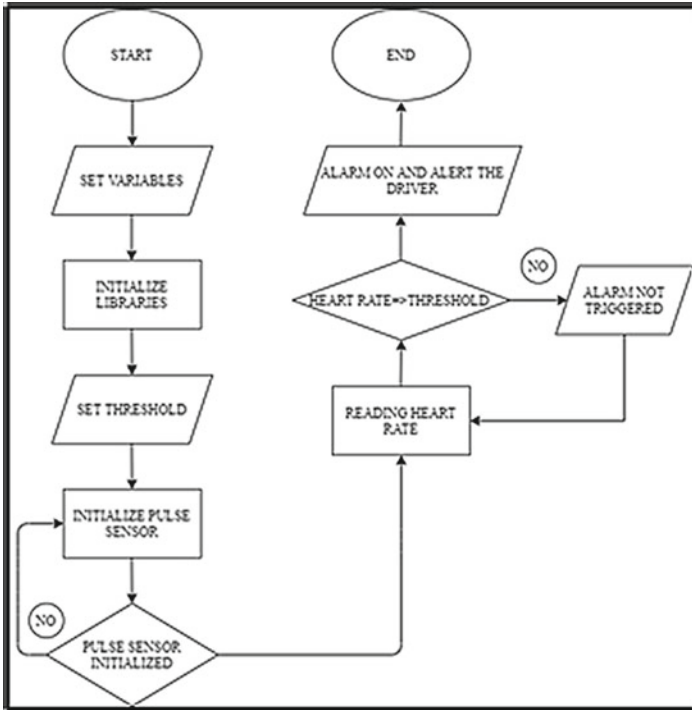
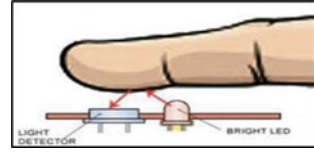


Fig. 2 Arduino flowchart

measurement value is translated to the voltage variation due to the op-amp function, and the controller received the output value in DC voltage from the sensor.

2.3 Experiment Setup

This experiment was conducted at Ergonomic Laboratory, School of Mechanical Engineering, Universiti Teknologi Mara (UiTM). Figure 3 shows a driving simulator been used in the experiment.

Table 1 shows the list of the components and the functions respectively. The pulse sensor and the Arduino UNO are the main components to develop the heart rate warning system device.

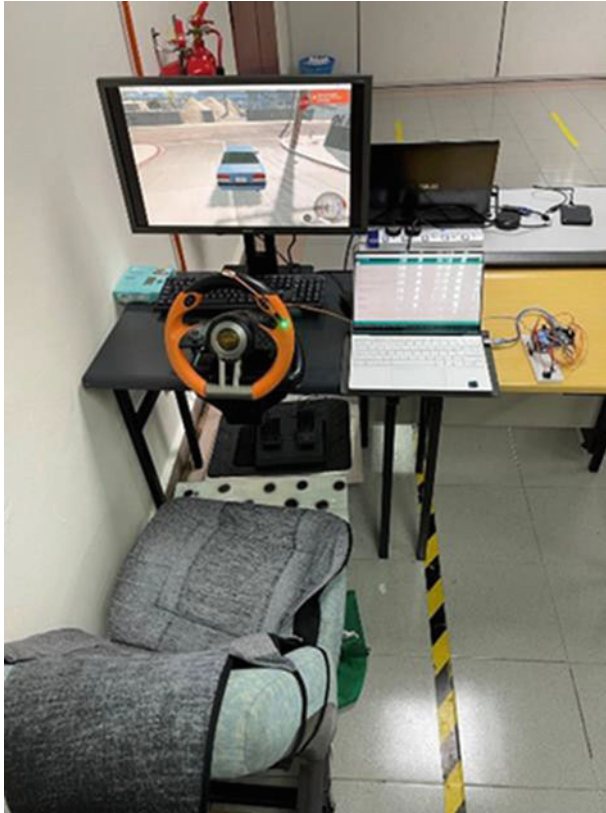









Fig. 3 Driving simulator set-up

The respondent's blood pressure, temperature, and initial pulse rate were recorded to make sure they were in good condition so that the result of the experiment would not be affected. Then, the respondent was briefly explained about the objective and the procedure of this experiment. Next, the respondent needs to declare their personal back- ground like age, height, weight, health condition and driving experience in the Google Form. The consent form was given to the respondent to sign after they agreed to participate in this experiment and understood that all of the data taken was for research purposes only. At the end of the experiment, the respondent needed to answer the feed- back form for the device's performance throughout the whole experiment.

Table 1 Components of device

Component	Function
Bread Board 	As a medium to assemble all of the components of a system for heart rate sensor monitoring system
Jumper Wire 	Connecting wire for connect all the components to enable current going through
Pulse Sensor 	To detect the pulse from the user
Arduino UNO 	As the open-source medium to read the input and turn it to the output by using the programming
LCD Display 	Will the display the heart rate reading of the user
Potentiometer 	Use to control the current that going through the LCD display
Buzzer 	Produce the alarm when the heart rate exceeds the threshold

2.4 Experiment Procedure

There are two tasks that each driver needs to be done. For the main task, the driver needs to focus and drive safely and for the secondary task driver will be given the N- Back Task while driving and answer it correctly. All drivers will be doing four sessions of driving test. Each session consists of 3 repeated tests. The first session driver will be given with 0-Back Task, the second session will be 1-Back Task and the third session will be 2-Back Task. The last session will be 2-Back Task that will be conducted with alarm device to alert the driver when their heart rate exceeds the threshold that will be set. In the 1-Back Task driving session, driver's heart rate will be taken to set the heart rate threshold for the last session. The heart rate threshold for each driver will be set by taking the average heart rate in this session. The duration of each test is 6 min. First two minutes, the driver just needs to drive normally. Then, for the next 2 min driver needs to drive and correctly answer the N-Back Task given. After that, for the last 2 min driver need to drive normally and complete the test. After the respondent's finished each test in each session, they will need to answer the NASA Task Load Index immediately to rate their driving performance. Finally, after all the sessions have been done, the driver needs to answer the feedback form regarding the alarm device used in the last session.

2.5 Statistical Analysis

IBM SPSS Statistic was used to analyse all the respondents' data after being sorted out in Microsoft Excel. IBM SPSS Statistic analyzed the data to reveal the significant result in each measurement. Repeated measure ANOVA and univariate measure ANOVA were used to analyze the physiological, performance and subjective measurement (Table 2) as the software will determine the significance for the data that have been entered is less than ($p < 0.05$). The significance level or alpha level of 0.5 (5%) has been selected to run the test of all results in this experiment. Repeated measure ANOVA was used to determine the statistical significance in all gender results and univariate measure ANOVA was used to determine the statistical significance in the overall result of eight respondents.

Table 2 List of dependent variables

Measurement groups	Dependent variables
Physiological	Heart rate reading using pulse sensor
Performance	N-Back task result
Subjective	Nasa task load index and device Feedback form

3 Result and Discussion

3.1 Physiological Measurement

A repeated measure ANOVA was applied with the heart rate of the first, second, third and fourth sessions as the within-subjects factor. Besides, gender was declared as the between-subjects factor. The average value of heart rate of an experiment for two genders was shown in Fig. 4. Repeated measure ANOVA analysis showed a significant heart rate value for each session, Wilks' Lambda = 0.99, $F(3, 2876) = 14.18$, $p < 0.001$, $\eta^2 = 0.012$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. The female average heart rate results between (M:81) from first session until third session. The male average heart rate value was (M:72) to (M:78) in the third session. Significant drop of heart rate, male (M:72) and female (M:77) in the fourth session. The female was more nervous when handling secondary tasks while driving because gender can significantly affect stress levels, lifestyles and all aspects of human behavior [9].

A univariate ANOVA was applied with the heart rate of the first, second, third and fourth sessions as the dependent variable. Besides, four tasks were declared as the fixed factor. The average value of the heart rate of an experiment for the entire session was showed in Fig. 5. Univariate ANOVA analysis showed a significant heart rate value for each session, $F(3, 11,517) = 47.96$, $p < 0.001$, $\eta^2 = 0.012$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. First session (M:76) and reached (M:80) in the third session. Significant drop in fourth session (M:74). Both added cognitive and the driving task compete

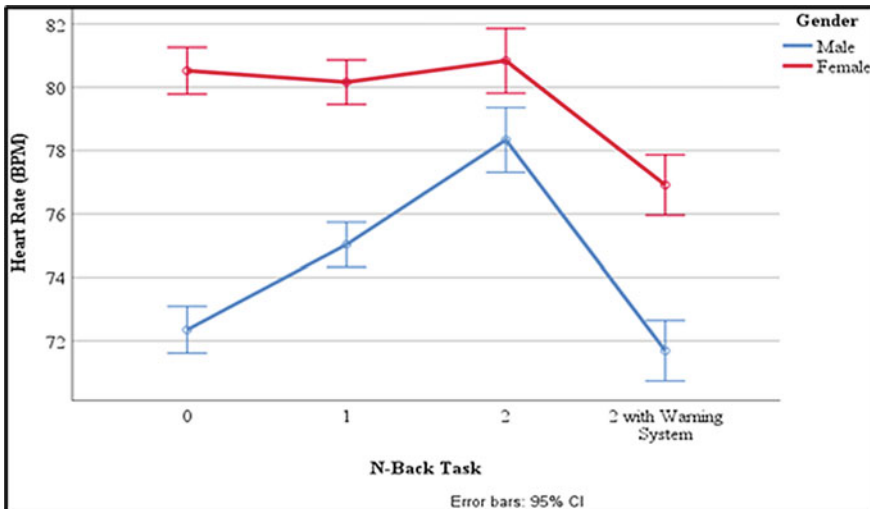


Fig. 4 Line graph average heart rate for gender

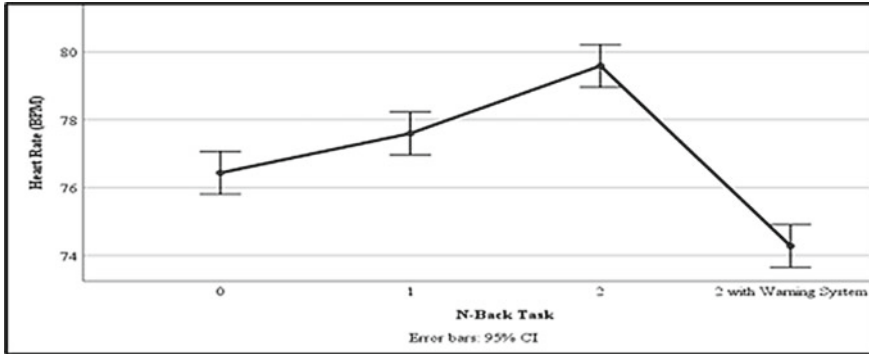


Fig. 5 Line graph average heart rate for entire session

for working memory components. Significant difference between session four and session three because of warning system device. Overall, the pairwise comparison revealed the significant difference between the average heart rate value and other tasks that has been proved by using the Bonferroni method.

3.2 Performance Measurement

A repeated measure ANOVA was applied with the driver performance of the first, second, third and fourth sessions as the within-subjects factor. Besides, four tasks were declared as the between-subjects factor. The average value of driver performance of an experiment for two genders were shown in Fig. 6. The repeated measure ANOVA analysis showed no significant driver performance value for each task,

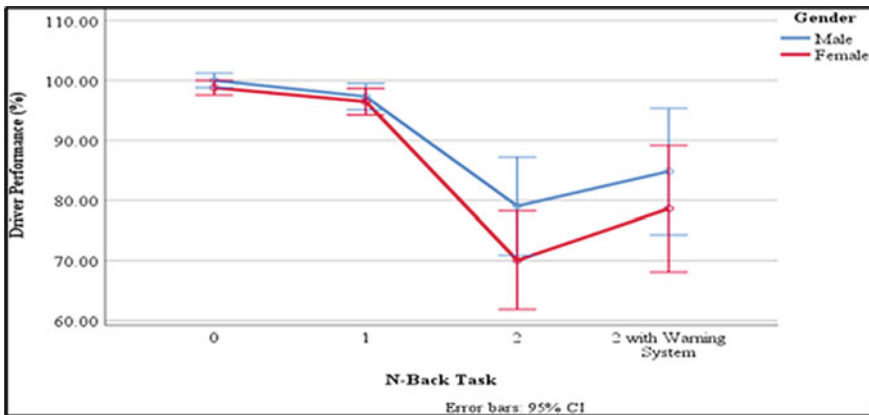


Fig. 6 Line graph average driver performance for gender

Table 3 Pairwise comparison by gender

(I) Gender	(J) Gender	Mean difference (I-J)	Std. error	Sig	95% Confidence interval for difference	
					Lower bound	Upper bound
Male	Female	4.307	2.826	0.142	-1.553	10.167
Female	Male	-4.307	2.826	0.142	-10.167	1.553

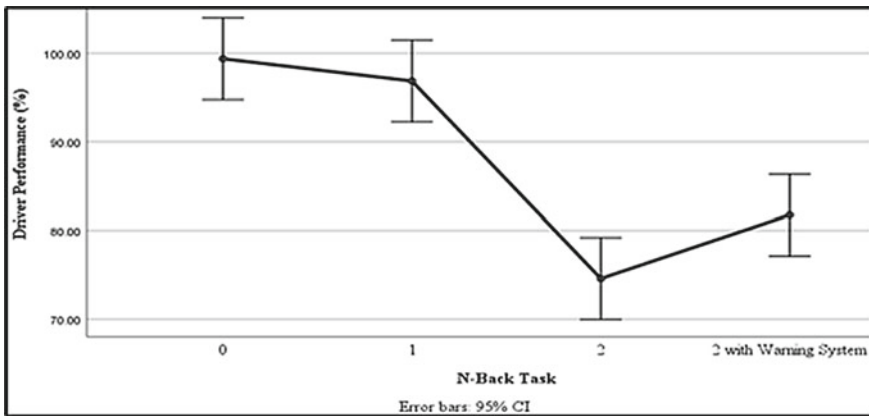


Fig. 7 Line graph average driver performance entire session

Wilks’ Lambda = 0.91, $F(3, 20) = 0.70$, $p > 0.05$, $\eta^2 = 0.095$. Overall, the pairwise comparison revealed no significant difference between the gender that has been proved by using the Bonferroni method as stated in Table 3.

A univariate ANOVA was applied with the driver performance of the first, second, third and fourth sessions as the dependent variable. Besides, four tasks were declared as the fixed factor. The average value of driver performance of an experiment for entire sessions was shown in Fig. 7. The univariate ANOVA analysis showed a significant driver performance value for each session, $F(3, 92) = 26.28$, $p < 0.001$, $\eta^2 = 0.461$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. First session (M:99.39) to the third session (M:74.57). Significantly increased in fourth session (M:81.73). The driver’s performance significantly improves when a warning system device is used in session four. When drivers attempt to maintain their composure and focus, their driving performance will improve. Overall, the pairwise comparison revealed no significant difference between the average driving performance value for session one and session two ($p = 0.447$), because both sessions had a quite similar difficulties level of tasks that drivers can withstand. The pairwise comparison revealed a significant difference besides first session and second session using the Bonferroni method.

3.3 Subjective Measurement

A repeated measure ANOVA was applied with the NASA TLX of the first, second, third and fourth sessions as the within-subjects factor. Besides, gender was declared as the between-subjects factor. The average value of NASA TLX of an experiment for two genders was shown in Fig. 8. The repeated measure ANOVA analysis showed a significant NASA TLX value for each session, Wilks' Lambda = 0.47, $F(3, 20) = 7.97$, $p < 0.05$, $\eta^2 = 0.544$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. First session for males was (M:12.99, increase to (M:74.31) in third session. First session for females was (M:5.35), increased to (M:40.49) in the third session. The males need more effort to maintain their driving performance because males were more aggressive while driving than females [10]. The pairwise comparison revealed a significant difference ($p = 0.00$) between the gender that has been proved by using the Bonferroni method. No significant difference between effort in the fourth session with third session because both tasks had the same level of difficulties.

A univariate ANOVA was applied with the NASA TLX of the first, second, third and fourth sessions as the dependent variable. Besides, four tasks were declared as the fixed factor. The average value of NASA TLX of an experiment for overall sessions was shown in Fig. 9. The repeated measure ANOVA analysis showed a significant NASA TLX value for each session, $F(3, 92) = 32.95$, $p < 0.00$, $\eta^2 = 0.518$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. The first session was (M:9.17) to (M:57.40) in the third session. Cognitive distraction needs more effort from the driver to maintain their performance during driving. Overall, the pairwise comparison revealed a significant difference that has been proved by using the Bonferroni method. No significant difference between effort in the fourth session and third session because both tasks had the same level of difficulties.

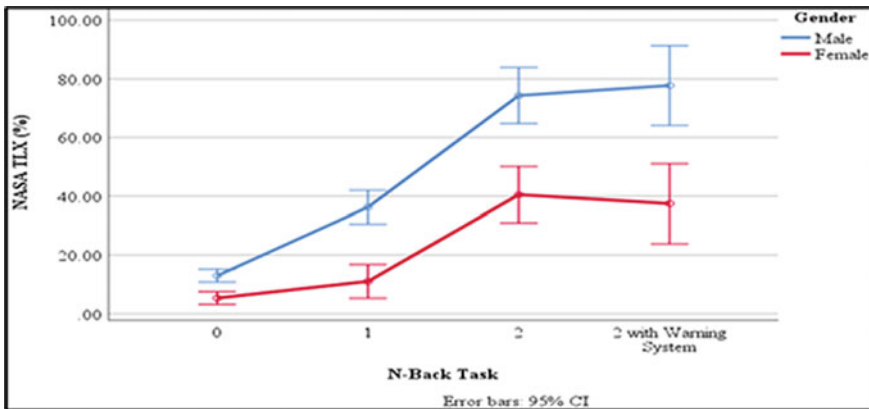


Fig. 8 Line graph average NASA TLX for gender

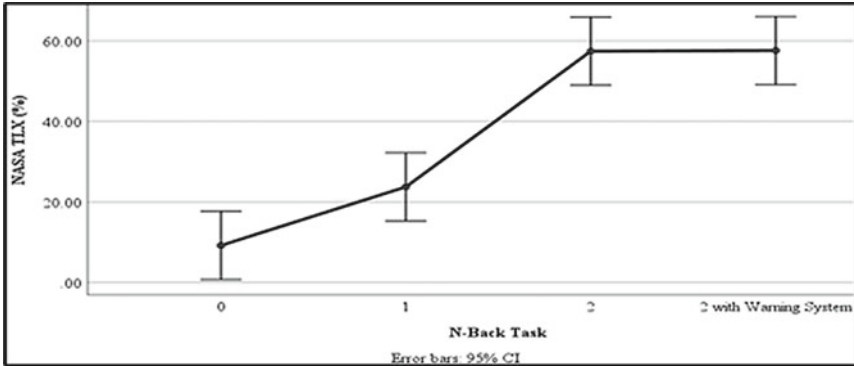


Fig. 9 Line graph average NASA TLX for entire session

3.4 Performance of Device

Figure 10 above showed the average responses to the questionnaires during the experiment. Repeated measures ANOVA was performed to analyze the questionnaire as a within-subjects factor. The repeated measure ANOVA analysis showed a significant value for the questionnaire answer, Wilks' Lambda = 0.13, $F(4, 4) = 6.49$, $p < 0.049$, $\eta^2 = 0.867$. Hence, the null hypothesis is rejected in this case, considering the significance level of $\alpha = 0.05$. Additionally, there was a significant difference between questionnaire answers that has been proved by using the Bonferroni method. The average value for response time for the alarm was (M:4), respondents agreed that the

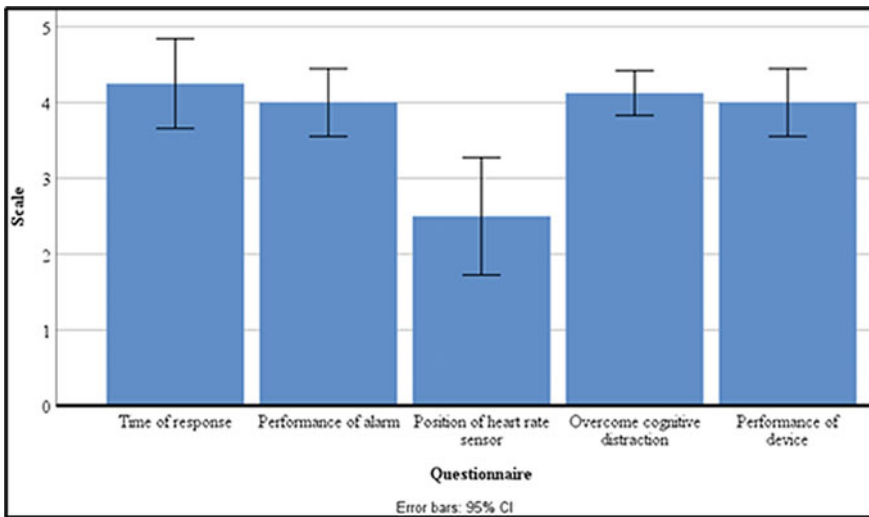


Fig. 10 Bar chart for the device feedback performance

warning system quickly warned them. Respondents agreed that the alarm sound did not annoy and not too loud as the average value for performance of alarm was (M:4). The average value for the position of heart rate given was (M:2), because respondents dissatisfied only one heart rate sensor was provided. The average value given was (M:4) because respondents agreed the device allows them to focus and maintain their performance when distracted. Lastly, all respondents agreed with overall device performance as the average value showed (M:4). Device performance can be justified by the warning alarm that is only triggered when the heart rate exceeds the threshold that has been set.

4 Conclusion and Recommendations

The development of a heart rate sensor warning system was completed successfully. The physiological and performance measurement analysis was obtained by using the device in this experiment. Physiological measurement indicated the statistical significance for gender and overall result. Next, performance measurement indicated no statistical significance in gender but significance in overall result. Then, subjective measurement indicated that both gender and the overall result were significant. Some improvements can be highlighted in this research. The application of the pulse sensor should be increased to prevent fatigue to the driver because they need to put their finger all the time on the pulse sensor. Lastly, the number of respondents should be increased to 20–30 respondents to get more accurate and precise data following the various re- search experiment.

References

1. Naim HM (2020) Kadar kemalangan jalan raya pada musim perayaan di malaysia dalam tempoh 10 tahun (2010–2019)
2. Hurts K, Angell LS, Perez MA (2011) The distracted driver: mechanisms, models, and measurement. *Rev Hum Factors Ergon* 7(1):3–57. <https://doi.org/10.1177/1557234X11410387>
3. Heart rate: What is a normal heart rate? <https://www.medicalnewstoday.com/articles/235710>, Accessed 20 July 2021
4. Hashem MMA, Shams R, Kader MA, Sayed MA (2010) Design and development of a heart rate measuring device using fingertip. In: International conference on computer communication engineering, ICCCE 2010, no April. <https://doi.org/10.1109/ICCCE.2010.5556841>.
5. Khan MQ, Lee S (2019) A comprehensive survey of driving monitoring and assistance systems. *Sensors (Switzerland)* 19(11). <https://doi.org/10.3390/s19112574>
6. Doctoral P, Diversity E (2010) A review of postwar. *Rugoscopy Sci Palatal Rugae* 1(4):1–31
7. Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M (2022) Review on psychological stress detection using biosignals. *IEEE Trans Affect Comput* 13(1):440–460. <https://doi.org/10.1109/TAFFC.2019.2927337>
8. da Silva FP (2014) Mental workload, task demand and driving performance: what relation? *Procedia Soc Behav Sci* 162:310–319. <https://doi.org/10.1016/j.sbspro.2014.12.212>

9. Soares S, Monteiro T, Cunha L (2020) Sustainability analyzing driver drowsiness : from causes to effects, no 2008, pp 1–12
10. Rodrigues HS, Fonseca MJ, Cardoso PR (2015) The effect of drivers gender on the perception of Portuguese road safety communication campaigns. *Int J Bus Manag* III(4):104–123. <https://doi.org/10.20472/bm.2015.3.4.006>
11. Applications A (2021) A review of heartbeat detection systems for automotive applications
12. Strayer DL, Turrill J, Cooper JM, Coleman JR, Medeiros-Ward N, Biondi F (2015) Assessing cognitive distraction in the automobile. *Hum Factors* 57(8):1300–1324. <https://doi.org/10.1177/0018720815575149>
13. Srinivasan P, Khan AA, Prabu T, Manoj M, Ranjan M, Karthik K (2020) Heart beat sensor using fingertip through Arduino. *J Crit Rev* 7(7):1058–1060. <https://doi.org/10.31838/jcr.07.07.192>

Gain Scaling Tuning of Fuzzy Logic Sugeno Controller Type for Ride Comfort Suspension System Using Firefly Algorithm



Mat Hussin Ab Talib, Intan Zaurah Mat Darus, Hanim Mohd Yatim, Muhamad Sukri Hadi, Mohd Syahril Ramadhan Mohd Saufi, and Nor Hasrul Akhmal Ngadiman

Abstract A control system based on fuzzy logic (FL) is one of the effective controllers which operates using an inference mechanism rule base that requires a knowledge database. The system itself can remotely able to produce good linguistic variables depending types of output required. Nevertheless, the FL controller design still has a drawback that requires an improvement to give a very high capability in controlling a dynamic ride comfort of the vehicle suspension system. This study aims to improve the FL controller design by adding a gain scaling value for each input and output of the FL system. A metaheuristic-based firefly algorithm (FA) is used to optimize the value of each input and output of the FL system. Taking an acceleration of the suspension system response as an objective function, the FA strategy is an attempt to find and search for an optimum value of the gains that able to be as a sort of contact information for improving the targeted value obtained from the FL controller. In this work, an external disturbance in the form of sinusoidal waves is applied to the system to verify the sensitivity and durability of the proposed control schemes. Consequently, a comparative assessment between FL controller without having gain scaling and with the gain scaling tuned by FL strategy is investigated an analysis in the form of the amplitude reduction for both body displacement and acceleration responses. Simulation results indicated that the FL with gain scaling shows a good response compared to the FL without gain and its performance is improved by up to 52.1% compared to others.

Keywords Firefly algorithm · Fuzzy logic · Gain scaling · Metaheuristic algorithm · Suspension system · Ride comfort

M. H. Ab Talib (✉) · I. Z. Mat Darus · H. Mohd Yatim · M. S. R. Mohd Saufi · N. H. A. Ngadiman

School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
e-mail: mathussin@utm.my

M. S. Hadi
Faculty of Mechanical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

1 Introduction

The use of fuzzy logic (FL) controller becomes a major study that has been introduced by many researchers for any application recently and it is one of the systems that can imprecisely the inputs, as well as able to handle any nonlinearity problems [1–3]. It has been used in many type of application not only for controller purposes but also for modeling development. Ali et al. in 2018 have developed an FL controller for an agricultural greenhouse modeling using MATLAB Simulink environment to promote a suitable microclimate [4]. Al Badwawi and his team proposed the FL controller as the supervisory control in controlling the power management of an islanded AC microgrid using a frequency signalling [5]. The FL controller has been also widely used in engineering application that gives a remarkable performance for the system. For instance, Errouha et al., in 2019 also present the FL controller to improve the conventional direct torque control. The results show the effectiveness of the said controller in terms of torque and flux ripples reduction and daily pumped water [6]. Qais et al., in 2020 introduced the whale optimization algorithm-based Sugeno fuzzy logic controller to improve the fault ride-through of the wind generator. The simulation results revealed that by using the proposed controller can able to have less overshoot, fast time response, and small steady-state error [7].

In the application of the semi-active suspension system, many researchers have proposed good ideas in developing the FL logic controller with different approaches such as by integrating it with intelligent optimization, increase the number of input or output, and change the different types of fuzzy rules and membership functions [8–10]. Nevertheless, the improvement of the FL controller still has areas that could be investigated to increase the performance of the system. By adding gain scaling for every input and output of the controller, it is becoming one of the strategies that could give a good signal for the fuzzy logic data as a sort of context information. Thus, the information received from the fuzzy rules can be sorted for the better controller design after the gain scaling is included as a value added.

To optimize the value of FL gain scaling, a metaheuristic algorithm or so-called bio-inspired algorithm is used due to a powerful method that is able to solve the problem intelligently. Algorithm such as the particle swarm optimization (PSO), the artificial bee colony (ABC) and the cuckoo search algorithm (CSA) have been extensively used as a good optimizer in solving a nonlinear problem for any applications, recently [11–13]. Similarly, the firefly algorithm (FA) is the metaheuristic algorithms that is also highly being used as an optimizer in various applications [14, 15]. Thus, this study is an attempt to use the FA algorithm to compute the value of FL gain scaling and then to be integrated with the FL controller in order to improve the performance of the semi-active suspension system in terms of the amplitude reduction for both body acceleration and displacement analysis.

The paper will be organized as follows. Section 2 describe semi-active suspension modeling for a quarter vehicle using MATLAB simulation environment. In Sect. 3, the configuration of the controller system using the FL controller and the integration process with the FA algorithm is developed. Section 4 shows the analysis and

discussion for the FL controller without gain, the FL controller with gain tuned by FA algorithm, and the passive system. In Sect. 5, the conclusion is shown.

2 Semi-Active Suspension Modelling

Taking into consideration of second-newton law, the quarter car semi-active suspension system model is developed through a MATLAB simulation diagram. The block diagram model is shown in Fig. 1 and the mathematical equation of the said model is described as in Eqs. 1 and 2.

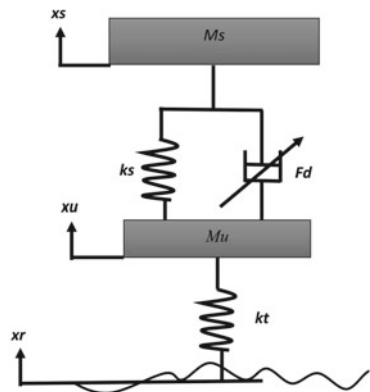
$$M_s \ddot{x}_s + F_d - k_s(x_u - x_s) = 0 \tag{1}$$

$$M_u \ddot{x}_u + F_d - k_s(x_u - x_s) - k_t(x_r - x_u) = 0 \tag{2}$$

where the sprung and unsprung mass is symbolized as M_s and M_u , respectively. The damper is represented as F_d , k_s is a spring stiffness and k_t is a tire stiffness. Other parameters like body acceleration, \ddot{x}_s , tire acceleration, \ddot{x}_u , body displacement, x_s , tire displacement, x_u , and road profile displacement, x_r are the important elements that can be defined as the main parameters of interests that were investigated in this study. The parameter value for sprung mass and unsprung mass are set as 80.5 and 18.5 kg, respectively whereas for spring stiffness and tire stiffness, the values are defined as 45,409 and 274,680 N/m, respectively. These scale-down values are taken experimentally regarding the real elements in the suspension structure.

The damper system used in this study is based on the magnetorheological (MR) system and s modeled using parametric approach called the Spencer model. The said model is also developed using MATLAB simulation block diagram at its parameters are defined from the previous study by other researchers [12].

Fig. 1 Semi-active suspension model



3 FL Controller Design and Optimization

The Fuzzy logic system is among the control system that is able to change the body and tire movements. It is first introduced by Lotfi A. Zadeh back in 1965 [16] by proposing the concept of linguistic variables and later to be defined as a fuzzy set. In this study, two sets of fuzzy inputs known as the body velocity and the relative velocity are defined in order to analyze their moving direction depending on the rules conducted in the system. As the input values are defined, the fuzzy inference rules with membership function are set up during fuzzification process before it is ending with the defuzzification process as important output values that can be sent to the next model. The fuzzy output in this study is defined as damper coefficient and later to be integrated with the relative velocity in order to produce desired force for the system. Overview of the fuzzy system that has been developed in MATLAB Simulink is shown in Fig. 2.

As for membership function, the gaussian-bell type is selected in this study as it is one of the most common types being used recently. The three linguistic variables are defined for each fuzzy input. Those are, Negative (N), Zero (Z), and Positive (P), and it is symmetric about zero. The function of membership is shown in Fig. 3.

The investigation of the fuzzy rules is the next step in executing it based on the IF–THEN rules. The fuzzy rule is as follows:

$$\text{If } \dot{z}_s \text{ is } (A) \text{ and } \dot{z}_{rel} \text{ is } (B) \text{ then } C_d \text{ is } (C)$$

where the values of A , B , and C are representing as the sprung mass velocity, the relative velocity, and the desired damping coefficient, respectively. By using Sugeno Fuzzy Type, the damping coefficient which representing the fuzzy output is determined by the range of the low and high damping state as shown in Table 1 and its

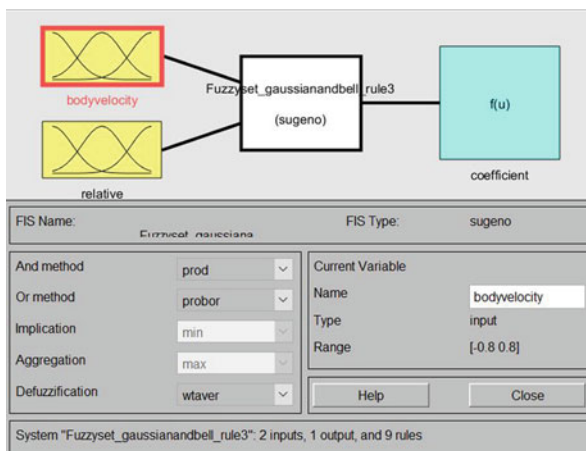


Fig. 2 The fuzzy logic system developed in MATLAB Simulink

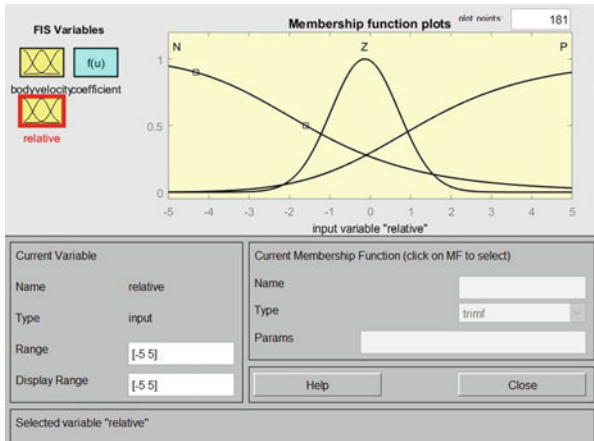


Fig. 3 The set of fuzzy membership function

Table 1 Fuzzy coefficient

C_{min}	C_{d1}	C_{d2}	C_{d3}	C_{d4}	C_{max}
700	3000	6000	9000	12,000	15,000

rule is also depicted in Table 2. The fuzzy coefficient values as presented in Table 1 are obtained regarding on a number of the damping values in the range of low and high damping state [17].

As the FL controller itself has many varieties to be improved and working as high performance for particular purposes, it has some ideas to be proposed in this study by incorporating the gain scaling for each input and output of the fuzzy system. The gain scaling known as the gain sprung velocity (GSV) and the gain relative velocity (GRV) is used to be included in the fuzzy system as scale values for the fuzzy input while the fuzzy output is also adopted with the gain scaling known as the gain coefficient (GC). All these parameter gains will be optimized using the FA optimization technique.

Table 2 Fuzzy rule

		Relative velocity (RV)		
Spring velocity (SV)	Fuzzy rule			
		N	Z	P
N		C_{max}	C_{d2}	C_{min}
Z		C_4	C_{d3}	C_{d1}
P		C_{min}	C_{d4}	C_{max}

3.1 Firefly Algorithm

Firefly Algorithm is a group of metaheuristic algorithms inspired by the flashing behavior of fireflies based on nature and it is introduced by Yang, back in 2010 [18]. The FA uses brightness or light intensity as the main objective function and the main concept of the algorithm is based on the pair of two fireflies, x_i and x_j . The lower brightness of the particle is attempt to find the brightest particles in order to increase their brightness. As the distance between two particles is far apart, the brightness will decrease. A group of fireflies will be sorted out according to higher brightness to lower brightness. The process of the particle to find their best mate or partners will remain continues until the number of generations is limited to particular values. The main three equations involved during the generation process are known as the attractiveness (β), the cartesian distance (r_{ij}), and the particle movement (x_i) and it is shown as in Eqs. 3, 4, and 5, respectively.

$$\beta(r) = \beta_0 e^{-\gamma r^m}, m \geq 1 \quad (3)$$

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (4)$$

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha \varepsilon_i \quad (5)$$

where γ is known as the fixed light absorption coefficient, the k^{th} component is the spatial coordinate x_i of i^{th} firefly, d is the dimension number, α is the randomization parameter and ε_i is a vector of random numbers defined by the uniform distribution. The summary of the algorithm process can be depicted in the form of pseudo-code as shown in Fig. 4.

The block diagram of the system is shown in Fig. 5. The MSE of the system is based on the body acceleration error and it would be defined as an objective function for the FA algorithm process.

4 Result and Discussion

The evaluation process in determining the system performance can be discussed in terms of body acceleration and body displacement by comparing the proposed FL gain scaling by FA with the FL without gain scaling and the passive system. The selection of FL without gain to be as a competitor as it is a normal controller type that is widely used recently. A 4 Hz frequency and 0.005 m amplitude of sinusoidal wave road profile are applied to be the disturbance input to the system. Table 3 shows the optimized result of the gain scaling values for the FL controller and it is worth to be mentioned that the mean square error (MSE) value is decreasing and

```

Objective Function,  $f(x)$ ,  $x = (x_1, \dots, x_d)$ 
Generate initial population of fireflies,  $x_i (i = 1, 2, \dots, n)$ 
Light intensity  $I_i$  at  $x_i$  is determined by  $f(x_i)$ 
Define light absorption coefficient
while ( $t < MaxGeneration$ )
    for  $i = 1 : n$  (all n fireflies)
        for  $j = 1 : i$  (all n fireflies)
            if ( $I_j > I_i$ ), move firefly  $i$  towards  $j$  in d-dimension; end if
            Attractiveness varies with distance  $r$  via  $exp [-r]$ 
            Evaluate new solutions and update light intensity
        end for  $j$ 
    end for  $i$ 
Rank the fireflies and find the current best
end while
Post process results and visualization
    
```

Fig. 4 The pseudo-code of FA

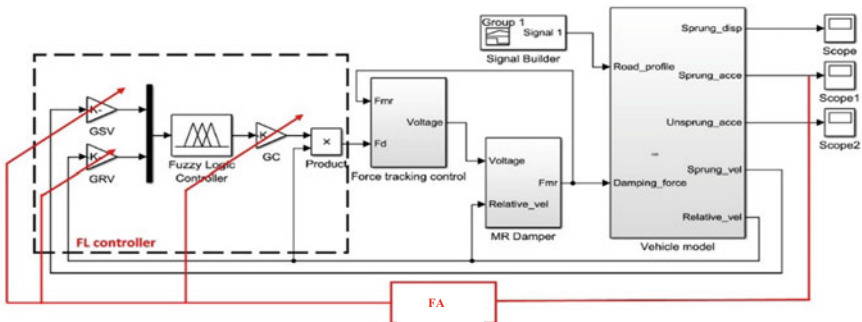


Fig. 5 Full configuration of the block diagram

Table 3 The optimized result and MSE

Controller	No. of iteration	Optimized value	MSE
FL gain scaling tuned by FA	50	Gsv: 3.666	9.233
		Grv: 3.243	
		Gc: 0.3577	

converged at a particular value as the number of generations increases. The overall result performance of the optimization process is also shown in Fig. 6.

Figures 7 and 8 show the body acceleration and displacement analysis responses, respectively. The body acceleration analysis becomes an indicator to investigate the vehicle ride comfort performance. Referring to these figures, the improvement of

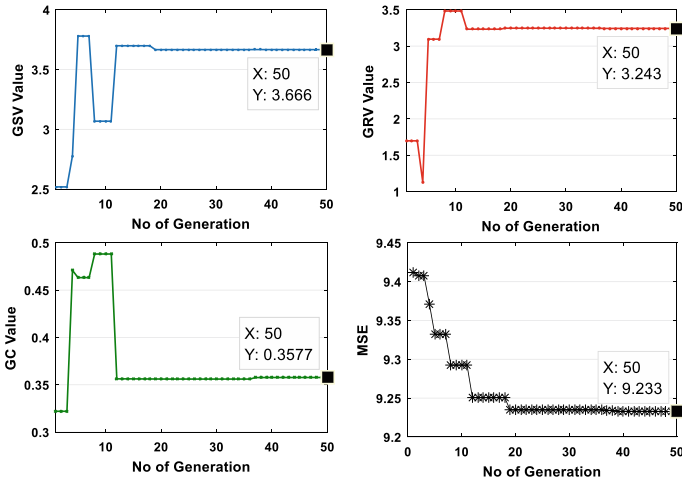


Fig. 6 Convergence curve and the gain scaling values for AF algorithm

Fig. 7 Analysis result for body acceleration

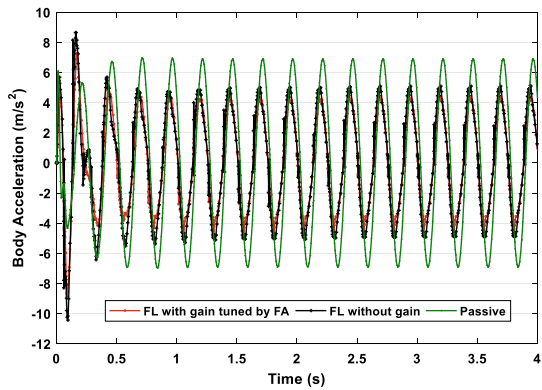


Fig. 8 Analysis result for body displacement

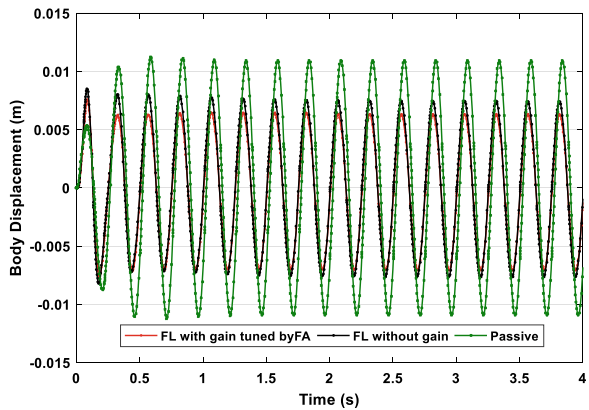


Table 4 MSE and percentage improvement of all parameters of interest

Index	Passive	FL without gain	FL with gain tuned by FA
Body acceleration (m/s ²)	17.6	12.67 (28.0%)	9.23 (47.5%)
Body displacement (m) (10 ⁻⁵)	4.43	2.71 (38.9%)	2.12 (52.1%)

vehicle ride comfort can be seen when the FL controller with gain scaling managed to increase the system performance by up to 47.5 and 52.1% for both body acceleration and body displacement, respectively. The full results can be obtained in Table 4.

5 Conclusion

The suspension system model with the controller configuration has been well developed using MATLAB Simulink. Referring to results, the FL with gain scaling tuned by FA managed to improve the vehicle ride performance (body acceleration analysis) by up to 47.5% compared to the FL controller without gain scaling and the passive system. It has also proven the vehicle amplitude (body displacement analysis) shows the same trend as the body acceleration analysis when the performance of the FL with FA is better than the FL controller without gain and the passive system by up to 52.1%.

Acknowledgements The authors would like to express their gratitude to Minister of Education Malaysia (MOE) and Universiti Teknologi Malaysia (UTM) for funding and providing facilities to conduct this research. This research is supported by UTM Research University grant, Vote No. 4B447.

References

1. Dzitac I, Filip FG, Manolescu MJ (2017) Fuzzy logic is not fuzzy: world-renowned computer scientist Lotfi A. Zadeh. *Int J Comput Commun Control* 12(6):748–789
2. Sharma S, Obaid AJ (2020) Mathematical modelling, analysis and design of fuzzy logic controller for the control of ventilation systems using MATLAB fuzzy logic toolbox. *J Interdisc Math* 23(4):843–849
3. Kambalimath S, Deka PC (2020) A basic review of fuzzy logic applications in hydrology and water resources. *Appl Water Sci* 10(8):1–14
4. Ali RB, Bouadila S, Mami A (2018) Development of a Fuzzy Logic Controller applied to an agricultural greenhouse experimentally validated. *Appl Therm Eng* 141:798–810

5. Al Badwawi R, Issa WR, Mallick TK, Abusara M (2018) Supervisory control for power management of an islanded AC microgrid using a frequency signalling-based fuzzy logic controller. *IEEE Trans Sustain Energy* 10(1):94–104
6. Errouha M, Derouich A, Motahhir S, Zamzoum O, El Ouanjli N, El Ghzizal A (2019) Optimization and control of water pumping PV systems using fuzzy logic controller. *Energy Rep* 5:853–865
7. Qais MH, Hasanien HM, Alghuwainem S (2020) Whale optimization algorithm-based Sugeno fuzzy logic controller for fault ride-through improvement of grid-connected variable speed wind generators. *Eng Appl Artif Intell* 87:103328
8. Ab Talib MH, Mat Darus IZ (2014) Development of fuzzy logic controller by particle swarm optimization algorithm for semi-active suspension system using magneto-rheological damper. *WSEAS Trans Syst Control* 9(1):77–85
9. Mustafa GI, Wang HP, Tian Y (2019) Vibration control of an active vehicle suspension systems using optimized model-free fuzzy logic controller based on time delay estimation. *Adv Eng Softw* 127:141–149
10. Koc MA (2020) Fuzzy logic control of vibrations due to interaction one DOF vehicle suspension and flexible structure with tuned mass damper. *J Smart Syst Res* 1(1):1–10
11. Priyadarshi N, Azam F, Sharma A, Bhoi AK, Kumar MA (2018) Particle swarm optimization based fuzzy logic control for photovoltaic system. *Int J Eng Technol* 7(3.24):491–496
12. Ab Talib MH, Darus IZM, Samin PM, Yatim HM, Ardani MI, Shaharuddin NMR, Hadi MS (2021) Vibration control of semi-active suspension system using PID controller with advanced firefly algorithm and particle swarm optimization. *J Ambient Intell Humaniz Comput* 12(1):1119–1137
13. Wang H, Wang W, Xiao S, Cui Z, Xu M, Zhou X (2020) Improving artificial bee colony algorithm using a new neighborhood selection mechanism. *Inf Sci* 527:227–240
14. Hassan BA (2021) CSCF: a chaotic sine cosine firefly algorithm for practical application problems. *Neural Comput Appl* 33(12):7011–7030
15. Sisi ZA, Mirzaei M, Rafatnia S, Alizadeh B (2021) Design of a constrained nonlinear controller using firefly algorithm for active suspension system. *Comput Methods Eng* 39(2):23–44
16. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
17. Hudha K, Jamaluddin H (2011) Simulation and experimental evaluation on a skyhook policy-based fuzzy logic control for semi-active suspension system. *Int J Struct Eng* 2(3):243–272
18. Yang XS (2010) *Nature-inspired metaheuristic algorithms*. Luniver press, London

Evaluation of a DC Motor Temperature Response Characterization Method Under Different Sampling Interval



I. A. Kamaruddin, M. A. H. Rasid, N. F. Abdullah, and A. Abdul Wahab

Abstract A fast and simple diagnostic tool is necessary to reduce the cost of preventive maintenance in industrial settings. Being the main actuator in industry, motors and generators need regular monitoring on their state of health. Temperature response characterization of the motor can be a promising solution if reference temperature characterization is documented in the early phase of assembly and commissioning. Temperature response characteristics of a motor that can be considered as a first order response are characterized by its steady-state temperature T_{ss} and time constant τ . These characteristics can be collected and compared to the reference at regular time intervals to diagnose any eventual fault in the motor. This study analyzes an algorithm that characterizes the temperature response collected from a Brush DC motor. The time interval δt , and the temperature resolution $\delta Temp$ are the two determining parameters in the algorithm. The results found that the larger δt and the smaller $\delta Temp$ give a more precise steady-state temperature deduction. With the choice of $(\delta t, \delta Temp) = (300 \text{ s}, 0.05 \text{ }^\circ\text{C})$, the steady-state temperature was deduced by the algorithm with an error of 0.235%. However, the opposite, smaller δt and the larger $\delta Temp$ can give a percentage of that steady-state temperature that could potentially be used to reduce the time of diagnosis. For the MY1016 DC motor, the choice of $(\delta t, \delta Temp) = (50 \text{ s}, 1.5 \text{ C})$ allows a diagnosis reduction time of 5895 s, attaining 64% of the steady-state temperature.

Keywords Temperature response · Electrical machine · Steady state · DC motor · Fault diagnosis

1 Introduction

In industrial settings, it is necessary to identify any fault in machines and equipment. It is much better if it is part of a preventive regular process to avoid lengthy and costly downtime. Electrical motors and generators are known to be the main

I. A. Kamaruddin · M. A. H. Rasid (✉) · N. F. Abdullah · A. A. Wahab
Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia
Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: mahizami@ump.edu.my

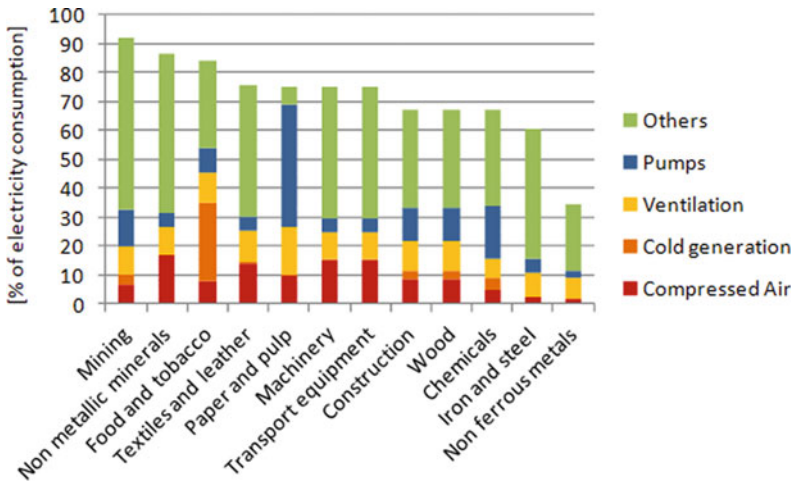


Fig. 1 Assumptions on the share of motor systems in total electricity consumption by sub-sector (values in percent)

actuators that moves the majority of equipment in industries, ranging from manufacturing, energy generation and auxiliary functions like ventilation, heating and cooling. Figure 1 shows the industry and applications in which electric motors are employed [1]. This huge market open a possibility for different tools of diagnosis to be applied, depending on the requirements of the maintenance of each application.

In this study, concentrating on thermal diagnosis, an algorithm to characterize the temperature response measured on a motor is analyzed. The introduction will clarify the advantage of a thermal analysis compared to other method and define the temperature response characteristic. The methodology of the analysis and the results are presented in separate section.

1.1 Electric Motor Fault Diagnosis

The environment and the loading of each application may induces more damages to certain components than others. These situations include overload, lacking lubrication, repeated motor starts/stops, and insufficient cooling. Under these conditions, electric motors are exposed to undesirable stresses, which put the motors under risk of faults or failures [2, 3]. Due to the commonness of electric motors and their major positions in applications, many efforts have been made to improve a motors consistency. According to IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems on 1997 provides in Table 1 , i.e., bearing (41%), stator (37%), rotor (10%) and other (12%) [4].

Methods employed to diagnose fault in electrical machines are numerous, including regular visual inspection, vibration analysis, current analysis (namely

Table 1 Statistical study by IEEE recommended practice for the design of reliable industrial and commercial power systems

Number of faults					
Type of fault	Induction motors	Synchronous motors	Wound rotor motors	DC motors	All motor types (total)
Bearing	152	2	10	2	166
60.42	60.54	16	6	–	97
Rotor	8	1	4	–	13
Shaft	19	–	–	–	19
Brushes or slip rings	–	6	8	2	16
External devices	40	7	1	–	18
Others	10	9	–	2	21

MCASA method) and in rare cases, thermal analysis. Some of the prominent method analyzed can be seen in Table 2.

Among the methods cited, MCASA and vibration analysis are the most employed in industries that allows higher spending for preventive maintenance due to the high cost. These methods usually needs specialist to employ the tools necessary and make the analysis necessary in order to come out with a conclusive report. The regular visual inspection is the common practice. It might seem less expensive, however with a downtime to allow machines dismantling and reassembling, the shutdowns cost might be higher even if they are planned and predetermined.

Thermal analysis has a certain important advantage. In previous study, most employed method is thermography and this analysis is only available for surface temperature. Thermocouple instrumented inside the components are more interesting. The implementation of the thermal diagnosis tools are far cheaper and the characterization of temperature is far simpler and can be performed with minimum training. Thermocouple are not expensive and can be instrumented on critical components during assembly and installation of the equipment. The data acquisition can be automated and the temperature can be observed in real time either on site or through IoT system. These lead to much cheaper monitoring and diagnosis tool system. Thermal analysis is however rarely employed due to several inconveniences. The first one is that the hot spots are only obvious to observations when the damage is in an advance state. Secondly is due to the lack of reference temperature characterization of the machines used. Both reasons can be traced back to one root cause, which is the lengthy delay to perform the thermal characterization of an electrical machine.

To remedy the problem, it is necessary to optimize the duration of temperature characterization. In this study, we will analyze the influence of temperature sampling resolution in temperature response characterization. To do that, in this introduction section, we will briefly present the definition of temperature response and its main characterization parameters. In the methodology, we will present an algorithm that we created to evaluate the characteristics of a temperature response data extracted

Table 2 Examples of different fault diagnosis techniques employed

Study	Diagnosis method	Fault diagnosed	Specificity
Xiao [5]	MCASA	Stator winding fault, rotor broken bar and bearing fault	Stator current analysis using Stacked Denoising Auto-encoders (SDAE) for feature extraction in Induction Machine
Ben Salem [6]	MCASA	Static, dynamic and mixed air gap eccentricity	Motor current signature analysis to diagnosis eccentricities fault in induction motor with analyze frequency using FFT algorithm
Huang [7]	Vibration	Air gap eccentricity, bearing failure, and broken rotor bar	Extension neural network (ENN) is used to analyze the frequency spectrum of vibration signal to detect the fault on induction motor
Delgado-Arredondo et al. [8]	Vibration	Two broken rotor bars, mechanical unbalance bearing defects	Acoustic and vibration signal analysis and to detect common fault of induction motor using Complete Ensemble Empirical Mode Decomposition
Redón et al. [9]	Thermal	Ventilation problem and bearing defect	Infrared thermography coupled with signal processing algorithms is used in FDD of induction motor to detect fault by pattern of temperature profile on motor part
Garcia-Ramirez et al. [10]	Thermal	Rotor broken bar, bearing fault and misalignment	Thermographic analysis technique has used in fault diagnosis with focus only on the detection of common faults occur the effect of the induction motor parts and the elements of their kinematic chain

for an experimental test bench, and the method that was used to try to minimize the time taken to make the characterization. Finally, we will analyze the influence of different temperature data sampling resolution on the characterization precision and the time taken.

1.2 Temperature Response Characteristics

What we define by temperature response here is the temperature recorded on a point when the object of study is subjected to a step of heat flux. In control engineering, it is equivalent to the step response [11]. The input being a step of heat flux at a given amplitude, the output is the temperature hence, it is called temperature response. The subject or the system transforming the heat flux into temperature is the thermo-physical characteristic of the motor (Fig. 2).

In most cases [12–16] the temperature response is an under-damped response, which may lead us to make an assumption that the thermal behavior of a motor is a first order system. Following that assumptions, the temperature response can therefore be characterized following a first order system performance criteria. The performance criteria are the steady-state temperature and the time constant [17].

Steady-state temperature T_{ss} can be attained when the studied object is subject to a steady-state heat transfer (conduction, convection or/and radiation). This happens when the temperature difference driving the conduction are constant, so that (after an equilibration time), the spatial distribution of temperatures (temperature field) in the conducting object does not change any further. Thus, all partial derivatives of temperature concerning space may either be zero or have nonzero values, but all derivatives of temperature at any point concerning time are uniformly zero. The time constant τ in other hand is the product of thermal resistance and thermal conductance, and this product is equal to around 63% of the steady-state temperature.

By gathering and inventorying these data on different components of the machine, we can build a classification of thermal characteristics. Later, by comparing these two criteria of a diagnosed motor with references values, we can hopefully expect to observe variations when there are anomalies in the motor such as faulty components.

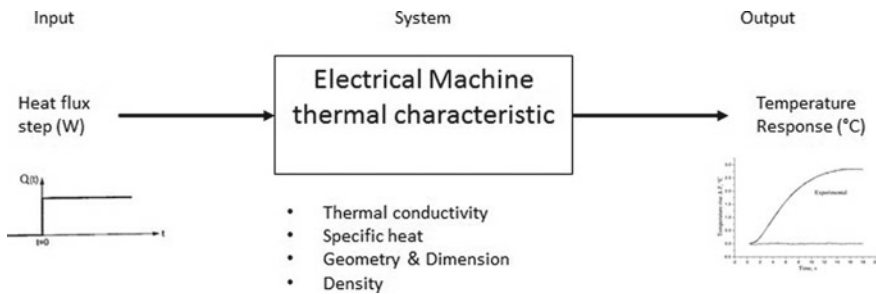


Fig. 2 System, input and output in thermal behavior of an electrical machine

2 Methodology

From the introduction, we know that the time taken to do thermal characterization depends on the time it takes to reach steady-state temperature. Depending on the size of the motor, the thermal conductivity and the specific heat of the components, it may take from 2 h and higher. To minimize the time taken to conduct the temperature diagnosis, two sampling parameters of the temperature response can be manipulated: the time interval, δt and the temperature resolution, $\delta Temp$. The size of the sampling is here known as the sampling resolution.

In this section, we will present the algorithm that we used to deduce the steady-state temperature, T_{ss} and time constant, τ , the definition of sampling resolution, and the experimental set up that was used to get a temperature response.

2.1 Temperature Response Characterization Algorithm

In order to deduce the T_{ss} , an algorithm was developed. The program consists of three main steps. The main function of the algorithm is to search for the steady-state temperature from the temperature response data provided to it. To do it, the algorithm sweep through the data and calculate the temperature difference in a defined interval of time. We define the two resolution parameters as the time sampling, δt and the temperature sampling, $\delta Temp$. δt represents the time discretization in which the slope of the temperature response curve is computed, and $\delta Temp$ represents the temperature discretization which serves as condition to stop the steady-state temperature searching algorithm. Its' definition can be illustrated in Fig. 3. The algorithm will stop when the temperature difference reach a value $\delta Temp$ set by the user.

In an ideal case, the temperature difference when we are at steady state should be null, therefore the slope should equal to zero. However, in practical, fluctuations will always exist, leading to a non-terminating algorithm. Furthermore, the more important issue is, the smaller the temperature resolution $\delta Temp$ we define, the longer it takes for the algorithm to reach the defined steady state. As for the time interval δt , a higher precision can be obtained if we have a large time interval. On the other hand, a smaller time interval will be sensitive to fluctuations of the measurement.

In summary, the algorithm involves three main steps. The first step is to load the temperature response data and pre-process the data. This involving in converting the format of the data that come from the DAQ in csv or txt file to a double precision floating points to be treated in Matlab, removing erroneous data and smoothing the fluctuation. Then, starting at the initial temperature, the algorithm will start to compute $\delta Temp$ in each δt interval, continuously in loop. Finally, the loop will only stop when $\delta Temp$ reaches the conditional value set by the user. Only then, it will display the steady-state temperature and compute the time constant. The flowchart

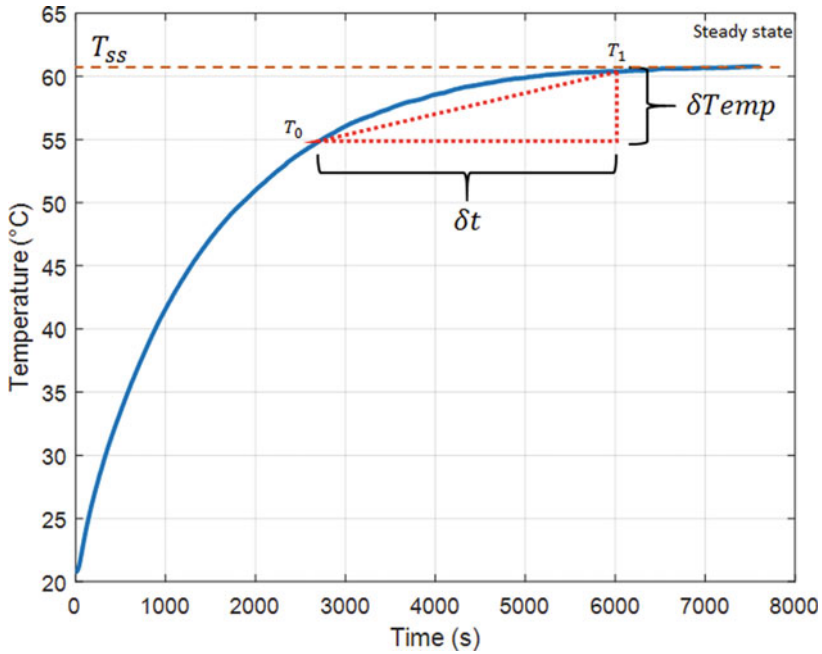


Fig. 3 Definition of time interval δt and temperature resolution $\delta Temp$ of a temperature response

representing the process in the algorithm can be found in Fig. 4 and the detail script at the website found in the Appendix.

In this study, we are evaluating the influence of the definition of both the $\delta Temp$ and δt on the precision of the deduced steady-state temperature. The steady-state temperature and time constant will be computed using a range of values as shown in Table 3.

The range for temperature resolution is limited by the precision of the measurement, while the time interval can be changed arbitrarily.

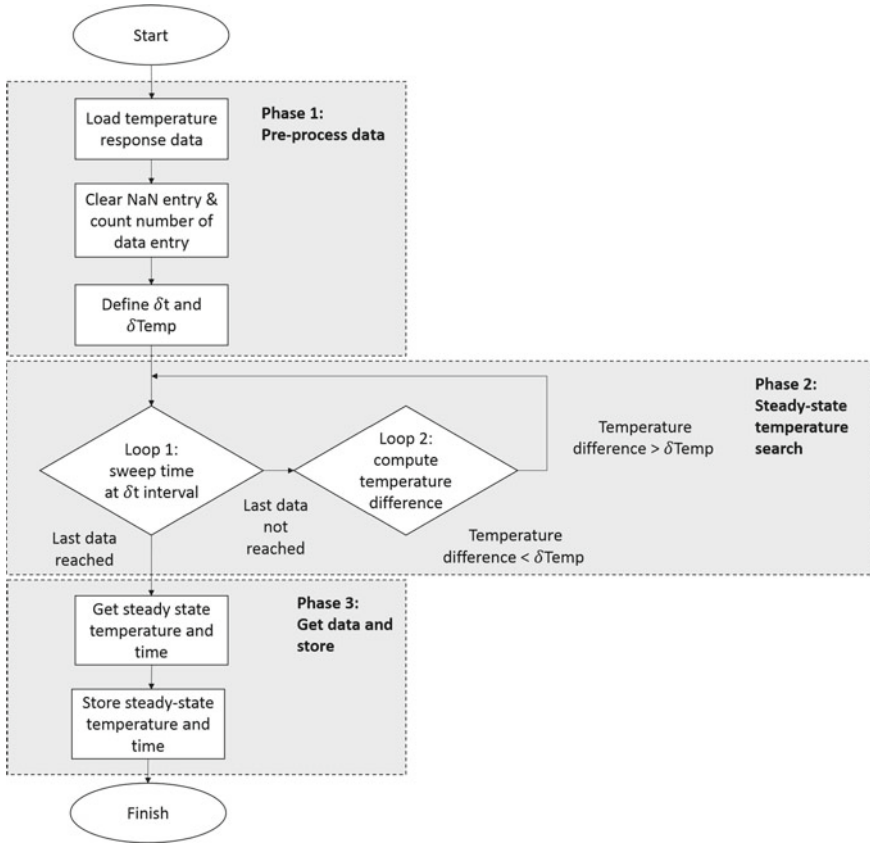


Fig. 4 The flowchart representing the algorithm. The program can be found at the website with link in the Appendix

Table 3 The range of values tested for analysis of the algorithm

Parameters values	
δt , (s)	50, 100, 150, 200, 250, 300, 400, 500
$\delta Temp$, ($^{\circ}C$)	0.05, 0.1, 0.5, 1.0, 1.5

2.2 The Experimental Test Bench

The experimental test bench the temperature response was gathered from a DC Brush Motor test bench. The motor on which the temperature was recorded is a 24 V DC Motor with the specification as shown in the Table 4. The motor was instrumented with type-K thermocouple on several components: casing, bearing, brush, and permanent magnet (Fig. 5). These are the components on which temperature response characterization can be made. However, for the purpose of this study which

Table 4 Specification of the DC motor used

Model	MY1016
Operating voltage	24 VDC
Rated current	13.5 A
Rated speed	2650 Rpm
Operating power/output	250 W
Rated torque	100 N-cm
No load current	<2.2 A
Weight	2.0 kg
Dimension	(20 * 15 * 10) cm

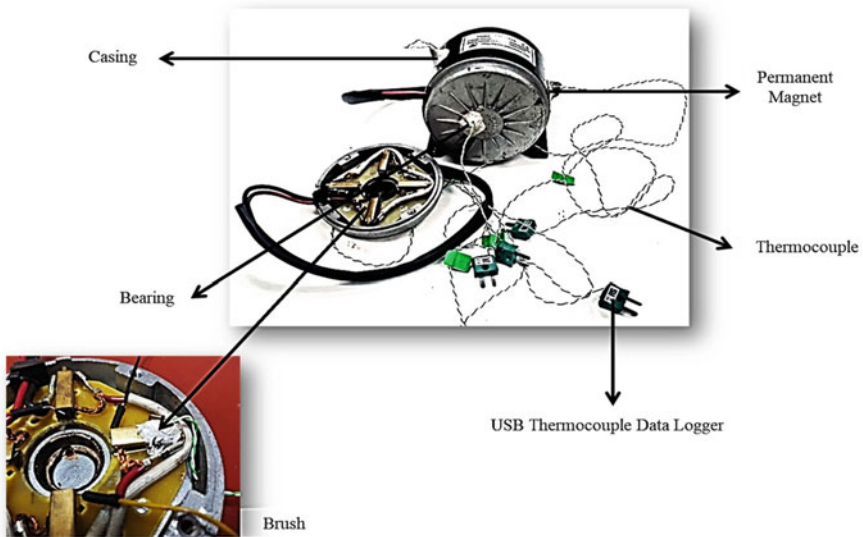


Fig. 5 The MY1016 DC motor components that are instrumented with type-K thermocouple

is evaluating the sampling resolution, we are going to concentrate only on the most important temperature point in the motor. The most important hot spot in a motor is the armature winding. Therefore, the closest non-moving component that can be monitored here is the brush.

The motor is then installed in its setup, connected to a load simulated by a mechanical brake (Fig. 6). The thermocouples are connected to a DAQ that is sampling at 1 sample per second and send, store and display the data on a host PC. The motor is controlled by a power supply. To monitor the constant torque and speed during operation, an encoder measures its speed and a clamp meter measures the DC bus current and display it on an oscilloscope. A constant torque and speed is imaged by the constant encoder measurement and constant current.

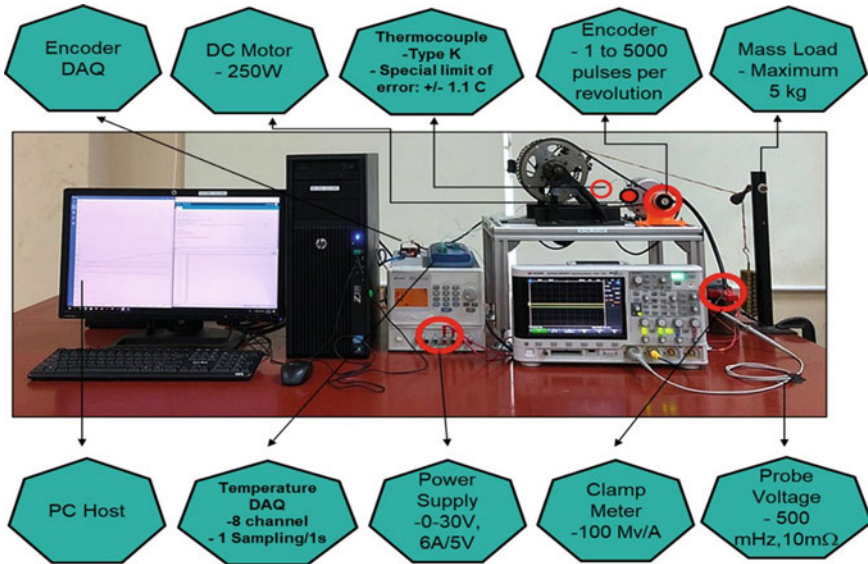


Fig. 6 The complete experimental test bench to gather the temperature response

The data sampled to test the algorithm was taken for a speed at 1620 rpm and load that corresponds to 3 A DC bus current. The motor run until there is no variation of temperature for 10 s which we consider enough to be in steady state. The steady-state temperature found here will be used as reference steady-state temperature to which the algorithm-deduced steady-state temperature will be compared to.

3 Results and Analysis

This section is divided into two parts: The first one presenting the temperature response general form of the tested DC motor in experimental setup. The second part will deal with the analysis of the time and temperature sampling resolution (δt and $\delta Temp$) towards searching the steady-state temperature.

3.1 Experimental Temperature Response

Figure 7 shows the temperature response of all the components in the motor instrumented with thermocouple. The complete steady state is reached at 7620 s, which is more than 2 h. All the components express the behavior that can be suggested as a first order response. The highest temperature is reached by the brush at 60.76 °C followed by the other components in cluster (in descending temperature order: bearing, casing,

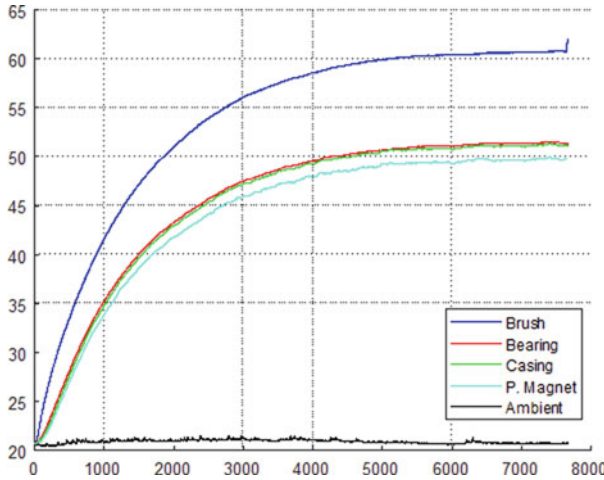


Fig. 7 Temperature response of all the component of DC motor that were instrumented

and permanent magnet). With an ambient temperature recorded at around 20 °C, the temperature rise in brush is at more than 40 °C with a heat flux input generated by a load equivalent to 3 A in the motor DC bus. With this observation, we select the component with the highest temperature response, which is the brush to be subjected to the sampling resolution analysis. The steady-state temperature of the brush at 60.76 °C will serve as the reference temperature T_{ref} to evaluate the precision of the algorithm-deduced steady-state temperature.

3.2 Analysis of the Sampling Resolution Influence

The temperature response data set of the brush is then used as input for the steady-state temperature search algorithm presented in Fig. 4 and Appendix A. For the combination of δt and $\delta Temp$ presented in Table 1, the steady-state temperature and the time taken to reach that steady-state temperature is computed. Figure 8 shows the results: 2 surface mapping the steady-state temperature T_{ss} and the time taken to reach steady state, with regards to the two algorithm resolution parameters δt and $\delta Temp$ (Tables 5 and 6).

It can be observed that the steady-state temperature deduced by the algorithm is lower as the temperature resolution increase. Meanwhile, the steady-state temperature deduced is higher as the time interval is increased. Naturally, the right surface also shows that the time taken to reach that steady state also evolves in the same pattern. The bigger the temperature resolution set in the algorithm, the faster it takes to deduce the steady-state temperature, and also the lower the steady-state temperature deduced.

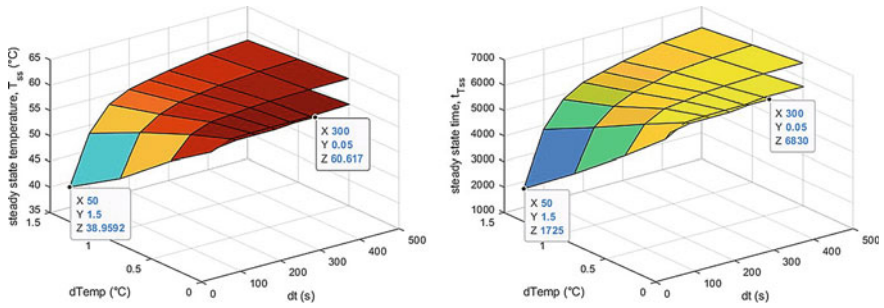


Fig. 8 Left: The steady-state temperature deduced by the algorithm; Right: The time to reach the steady-state temperature deduced by the algorithm

Table 5 Steady-state temperature deduced by the algorithm

dTemp (°C)	dt							
	50	100	150	200	250	300	400	500
0.05	59.78	60.28	60.31	60.42	60.54	60.62	NaN	NaN
0.10	58.56	60.09	60.22	60.25	60.28	60.30	60.59	NaN
0.50	53.19	57.29	58.73	59.07	59.43	59.66	59.84	59.88
1.00	45.18	53.12	55.68	57.01	57.63	58.22	58.86	59.17
1.50	38.96	48.39	53.02	54.93	55.77	56.56	57.71	58.32

Table 6 Time at which the steady-state is reached

dTemp (°C)	dt							
	50	100	150	200	250	300	400	500
0.05	6580.00	6730.00	6735.00	6740.00	6830.00	6830.00	NaN	NaN
0.10	6220.00	6665.00	6710.00	6725.00	6730.00	6820.00	6815.00	NaN
0.50	4840.00	5870.00	6270.00	6360.00	6475.00	6530.00	6590.00	6600.00
1.00	3225.00	4825.00	5445.00	5795.00	5965.00	6165.00	6305.00	6390.00
1.50	1725.00	3830.00	4805.00	5265.00	5470.00	5670.00	5990.00	6160.00

For the range tested, the shortest time taken to reach steady-state temperature is at 345 s, with the deduced steady-state temperature of 38.96 °C for $(\delta t, \delta T_{emp}) = (50 \text{ s}, 1.5 \text{ }^\circ\text{C})$. The longest time taken is at 1366 s, with the deduced steady-state temperature of 60.62 °C for $(\delta t, \delta T_{emp}) = (300 \text{ s}, 0.05 \text{ }^\circ\text{C})$, which is also the closest to the reference steady-state temperature T_{ref} . Therefore, the larger the time interval and the smaller the temperature resolution, the closer the deduced steady-state temperature to the reference temperature. The increase of time interval and reduction of temperature resolution have however limitations, where we observed that the steady state cannot be found by the algorithm for $(\delta t, \delta T_{emp})$ equals to (400 s, 0.05 °C), (500 s, 0.05 °C) and (500 s, 0.1 °C). This occurs because the slope is computer for a very large time

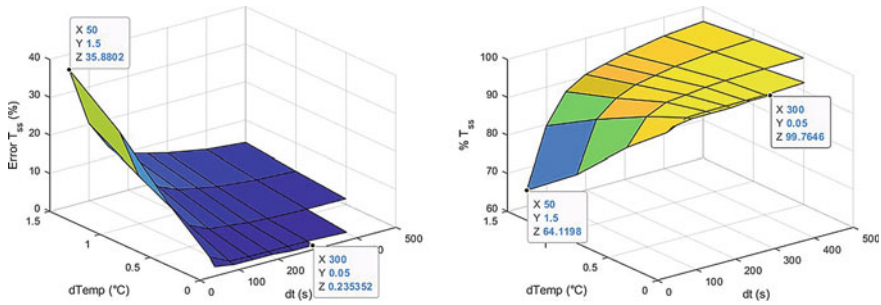


Fig. 9 Left: Error of the deduced steady-state temperature; Right: The deduced temperatures actually a percentage of the actual steady-state temperature

interval, yet very small temperature variation. It shows that the temperature is still increasing, but very slowly at around 0.05 °C for every 400 s (6.7 min)

In terms of temperature resolution, it is obvious that it is better to have the smallest value possible for $\delta Temp$. Its variation only have important impact when the time interval becomes smaller. In diagnosis application, we would want to have a shorter diagnosis time, yet a sufficient precision of steady state characterization. As the choice of time interval is shown to impact the deduced steady-state temperature, it is necessary to evaluate the gain in time reduction due to the choice of a larger time interval, compared to the error on the deduced steady-state temperature.

Compared to the T_{ref} defined previously in Sect. 3.2, the error on steady-state temperature deduction by algorithm is plotted on the left of Fig. 9. As expected, the steady state deduced with $(\delta t, \delta Temp) = (50 \text{ s}, 1.5\text{C})$ leads to the highest error of 35.88%. When normalized at 100%, the error represents in fact the percentage of time response data that is not evaluated by the algorithm due to the condition of temperature resolution to stop the steady state search is already achieved. Consequently, we can also see the compliment of the error as the percentage out of the reference steady state. The right of Fig. 9 shows the percentage of the reference steady-state temperature attained by the algorithm for a given $(\delta t, \delta Temp)$ pair. For the same pair of $(\delta t, \delta Temp) = (50 \text{ s}, 1.5 \text{ °C})$, the algorithm manage to attain 64.12% of the reference steady-state temperature. Therefore instead of focusing on the error, the temperature reached can be referred as the 64.12% of the steady-state temperature. The larger the time interval and the smaller the temperature resolution, the closer the algorithm will get to the reference steady-state temperature. However, as shown in Fig. 8 right previously, it will also take longer time

For a machine that was tested to get the reference steady-state temperature during the installation and commissioning, comparing the partial percentage of steady-state temperature $\%T_{ss}$ for diagnosis may be sufficient. This will reduce the time for diagnosis. To evaluate the time that can be saved if the temperature is recorded up to a certain $\%T_{ss}$, we plotted the graph of $\%T_{ss}$ against the diagnosis reduction time for a given δt interval in Fig. 10, and for a given $\delta Temp$ resolution in Fig. 11.

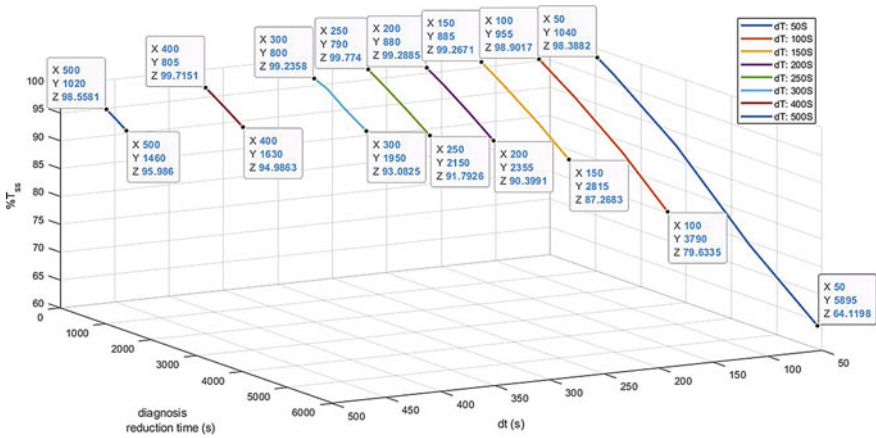


Fig. 10 The percentage of steady state reached vs. diagnosis reduction time for a given algorithm δt interval

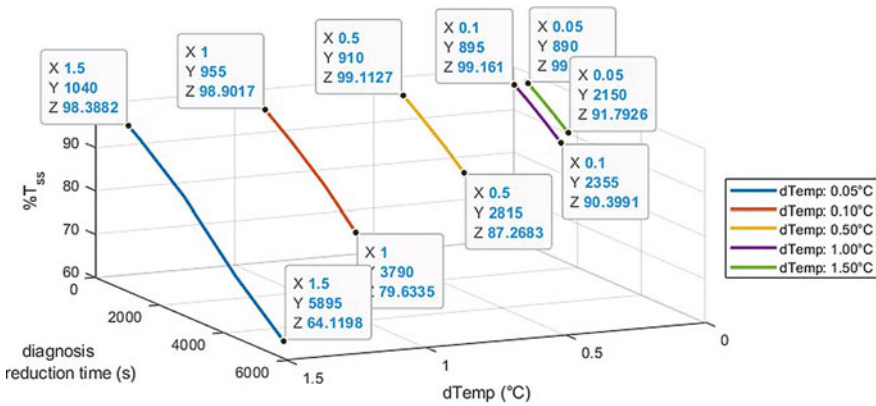


Fig. 11 The percentage of steady state reached vs. diagnosis reduction time for a given algorithm $\delta Temp$ resolution

Figure 11 shows that for our particular MY1016 Brush DC motor, the minimum diagnosis time reduction can be obtained is 5895 s, but will only attain 64% T_{ss} with the largest $\delta Temp$ resolution at 1.5 °C. Referring to Fig. 10, this diagnosis time reduction also needs δt interval to be at the smallest value of the tested range which is 50 s. Both figures show that the choice of $\delta Temp$ resolution and δt interval is critical to allow diagnosis time reduction from the point of view of algorithm that evaluate the temperature response characteristics. The lower points of each curves in both graph shows the maximum diagnosis time reduction that can be allowed for the respected choice of $\delta Temp$ resolution and δt interval.

4 Conclusion

The algorithm to characterize the temperature response of a motor has been developed. Using data from the MY1016 brushed DC motor, it has been shown that the larger δt and the smaller $\delta Temp$ gives a more precise steady-state temperature deduction. With the choice of $(\delta t, \delta Temp) = (300 \text{ s}, 0.05 \text{ }^\circ\text{C})$, the steady-state temperature was deduced with an error of 0.235%. However, the opposite, smaller δt and the larger $\delta Temp$ can give a percentage of that steady state that could potentially be used to reduce time of diagnosis. The choice of $(\delta t, \delta Temp) = (50 \text{ s}, 1.5\text{C})$ allows a diagnosis reduction time of 5895 s, but only attain 64% of the steady-state temperature.

In perspective, in order to develop a reliable thermal diagnostic tool, two question need to be answered in the near future. The first one is to verify and validate that temperature response of a motor is indeed a first order time response, where comparing the T_{ss} and τ is sufficient. If not all, we need to test and create a database of motors that are first order in terms of temperature response. The second study that we need is to identify the types of fault that can be distinguished using the first order temperature response characteristics, and classify them.

Acknowledgements The financial support by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS) FRGS/1/2018/TK03/UMP/02/27 as well as facilities from Universiti Malaysia Pahang are gratefully acknowledged.

A Appendix

Website to retrieve the script to deduce the steady-state temperature shown in Fig. 4 and its analysis: <https://sites.google.com/ump.edu.my/mahrasid/application-software/algorithm-to-deduce-steady-state-temperature-from-temperature-response-dat>.

References

1. Eichhammer W, Fleiter, T (2012) United Nations Industrial Development Organization: energy efficiency in electric motor systems: technology, saving potentials and policy options for developing countries
2. Basak D, Tiwari A, Das SP (2006) Fault diagnosis and condition monitoring of electrical machines a review. In: Proceedings of the IEEE international conference on industrial technology, pp 3061–3066
3. Nandi S, Toliyat HA, Li X (2005) Condition monitoring and fault diagnosis of electrical motors—a review. IEEE Trans Energy Conv 20(4):719–729. <https://doi.org/10.1109/tec.2005.847955>. <https://dx.doi.org/10.1109/tec.2005.847955>
4. Power Systems Reliability Subcommittee of the Power System Engineering Committee: IEEE recommended practice for the design of reliable industrial and commercial power systems. Gold Book(1997)

5. Xiao N (2016) Adaptive feature extraction based on Stacked Denoising Auto-encoders for asynchronous motor fault diagnosis. In: Proceedings - 2016 9th international congress on image and signal processing, pp 854–859
6. Salem SB, Salah M, Bacha K, Chaari A (2017) Experimental investigation of the eccentricity impact on the line current spectrum for induction motors fault diagnosis purposes. In: 2016 17th international conference on sciences and techniques of automatic control and computer engineering, pp 205–210
7. Huang SR, Huang KH, Chao KH, Chiang WT (2016) Fault analysis and diagnosis system for induction motors. *Comput Electr Eng* 54:195–209
8. Delgado-Arredondo PA, Morinigo-Sotelo D, Osornio-Rios RA, Avina-Cervantes JG, Rostro-Gonzalez H, de Jesus Romero-Troncoso R (2017) Methodology for fault detection in induction motors via sound and vibration signals. *Mech Syst Sig Process* 83:568–589. <https://doi.org/10.1016/j.ymsp.2016.06.032>. <https://dx.doi.org/10.1016/j.ymsp.2016.06.032>
9. Redón P, Picazo-Ródenas M, Romero-Troncoso RJ, Antonino-Daviu J (2017) Reliable methodology for online fault diagnosis in induction motors using passive infrared thermography. In: 2017 IEEE 11th International symposium on diagnostics for electrical machines, power electronics and drives
10. García-Ramírez AG, Morales-Hernández LA, Osornio-Rios RA, Benitez-Rangel JP, García-Pérez A, de Jesus Romero-Troncoso R (2014) Fault detection in induction motors and the impact on the kinematic chain through thermographic analysis. *Electr Power Syst Res* 114:1–9. <https://doi.org/10.1016/j.epsr.2014.03.031>. <https://dx.doi.org/10.1016/j.epsr.2014.03.031>
11. Rahman NA, Gu L, Bostanci E, Fahimi B (2016) Temperature estimation of switched reluctance machines using thermal impulse response technique. In: 2016 IEEE conference on electromagnetic field computation (CEFC), p 1
12. Rasid MAH, Ospina A, Benkara KB, Lanfranchi V (2021) A 3D thermal model of SynRM with segmented rotor considering anisotropic conductivity. In: 2021 IEEE 12th energy conversion congress & exposition – Asia, pp 1816–1822
13. Rasid MAH et al (2018) Preliminary thermal evaluation of actuator for steer-by-wire vehicle. *IEEE Trans Veh Technol* 67:11468–11474
14. Abdullah NF, Wahab AA, Rasid MAH (2018) Study of temperature rise of small brushed DC motor under different load. *MATEC Web Conf* 225:03010–03010. <https://doi.org/10.1051/mateconf/201822503010>. <https://dx.doi.org/10.1051/mateconf/201822503010>
15. Wahab AA, Abdullah NF, Rasid MAH (2018) Mechanical fault detection on electrical machine: thermal analysis of small brushed DC motor with faulty bearing. *MATEC Web Conf* 225:05012–05012. <https://doi.org/10.1051/mateconf/201822505012>. <https://dx.doi.org/10.1051/mateconf/201822505012>
16. Rasid MAH, Ospina A, Benkara KEK, Lanfranchi V (2014) Thermal model of stator slot for small synchronous reluctance machine. In: 2014 international conference on electrical machines (ICEM), pp 2199–2204
17. North-Holland (2003) Chapter 1 - The Stefan problem and its classical formulation. In: Gupta SC (ed) North-Holland series in applied mathematics and mechanics, vol 45. Elsevier, pp 1–38. [https://doi.org/10.1016/S0167-5931\(03\)80004-9](https://doi.org/10.1016/S0167-5931(03)80004-9)

Intelligent Optimization of Novel Particle Swarm Optimization with Explorer (PSOE) for Identification of Flexible Manipulator System



Hanim Mohd Yatim, Ahmad Nur Yussuf Zamri, Muhamad Sukri Hadi, Mat Hussin Ab. Talib, and Intan Zaurah Mat Darus

Abstract Flexible manipulator is widely used in robotics and mechanical systems. Its application have led to the development of systems which are lighter, less bulky, and provides greater system flexibility. However, the flexible manipulator has one drawback. It develops unwanted vibration during operation which reduced the efficiency of the flexible manipulator systems for accurate positioning requirements. Therefore, an intelligent optimizer, the Particle Swarm Optimization with Explorer (PSOE) was developed to model this highly non-linear and complex system. Initially, an experimental setup for the flexible manipulator was developed. Experimental input output data were acquired including hub angle and endpoint acceleration to fed into system identification method. Next, optimization was done using the proposed PSOE as compared to a standard Particle Swarm Optimization (PSO) algorithm via linear auto regressive with exogenous (ARX) model structure. Validations of the algorithms were attained on the basis of minimizing the value of mean-squared error (MSE) and correlation tests. The superiority of the added ‘explorer’ to the algorithm was confirmed as PSOE obtained the lowest MSE value of 2.8232×10^{-5} and 3.7364×10^{-7} for end-point acceleration and hub angle modelling, respectively. Both modelling also achieved good correlation values within the 95% confidence interval. Results obtained can be adapted for further analysis in implementing an active vibration control for flexible manipulator systems.

Keywords Flexible manipulator · System identification · Particle Swarm Optimization

H. M. Yatim (✉) · A. N. Y. Zamri · M. H. Ab. Talib · I. Z. M. Darus
School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: hanim.my@utm.my

M. S. Hadi
Faculty of Mechanical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

1 Introduction

The manipulator is a critical component of any robotic system. It enables the robotic system to perform its function, influence its environment, or accomplish its objective. Without a manipulator installed, development of robotic arms is not feasible. While rigid manipulators are capable of performing manipulative tasks, they are not the optimal solution in some applications since we now have another alternative to consider: flexible manipulators. When designing a robot arm for space exploration, for example, a flexible manipulator is the ideal choice since it is less bulky and lighter than a rigid manipulator [1].

However, there is one issue that must be resolved, and that is vibration acting upon the system. Due to the fact that the manipulator is flexible, it is susceptible to vibration. The inertia acting upon slender components such as the flexible manipulator becomes a problem as it results in vibration especially at the manipulator's endpoint, thus reducing the system's efficacy while executing the task [2]. If we want to utilize flexible manipulators in mechanical systems and harness its advantages, vibration suppression of the system is required. In order to effectively suppress vibration at precise positioning of the flexible manipulator, an accurate modelling to capture and understand the dynamics of the system needs to be developed [3].

Therefore, a system identification (SI) method was utilized to model this complex system. Using this method, a mathematical model is constructed which equivalent to the identified system [4]. With employment of intelligent optimization algorithm, system identification method has been extensively used and proven superior in solving identification problems. Presently, optimization alternatives using various evolutionary algorithms (EA) have become a trend due to its capability of solving global optimizations. Among them, gaining notable attention from many researchers is the Particle Swarm Optimization (PSO) that has shown to be promisingly suitable for both scientific and engineering applications [5–7].

The main advantages of PSO are it promotes easy implementation as well as its and fast convergence because it requires only a few parameters to adjust. Hence, PSO offers global optimization which can be effectively applied for system identification problems where the conventional search methods fail. Numerous research papers have highlighted the applications of PSO to solve system identification problems for linear and non-linear systems. Omar *et al.* [8] proposed identification of parameters in Permanent-Magnet Synchronous Machines (PMSMs) using PSO. The advantages of PSO algorithm was demonstrated with its ease of implementation, quick and stable convergence characteristics. Additionally, Sangdani and Tavakolpour [5] studied a PSO identification of a target tracker robot with flexible joints. Results revealed that the elastic joints were successfully modelled by the PSO-based identification in comparison with experimental data.

However, it is reported that PSO might undergo a stage of stagnation as there is loss of iteration and the swarm loses its diversity. Zhang and Xia [9] studied an improved algorithm of PSO to identify the parameters of nonlinear dynamic hysteretic models. They introduced a new way to translate inertia factor for global

exploration and prevent problems from being stuck at a local minima. The effectiveness of the proposed PSO compared to original PSO for corresponding system were highlighted.

A new approach of Particle Swarm Optimization with Explorer (PSOE) was developed and utilized for identification of a single-link flexible manipulator. Previously, Mohd Yatim [10] has proven that PSOE successfully modelled a flexible manipulator system. However, the input output data for identification was fed from a simulated environment using a finite difference model. Later, an experimental rig consisting of the flexible manipulator structure was developed. In addition to the more realistic experimental input output data with integration of sensors and actuators, it also contributed to a more complex data as disturbance from the actual environment was captured. Thus, in this study, PSOE based identification was carried out using the input–output data acquired experimentally in a real time environment. The validity of the obtained model was investigated using input/output mapping and correlation tests. This study also provides a good comparison of the proposed algorithm in identifying more complex systems acquired from an experimental environment.

2 Experimental Setup of Single-link Flexible Manipulator

An experimental rig of single-link flexible manipulator which is constrained to horizontal movement was developed. An integration between mechanical system, electric and electronics includes sensors and actuators and also a computer was set up. Experimental procedures of the flexible manipulator rig were developed in order to collect the input–output data of the system comprising of hub angle and end-point acceleration behaviors.

The flexible structure was made from a thin aluminium alloy with dimensions of $600 \times 40 \times 1.5$ mm was presented. One end of the flexible structure was rigidly attached to the motor that drive the flexible structure while the other end was free for movement. Figure 1 shows the system integration of the experimental setup of the flexible manipulator [11].

3 Particle Swarm Optimization (PSO)

PSO was originally formulated by Kennedy and Eberhart in 1995 which was inspired from the movement behaviour of schooling fishes and flocking birds [12]. Presently, PSO algorithm has been widely accepted and progressively developed for easy implementation and simple in terms of concept.

PSO is started with a swarm of random particles that ‘fly’ in the search space of an optimized solution. Unlike other nature-inspired algorithms, the evolution in PSO is driven by cooperation and competition among particles through iterations of the best position a particle has visited so far, *pbest* (P_{id}) and a global best or *gbest*

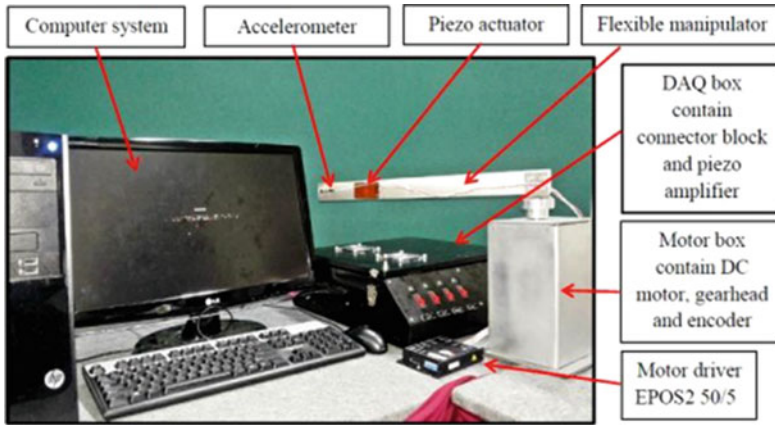


Fig. 1 Experimental setup for flexible manipulator structure [11]

(P_{gd}), obtained so far by any particle within the population. Information is shared among particles, and subsequently the particles are accelerated towards those two best values. Particles position and velocity are updated at each iteration using the following set of equations [13]:

$$v_{id}(t) = wv_{id}(t-1) + C_1 rand(P_{id} - x_{id}(t-1)) + C_2 rand(P_{gd} - x_{id}(t-1)) \quad (1)$$

$$x_{id}(t) = x_{id}(t-1) + v_{id}(t) \quad (2)$$

$v_{id}(t)$ and $x_{id}(t)$ represent the current velocity and position vector of the i -th particle in the d -dimensional search space respectively. C_1 is cognitive component which promotes the tendency of particles to return to their own previously found best positions and C_2 is called the social component which is identified as the cooperative effect of a particle flying towards the global solution, usually $C_1 = C_2 = 2$ [12]. $rand$ is the random number between 0 and 1. w is the velocity vectors or inertia that prevents the particle from drastically changing direction which serves as a memory of the previous fly direction. An equivalence coverage of particles between global exploitation and local search from the values of inertia is important to prevent loss of potential solutions. Prior to that, linearly decreasing the inertia over time is a good option where balance between global and local search can be achieved, expressed as

$$w(t) = w_{start} - \frac{(w_{start} - w_{end})}{T_{max}} \quad (3)$$

w_{start} and w_{end} defined as the start and end of inertia weight by linearly decreasing from 0.9 to 0.25. T_{max} is the maximum number of time step the swarm is allowed to search.

4 Particle Swarm Optimization with Explorer (PSOE)

PSO are capable of discovering reasonable and quality solutions in short time as compared to others classical evolutionary algorithms. However, despite its promising characteristics, it has known to lose its global exploring capabilities as iteration goes on and only flying within a small area. In this study, new elements are introduced to the standard PSO in order to overcome the weaknesses to improve the overall performance of the algorithm.

Taking inspiration from the Artificial Bees Colony (ABC), a modified algorithm was proposed in this study named PSO with explorer (PSOE). An ‘explorer’ is added in order to convey a similar behavior as the scout bees in ABC.

PSO can easily to be trapped into the local minima as iteration goes on. It may stop global exploration and only fly within a small area. When this happen, the whole swarm is directed into an equilibrium state which stagnates the evolution.

Hence, introduction of the explorer characteristic may satisfy the requirement to prevent the swarm achieving an equilibrium state. The explorer works by drawing the velocity component from a uniform distribution. This action results improves the diversity of the swarm and increase ability to explore new potential positions. Sustainable development using explorer seems promising due to its ability to share information in PSO algorithm. However, in order to avoid loss of good particles, the explorers will only take place after a certain limit of iteration is achieved. Particles will be given their *pbest* value so far and a new velocity to explore a new potential search space as in Eq. (1). Figure 2 depicts the pseudo code of PSOE.

From Fig. 2, errors between the fitness of *i*-th particle with the fitness of *gbest* is defined as $\Delta Fitness$. ξ is the constant according to the precision requirement which equals to 0.0001, array of *count(i)* is to memorize the counts that satisfy the conditions of $|\Delta Fitness(i)| < \xi$ for the *i*-th particle. *limit* is the limit of iteration that *i*-th particle has uniform fitness error which is set at 4. If the particle has a progressive fitness, thus it has less chances to become an explorer and will maintain its potential position. Figure 3 shows the flowchart of PSOE with the procedure of explorer as highlighted.

```

FOR i = 1: swarmsize ; for each particle
    IF  $|\Delta Fitness(i)| < \xi$ 
        THEN count(i) = count(i) + 1 ; add 1
        ELSE count(i) = 0 ; reset
    END
; Memorize the particles' position that has most count
IF count(i) >= limit
    THEN reinitialized (that i-th particle) ; execute new position to the
particle
    ELSE count(i) = 0 ; reset
    END
END

```

Fig. 2 Additional pseudocode of PSOE

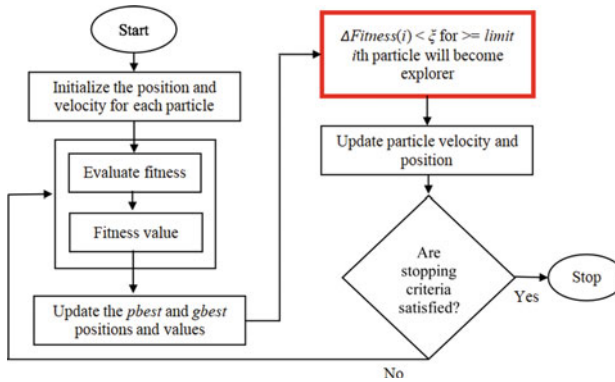


Fig. 3 Flowchart of PSOE

In this study, the performance of PSOE was observed in comparison with standard PSO modelling the flexible manipulator system. Experimental input output data includes hub angle and end-point acceleration behaviors. Modelling will be conducted using system identification via ARX model [4].

5 Results and Discussions

System identification of a flexible manipulator system via ARX structure was optimized using PSOE and PSO to obtain its approximate model. A total of 14,000 endpoint acceleration data with 2000 data chosen for training while the remainder was used for testing. Additionally, 8540 hub angle data was observed with 1600 data utilized for training and the rest was used for testing. Performance of PSOE-based identification was assessed in input/output mapping, lowest MSE value and correlation test in comparison to its standard counterpart.

5.1 Modeling Using Particle Swarm Optimization with Explorer (PSOE)

Experimental input output data of hub angle and end-point acceleration were fed into the PSOE-based identification. Then, it was tuned heuristically by changing the number of particles, number of iterations and model orders. The best results with the smallest MSE were captured.

PSOE-based identification achieved an order of 2 for both modellings and their best results are depicted in Table 1. Figure 4 shows hub angle modelling output using PSOE in time domain. Figures 5 and 6 illustrated pole zero stability diagram

Table 1 Comparative assessment

	Modeling domain	No. of particles	No. of iterations	MSE	Transfer function
Hub angle	PSOE	60	200	3.7364×10^{-7}	$H_1(z) = \frac{-2.209z^{-1} + 2.209z^{-2}}{1 - 1.26z^{-1} + 0.261z^{-2}}$
	PSO	100	300	5.2837×10^{-7}	$H_2(z) = \frac{0.01545z^{-1} - 0.01545z^{-2}}{1 - 3.048z^{-1} + 2.048z^{-2}}$
Endpoint acceleration	PSOE	80	200	2.8232×10^{-5}	$H_3(z) = \frac{0.03863z^{-1} + 0.06286z^{-2}}{1 - 1.491z^{-1} + 0.6943z^{-2}}$
	PSO	80	300	2.8236×10^{-5}	$H_4(z) = \frac{0.03563z^{-1} + 0.06609z^{-2}}{1 - 1.495z^{-1} + 0.6991z^{-2}}$

Fig. 4 PSOE modelling and actual output in time domain for hub angle

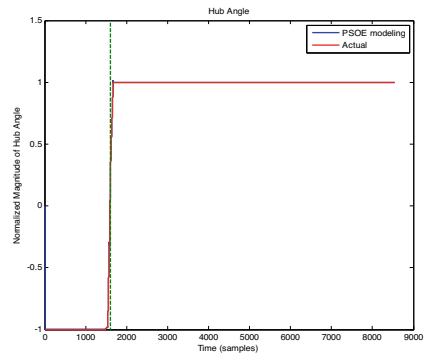


Fig. 5 Pole zero diagram of PSOE modelling for hub angle

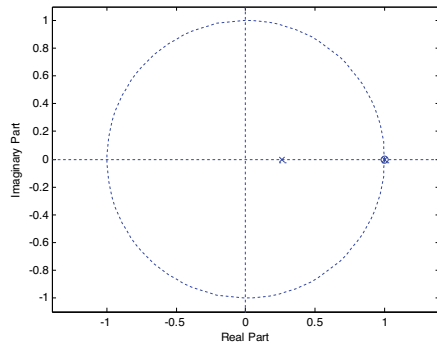
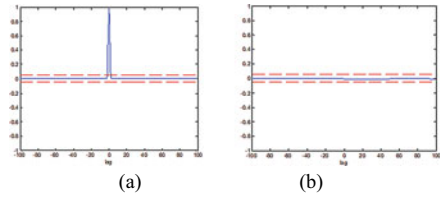


Fig. 6 Correlation tests of PSOE modelling for hub angle **a** auto correlation, **b** cross correlation



and correlation test, respectively. On the other hand, for end-point acceleration modelling, Figs. 7 and 8 represents time and frequency domains output, respectively. Figures 9 and 10 shown the pole zero mapping and its corresponding correlation test, respectively.

It is noted that PSOE modelling output closely matches the actual output in time domain for both hub angle and end-point acceleration as depicted in Figs. 4 and 7. The dynamics of the system was also successfully characterized by the PSOE algorithm as illustrated in frequency domain in Fig. 8. Results show that a stable model was achieved as all the poles lies within the circle unit as seen in Figs. 5

Fig. 7 PSOE modelling and actual output in time domain for end-point acceleration

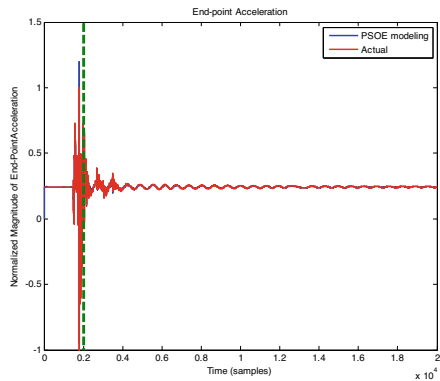


Fig. 8 PSOE modelling and actual output in frequency domain for end-point acceleration

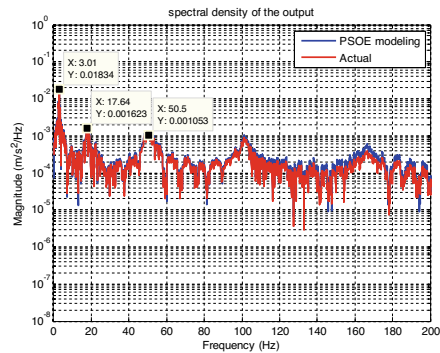


Fig. 9 Pole zero diagram of PSOE modelling for end-point acceleration

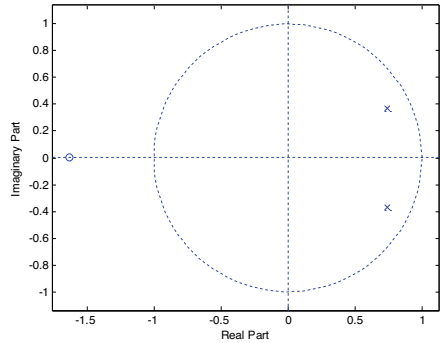
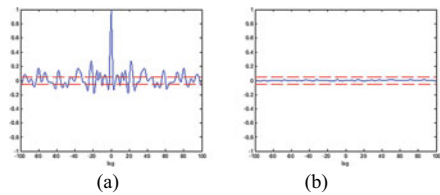


Fig. 10 Correlation tests of PSOE modelling for end-point acceleration **a** auto correlation, **b** cross correlation



and 9. Figures 6 and 10 represents unbiased correlation tests for both hub angle and end-point acceleration with 95% confidence interval.

5.2 Modeling Using Particle Swarm Optimization (PSO)

As a comparison, PSO was employed to model the flexible manipulator system. The optimum results acquired are depicted in Table 1 which achieved an order of 2 for both hub angle and end-point acceleration modelling.

For hub angle, the PSO modelling output in time domain is illustrated in Fig. 11. The pole zero mapping is shown in Fig. 12 while the correlation test is depicted in Fig. 13. Whereas, for end-point acceleration, the PSO modelling output in time and frequency domains are represented by Figs. 14 and 15, respectively. Figure 16 shown the pole zero stability and Fig. 17 illustrates the correlation test result.

5.3 Comparative Assessment

The performance achieved by both PSOE and PSO algorithms are represents in Table 1. It depicts the lowest MSE obtained at optimised particle position and iterations and also a model transfer function for both hub angle and end-point acceleration.

Fig. 11 PSO modelling and actual output in time domain for hub angle

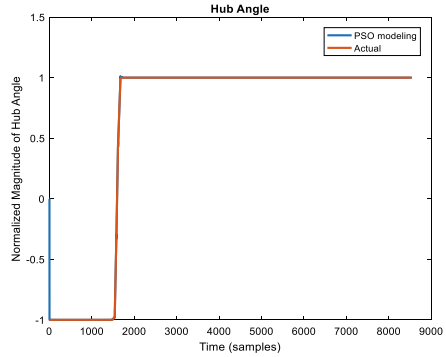


Fig. 12 Pole zero diagram of PSO modelling for hub angle

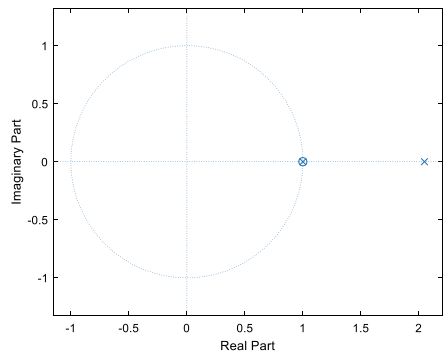
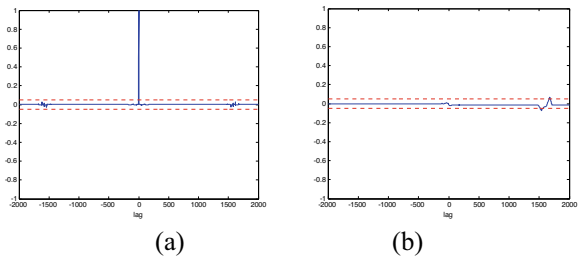


Fig. 13 Correlation tests of PSO modelling for hub angle: **a** auto correlation, **b** cross correlation



From Table 1, it was observed that PSOE was able to achieve the lowest MSE as compared to PSO. In addition, PSOE also achieved an optimum solution within a smaller swarm of particles with less iterations which indicates a simpler resolution coupled with faster convergence. This may due to the advantages of the explorer feature itself as if affects the parameters convergence which can be observed in Fig. 18.

The incorporation of the explorer feature within the algorithm make it possible for convergence to be active and increase diversity. As can be seen from Fig. 18(c), the new exploration of the explorer feature is not too random and fluctuates because

Fig. 14 PSO modelling and actual output in time domain for end-point acceleration

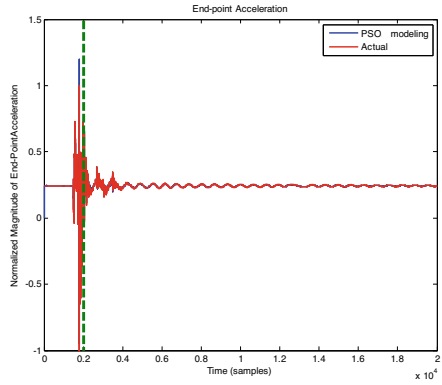


Fig. 15 PSO modelling and actual output in frequency domain for end-point acceleration

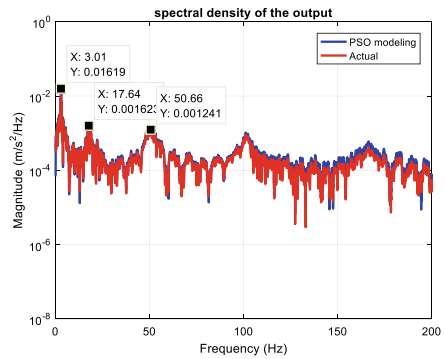


Fig. 16 Pole zero diagram of PSO modelling for end-point acceleration

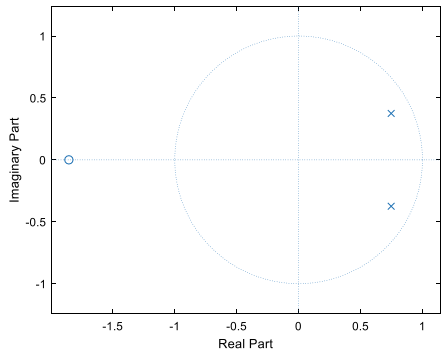
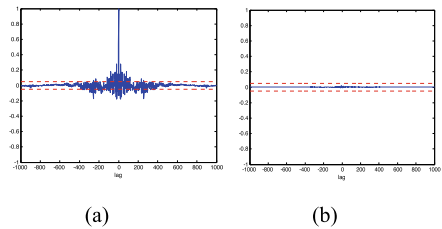


Fig. 17 Correlation tests of PSO modelling for end-point acceleration: **a** auto correlation, **b** cross correlation



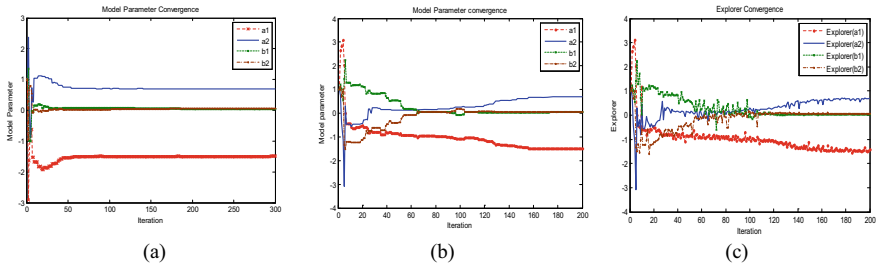


Fig. 18 Parameter convergence **a** in PSO **b** in PSOE **c** explorer convergence in PSOE

it is bound within its own previous *pbest* positions. Yet, particles are kept away from being trapped whilst promoting new potential solutions within its flying trajectory. From Fig. 18(a), it was noted that the model parameter of PSO already converged at an early stage. There is a high possibility of premature convergence which led stagnation at an equilibrium state. Figures 18(b) and (c) show the parameter convergence and explorer trajectory in PSOE. As particles start moving towards an optimum solution, passive particles will become an explorers to avoid stagnation which adds diversity to the swarm. As in Fig. 18(b), model parameters have an active path and only converge after their 100th iterations.

Therefore, superiority of PSOE can be seen in both hub angle and end-point acceleration modelling where the lowest MSE value was achieved with a simpler model that exhibits faster convergence, acceptable stability and unbiased correlation tests. Results concludes that PSOE is proven to perform well for complex experimental data of flexible manipulator system. This provides a good benchmark that can be used for future analysis of control methodologies.

6 Conclusion

In this study, Particle Swarm Optimization with Explorer (PSOE) was employed to model a single link flexible manipulator system in comparison with its standard PSO. Input output data were acquired experimentally for hub angle and end-point acceleration behaviors and fed to the system identification. PSOE and PSO models that are estimated were verified to be used for further analysis. From the results, it was established that the superiority of the added explorer feature in the PSOE in estimating the system behavior led to lower MSE value as well as experiencing faster convergence and improved diversity for better optimization.

Acknowledgements The authors wish to thank the Ministry of Higher Education and Universiti Teknologi Malaysia (UTM) for providing the research grant and facilities. This research was supported using the UTM Research Grant GUP Tier 2 Vote No. V17J07

References

1. Dwivedy SK, Eberhard P (2006) Dynamic analysis of flexible manipulators, a literature review. *Mech Mach Theory* 41(7):749–777
2. Zhao ZL, Qiu ZC, Zhang XM, Da Han J (2016) Vibration control of a pneumatic driven piezoelectric flexible manipulator using self-organizing map based multiple models. *Mech Syst Signal Process* 70–71:345–372
3. Mejerbi M, Zribi S, Knani J (2018) Dynamic modeling of flexible manipulator based on a large number of finite elements. In: *International conference on advanced systems and electric technologies*, pp 357–362
4. Baseri HH, Yatim HM, Hadi MS, Talib MHA, Darus IZM (2022) Firefly algorithm for modeling of flexible manipulator system. *Recent Trends Mechatron Towards Ind* 4:235–251
5. Sangdani MH, Tavakolpour-Saleh AR (2020) Particle swarm optimization based parameter identification applied to a target tracker robot with flexible joint. *Int J Eng* 33(9):1797–1802
6. Hadi MS, Darus IZM, Yatim HM (2013) Modeling flexible plate structure system with free-free-clamped-clamped (FFCC) edges using Particle Swarm Optimization. In: *Symposium on computers & informatics (ISCI)*, pp 39–44
7. Durmuş B, Gün A (2011) Parameter identification using particle swarm optimization. In: *International advanced technologies symposium (IATS 2011)*, pp 188–192
8. Sandre-Hernandez O, Morales-Caporal R, Rangel-Magdaleno J, Peregrina-Barreto H, Hernandez-Perez JN (2015) Parameter identification of PMSMs using experimental measurements and a PSO algorithm. *IEEE Trans Instrum Meas* 64(8):2146–2154
9. Zhang J, Xia P (2017) An improved PSO algorithm for parameter identification of nonlinear dynamic hysteretic models. *J Sound Vib* 389:153–167
10. Mohd Yatim H, Mat Darus IZ, Hadi MS (2014) Modeling of flexible manipulator structure using particle swarm optimization with Explorer. In: *IEEE symposium on industrial electronics & applications (ISIEA)*, pp 113–118
11. Mohd Yatim H, Mat Darus IZ (2017) Development of an experimental single-link flexible manipulator system. *Int J Eng Technol* 7:7–12
12. Settles M (2005) *An introduction to particle swarm optimization*. Department of Computer Science, University of Idaho
13. Hassan R, Cohanim B, De Weck O, Venter G (2005) A comparison of particle swarm optimization and the genetic algorithm. In: *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*

SCADA System for Industrial Manipulator PLC Trainer



Nur Hanisah Azman, Muchamad Oktaviandri,
and Muhammad Rizal Ramedan

Abstract In most industrial application, the implementation of SCADA which consists of an HMI, computer system monitoring, data acquisition and processing, and advanced visualization is able to make ease for the industries. It is common to utilize a computer to control an automated system from the microcontroller such as PLC, Arduino and Raspberry Pi. These controllers must be able to receive and analyze signals in any way possible. The main aim of this research is to establish a communication between the Arduino microcontroller and the field device located at the Industrial Trainer for the purpose of data monitoring. This trainer implemented pneumatic based system with the presence of additional field device which is the infrared sensor, ultrasonic sensor and load cell. Then, the Arduino reads the parameter for real-time value for distance, weight, force, pressure from the ultrasonic and load cell sensors and sends the value to the MySQL database through a LAN network. The user interface also receives the input value from the sensors through the serial port communication. The results obtained will validate either the system is reliable or not through the comparison on the accuracy of the live data recorded either it is tally with the data that is exported in excel file result and the data that is stored in database result.

Keywords SCADA · PLC trainer · Arduino

N. H. Azman · M. Oktaviandri (✉) · M. R. Ramedan
Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang,
Pekan 26600, Malaysia
e-mail: muchamad_oktaviandri@yahoo.com

M. Oktaviandri
Fakultas Teknik, Universitas Pembangunan Nasional Veteran Jakarta, Jakarta Selatan 16515,
Indonesia

Fakultas Teknologi Industri, Universitas Bung Hatta, Padang 25143, Indonesia

1 Introduction

A new era, also known as Industrial Revolution generation was inaugurated at the dawn of nineteenth century. This revolution gives a big impact to the global economy especially in industrial world [1]. Historically, the world has undergone three major growth period, Industry 1.0, Industry 2.0 and Industry 3.0 where disruptive technology has brought about huge productivity improvements especially in automation world. The latest Industrial Revolution 4.0 that currently takes place around this time is the kick start of the fully automated cyber-physical system. Simply put, Industry 4.0 aims at enhancing and developing innovative and outstanding systems towards a better management and control of industrial system. Machines are improving and keep getting smarter as they are able to access more data without the control supervision from human physically.

Recently, there has been a rapid increment in the research and development of control and automation technology for industrial application to meet the evolving needs of industry in the control systems [2]. As control in most industrial automation applications used the framework to run specific device program, these involve controllers like Programmable Logic Controllers (PLCs), Distributed Control System (DCS), Supervisory Control and Data Acquisition System (SCADA) and more [3]. With the rapid development of technology utilized in modern cycles of industrial processes, the automatic control system framework has become the mandatory basis of these developments [4]. Through this approach in controlling and computerize any modern framework, PLC has been extensively utilized in industrial production, automation industry and control system these days [5].

SCADA framework normally utilize using PLC using various methodologies. A prototype of cutting machine test bench that implement pneumatic controls, with the help of microcontroller Arduino, servo motor control and PLC/SCADA control is built [6]. Arduino as the main programmer for servo motor controller which ensure the smooth run of conveyer and control the accuracy of positioning of the piece to push the material forward. [7] proposed a new technique of integrating sensors with serial communication and a PLC using Arduino. The PLC is physically connected to the Arduino through a LAN connection and an ethernet shield. It is feasible to establish a multi-point remote monitoring system with this ethernet shield and an ethernet hub or switch.

TCP communication between Arduino and PLC has been developed with the presence of Ethernet shield to allow third-party devices to be connected to the PLC and saves I/O PLC modules [8]. Then, the data from ultrasonic and line-following sensors is pre-processed by the Arduino and combined into a single data packet that is sent over the network to the PLC server. Its sensor compatibility aids in providing the necessary feedback, and the Arduino controls the system by taking corrective action. It boosts product rate and decreases product time when compared to other traditional systems.

A PLC-based boiler automation system that compensates for the lack of a micro-processor is developed [9] with the employed of a sensor to measure the temperature

and control it by sending a signal to a PLC, which automatically starts and stops the boiler system based on the set value. Previous research created a mathematical model for industrial wastewater purification that included a PLC and advanced control application. The right level of purification is required for a satisfactory result and [10] successfully demonstrated a method for coal mine subsurface drainage utilizing an AT89S52 microprocessor.

For packaging line, a conveyor belt transports products to a packaging location is controlled by a Siemens LOGO! 230 RCE PLC [11]. The relays are controlled by a microcontroller (Arduino Uno) based on the output of Infra-red sensors that perceive the products. The relay states are fed into the PLC as an input. Remote control and monitoring are available through the interface. By replacing the commonly used supervisory control and data acquisition (SCADA) system with standard sensors and mobile application-based control, the proposed approach succeeds to reduce total costs.

Therefore, industry of today only needed human labor and operators to supervise and monitor the system operated in machines and devices from afar. Therefore, this thesis reports the findings on the fundamental principles and terminology used in the field of SCADA system. The research explores through the role of SCADA system in automation implementation on Industrial Manipulator PLC Trainer interfaced with Arduino based on various studies which best describes a SCADA based architecture for encompasses the collecting of the information, carrying out any necessary analysis and controlling and displaying that live information on a number of operator screens or displays.

2 Methodology

The most important step for the successful completion of this study is the technical planning of the project. Methodology explains the approach taken to setup the SCADA system for the manipulator. In developing a system, it is crucial to study on the best software options that should be emphasizes in producing the best system. Therefore, the proposed methodology will cover and discuss the best technique to be implemented in developing a system.

2.1 Proposed Methodology

The method proposed in completing this research is divided into three different stages as per Fig. 1. Phase I starts with the hardware configuration of the PLC and the setup of additional electrical components which includes ethernet shield and field devices. Since the hardware for the manipulator is readily installed, but still each connection needed to be checked out and reassured that there are no damaged wires which can cause short circuit for safety purposes. Moreover, with the presence of Arduino

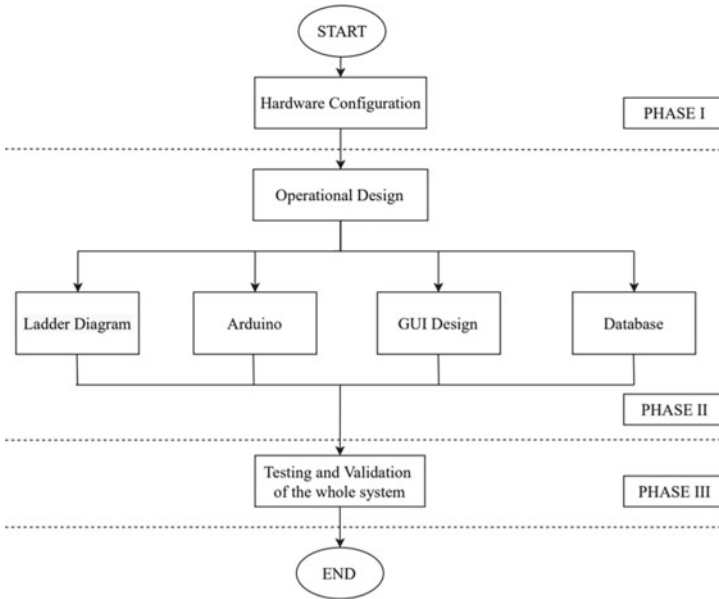


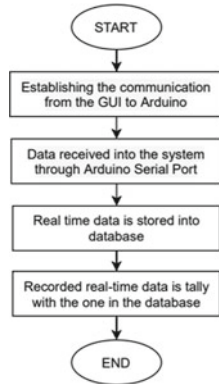
Fig. 1 Methodology for the development of SCADA system

Ethernet Shield, it is crucial to keep a detailed eye on the configuration from the PLC to the Arduino. The Arduino Ethernet Shield receive signals from the PLC and enable the communication in the SCADA system.

Phase II consist of operational design. The design in this system will be divided into four main programs. For PLC, ladder diagram is chosen to configure the work flow of the manipulator. For Arduino Ethernet Shield, the program is done in Arduino open-source software. GUI Design will make use of the Microsoft Visual Studio software for user interface. Next, database is setup in case of unexpected tragedies.

Phase III is to test the whole SCADA system and validate that the system is able to run smoothly like how it is intended to. The hardware allocation of the input and output must be accurate and tally with the programming sequel. After each designation, a test is done to ensure that each one of them is able to operate. If there is any mistakes or error encounter during the testing, an improvement is made to ensure the smooth operation of the system. In sequel to prove that the system is validated, a user should try to implement the system and their opinion should be taken in. At the end of the research, the database should be able to carry out to ensure that the system is able to preview the real-time monitoring analysis.

Fig. 2 Testing and validation of the system



2.2 Testing and Validation of the Whole System

In general, system testing is a level of testing that validates the complete and full integrated system. The purpose of this test is to evaluate the end-to-end system specifications. System testing will cover each part of the operational design competency. To verify that the system is applicable for the users, several test should be done as per Fig. 2.

The testing list starts with the inspection of the communication established on how they interact with one another and with the system itself as a whole. Then to validate that the system is user-friendly, every data received into the system window should be able to be previewed as real-time data. Lastly, the real-time data recorded in the system should be tally with the one stored in the database.

3 Result and Discussion

3.1 Hardware Configuration

The hardware specification for the Arduino is as per Table 1. As a safety precaution, it is important to identify the specification of the controller used to ensure they are suitable to be used for the purpose of developing a system.

The hardware configuration for this SCARA 3-axis Robotic Arm Industrial Manipulator is already assembled and installed in place. A small change is made with the presence of electrical components include Arduino Uno, Ethernet Shield and field devices, infrared sensor, ultrasonic sensor and load cell sensor. Figure 3 shows the circuit diagram for the additional sensor connection to the microcontroller. This amendment is made to establish a communication in measuring the parameters for data monitoring.

Table 1 Hardware specification of Arduino

No	Specification	Arduino
1	Model	Ethernet shield
2	MCU type	ATmega328P
3	Operating voltage	5 V
4	Flash memory	32 KB
5	MCU clock speed	16 MHz
6	Input and output	28 pins
7	Communication	Serial USART, TCP/IP

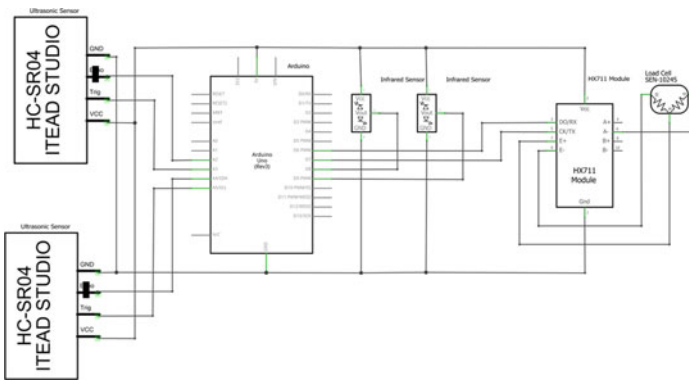


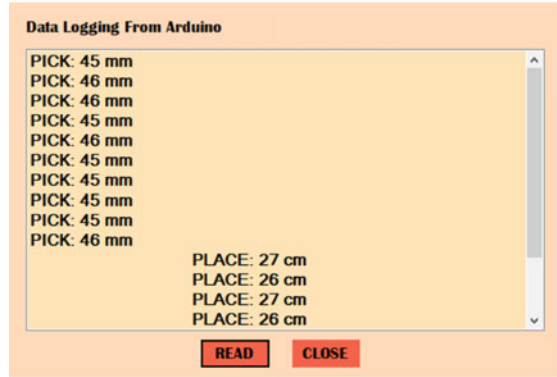
Fig. 3 Hardware connection to microcontroller

3.2 Operational Design

Arduino Uno and Arduino Ethernet Shield V2.0 play an important role in ensuring a secured communication. Arduino ethernet shield is mounted on top of Arduino Uno board. The cable is connected to the Ethernet Port of the Ethernet shield. Another importance of the presence of Arduino Ethernet Shield is to establish a connection from the system into the database. Arduino is connected to the field device and the parameter is measured from the Industrial Manipulator. The Arduino microcontroller plays an important role to send the data input into the GUI. The programming code for Arduino may seem easy but the higher the number of field devices connected, the higher the possibilities and obstacles in combining the code. Figure 4 shows the Arduino code for the established connection which needed to be tested for each field device before combining them.

Once a successful communication is established from the Arduino into the GUI, the GUI is configured as per design demanded shown in Fig. 5. GUI Flowchart resembles the SOP in machinery. A system flowchart should be as detailed as possible. The GUI consists of the parameters to be observed and measured. Starting from the Login Page, the user will have to input the mandatory username and password. If

Fig. 6 Distance measurement data logging



one complete cycle of pick and place. Therefore, if the distance measured increase uncontrollably or rapidly, the user will notice it right away and take action before it affects the whole monitoring system.

The System Management page display the real-time data monitoring system. A trend which shows a graphical form of data trending is able to be previewed once the data is exported from the database whenever they are requested as a method of data analysis. This page start viewing the live data once successful communication is established from the serial port of the microcontroller as per Fig. 7. The user may preview the specific value for each point in the graph at the exact time. The parameter measured for the real-time data includes the weight, the force and the pressure needed for the pick and place mechanism. In any case of a breakdown in factory, the stop button will avoid the system from keep on recording a redundant data. The file is then transferred into Excel sheet as per Fig. 8 to be previewed and analyzed. The data recorded in the graph is stored automatically in the database as per Fig. 9.

In terms of observing the performance of the pick and place mechanism, the pattern and trends of the real-time graph should be viewed by the user. A bad performance will show the presence of outliers in the graph which indicates the initiation of redundancy of data. Once this problem arises, analysis and troubleshooting in finding the root

Fig. 7 Real-time data logging

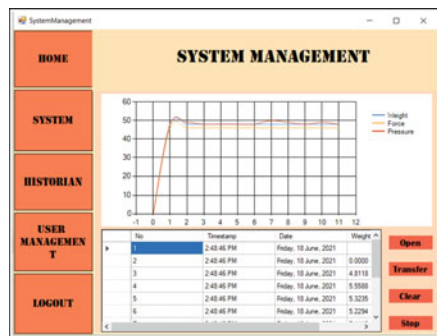


Fig. 8 Data transfer from the system into excel file

	A	B	C	D	E	F
1	No	Timestamp	Date	Weight	Force	Pressure
2	1	14:48:20PM	18/6/2021	0.0000	0.00	0.00
3	2	14:48:20PM	18/6/2021	5.3671	52.65	5.80
4	3	14:48:20PM	18/6/2021	4.8249	47.33	5.21
5	4	14:48:20PM	18/6/2021	4.9623	48.68	5.36
6	5	14:48:20PM	18/6/2021	4.8547	47.62	5.25
7	6	14:48:20PM	18/6/2021	4.8845	47.92	5.28
8	7	14:48:20PM	18/6/2021	4.8118	47.20	5.20
9	8	14:48:20PM	18/6/2021	4.7624	46.72	5.15
10	9	14:48:20PM	18/6/2021	4.8000	47.09	5.19
11	10	14:48:20PM	18/6/2021	4.7987	47.08	5.18
12	11	14:48:20PM	18/6/2021	4.8875	47.95	5.28
13	12	14:48:21 PM	18/6/2021	4.9125	48.19	5.31
14	13	14:48:21 PM	18/6/2021	4.5836	44.97	4.95
15	14	14:48:21 PM	18/6/2021	4.9871	48.92	5.39
16	15	14:48:21 PM	18/6/2021	4.5236	44.38	4.89
17	16	14:48:21 PM	18/6/2021	4.8781	47.85	5.27
18	17	14:48:21 PM	18/6/2021	4.2597	41.79	4.60
19	18	14:48:21 PM	18/6/2021	4.8876	47.95	5.28
20	19	14:48:21 PM	18/6/2021	5.0122	49.17	5.42

No	Weight	Force	Pressure	Date and Time
1	0	0.0000	0.00	2021-06-18 14:48:20
2	5.3671	52.65	5.80	2021-06-18 14:48:20
3	4.8249	47.33	5.21	2021-06-18 14:48:20
4	4.9623	48.68	5.36	2021-06-18 14:48:20
5	4.8547	47.62	5.25	2021-06-18 14:48:20
6	4.8845	47.92	5.28	2021-06-18 14:48:20
7	4.8118	47.20	5.20	2021-06-18 14:48:20
8	4.7624	46.72	5.15	2021-06-18 14:48:20
9	4.8000	47.09	5.19	2021-06-18 14:48:20
10	4.7987	47.08	5.18	2021-06-18 14:48:20
11	4.8875	47.95	5.28	2021-06-18 14:48:20
12	4.9125	48.19	5.31	2021-06-18 14:48:21
13	4.5836	44.97	4.95	2021-06-18 14:48:21
14	4.9871	48.92	5.39	2021-06-18 14:48:21
15	4.5236	44.38	4.89	2021-06-18 14:48:21
16	4.8781	47.85	5.27	2021-06-18 14:48:21
17	4.2597	41.79	4.60	2021-06-18 14:48:21
18	4.8876	47.95	5.28	2021-06-18 14:48:21
19	5.0122	49.17	5.42	2021-06-18 14:48:21
20	5.0000	50.00	5.00	2021-06-18 14:48:21
21	4.9000	48.00	5.0	2021-06-18 14:48:21

Fig. 9 Database for real-time data in MySQL

cause takes place before the problem worsen. The problem occur because of the connection issue with the controller which can affects the performance of the pick and place mechanism.

3.3 Testing and Validation of the Whole System

Validation of the system is the assessment of the system designed to ensure the system is able to meet the predefined requirements. The main purpose of the system is to be able to display the accurate value of the data. From the real-time data recorded from field device sensor integrated with microcontroller Arduino, display them in GUI, saved into database and exported into Excel sheet.

The load cell reads the weight of the object through serial communication port of Arduino. To observe the reliability of the system, the recorded data should be tally with the value displayed in GUI and database. The intriguing part normally can be seen from the accuracy of the data recorded. Since the data is tally as per Fig. 10,

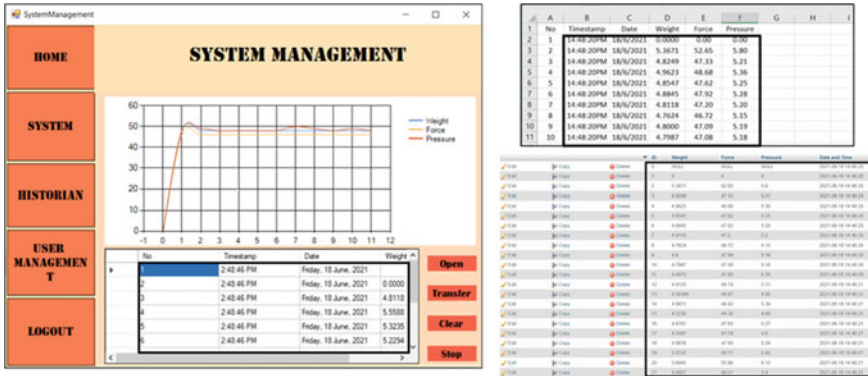


Fig. 10 Validation of the system

the system is validated as a reliable system due to the ability to produce the accurate value of data.

4 Conclusion

This paper reports the attempt to establish communication between the Arduino microcontroller and the field device, the infrared, ultrasonic and load cell sensors located at the Industrial Trainer. The program and the system are able to receive data from the input sensor through serial communication. Next, the flow of data transfer in sending and receiving those works efficiently since the established connection has been analyzed for any problem occurrence that can restrict the connection issue.

To fulfill the main purpose to enable the user to monitor the performance of the mechanism through system approach, the presence of the Historian form and User Management form was designed to give a chance for the user to observe the performance of the pick and place manipulator. If there is an outlier in the system, user may take a quick action by informing to the person in charge to analyze and figure out the problem. The recorded data which are saved to Excel sheet can be referred to easily without any problem.

Lastly, the system established succeed to analyze the data in the database to ensure it is tally and accurate with the real testing value. The real-time data from microcontroller was successfully stored in the database and any changes were updated and the database is able to display the live data and past data trending for analysis at the exact time and date.

References

1. Zawra LM, Mansour HA, Messiha NW (2019) Migration of legacy industrial automation systems in the context of industry 4.0-a comparative study. In: 2019 international conference on fourth industrial revolution (ICFIR)
2. Sahin S, Isler Y (2013) Microcontroller-based robotics and SCADA experiments. *IEEE Trans Educ* 56(4):424–429
3. Bagal KN, Kadu CB, Parvat BJ, Vikhe PS (2018) PLC based realtime process control using SCADA and MATLAB. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA)
4. Hadi HH, Sallom MY (2019) Pneumatic control system of automatic production line using SCADA implement PLC. In: 2019 4th scientific international conference Najaf (SICN)
5. Zhang W, Li Y (2020) Design of automatic feeding control system in tank area based on WinCC configuration software. In: 2020 Chinese control and decision conference (CCDC)
6. Durrani F, Riaz M, Ahmad MH, Durrani S, Ali H (2018) SCADA & PLC based fully automated pneumatic cutting machine: a test bench for industry and laboratory. In: 2018 international conference on engineering and emerging technologies (ICEET)
7. Rahadian H, Nugroho W, Izzhati DN (2016) Siemens CPU1215C input expansion and remote monitoring with Arduino Bridge. In: 2016 6th international annual engineering seminar (InAES)
8. Kajan M, Sovcik J, Duchon F, Mrafko L, Florek M, Beno P (2014) Sensoric subsystem of automated guided vehicle: TCP communication between SIMATIC S7 PLC and Arduino. In: 2014 23rd international conference on robotics in Alpe-Adria-Danube Region (RAAD)
9. Tahsin Labib SM, Ul Alam S, Hossain S, Hossain Patwary MI, Ahmed R, Islam MA (2019) Design and implementation of boiler automation system using PLC. In: 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)
10. Ying-mei Z, Jing-qian Z (2007) Design of coal mine drainage system based on the control of microcontroller. [En.cnki.com.cn](http://en.cnki.com.cn)
11. Mofidul RB, Sabbir MSH, Podder AK, Shaifur Rahman M (2019) Design and implementation of remote controlling and monitoring system for automatic PLC based packaging industry. In: 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)

A Self-adaptive Hybridized Algorithm with Intelligent Selection Scheme for Global Optimization



Zhi Chuan Choi, Koon Meng Ang, Cher En Chow, Wei Hong Lim, Sew Sun Tiang, Chun Kit Ang, and Balaji Chandrasekar

Abstract Hybrid optimization algorithms were proposed and widely implemented to solve various types of complex global optimization in the past decades. However, the random guess mechanism of the conventional initialization scheme adopted in existing hybridized approaches, tends to stochastically generate the initial population with questionable quality. In this article, a new hybrid optimization optimizer namely self-adaptive hybridized DE with PSO (SADE-PSO) is introduced to regulate the balancing of explorative and exploitative searches to deal with distinct global optimization more efficacious. The self-adaptive scheme is firstly designed into SADE-PSO, in order to propose better initial population. A hybridization framework is employed in SADE-PSO to amalgamate the superiorities of DE and PSO at the same time and facilitate the optimization manifestation with better quality. Extensive simulation results of the proposed SADE-PSO are performed among all five competitors using CEC 2014 test functions. The novel SADE-PSO is observed to have the best search performance due to its capability in tackling majority test functions with 22 best F_{mean} values.

Keywords Self-adaptive mechanism · Intelligent selection scheme · Hybridization · Particle Swarm Optimization (PSO) · Differential Evolution (DE)

1 Introduction

In the fourth Industry Revolution era, the complex mathematical modelling has gained significant attentions in deploying various engineering systems. Most frequently, metaheuristics search algorithms (MSAs) have yielded as a promising optimizer to tackle different real-world engineering optimization problems [1–6] by

Z. C. Choi · K. M. Ang · C. E. Chow · W. H. Lim (✉) · S. S. Tiang · C. K. Ang
Faculty of Engineering, Technology and Built Environment, UCSI University,
56000 Kuala Lumpur, Malaysia
e-mail: limwh@ucsiuniversity.edu.my

B. Chandrasekar
Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology,
Chennai, Tamil Nadu 603203, India

employing the working mechanisms inspired by various natural phenomena. Swarm intelligence (SI) algorithms and evolutionary algorithms (EAs) are the classes of MSAs that considered as popular decision-making tools with intelligent search mechanisms to solve the complex optimization problems. Differential evolution (DE) [7] and particle swarm optimization (PSO) [8] are the most prevalent EA and SI algorithm among all existing MSAs, due to their rival strengths such as simplicity implementation, formidable global search capability, high convergence rate etc. Various of these optimizers have been generated with further improvement of their search manifestation by implementing various approaches such as modified learning strategy and parameter adaptation as introduced in [9–13]. Nonetheless, some research conundrums are still remained as arduous challenges of MSAs. For example, random initialization is adopted as conventional method to generate the initial population of MSAs with questionable quality due to the random guess mechanism [14]. Candidate solutions with poor initial population are endangering the whole optimization manifestation of MSAs. The “No Free Lunch Theorem” [15] is also another prevalent issue of MSAs describing the imperfect of MSAs in tackling complex optimization manifestation. In this study, a novel hybridized algorithm known as self-adaptive hybridized DE with PSO (SADE-PSO) is introduced to overcome the arduous drawbacks. The technical contributions of SADE-PSO are highlighted:

- A self-adaptive mutation operator mechanism is incorporated as an intelligence selection scheme into SADE-PSO to generate promising solutions.
- A novel hybridization strategy is proposed to increase the convergence rate without sacrificing the population diversity by balancing advantages of DE and PSO.
- Detailed analyses are executed to compete the hybrid SADE-PSO with various metaheuristics by applying the CEC 2014 test problems.

The following parts of this article are highlighted. In Sect. 2, the related works of SADE-PSO is summarized. The working mechanisms of SADE-PSO are introduced in Sect. 3. Simulation settings and performance analyses are then presented in Sect. 4. Eventually, the conclusion and further research study of this work are reported in Sect. 5.

2 Literature Review

2.1 Conventional DE

The search mechanism of DE was enlightened by the natural evolution process in [7]. Every individual in DE with N population size is stochastically produced. Each n -th candidate of $X_n = [X_{n,1}, \dots, X_{n,D}]$ is premeditated as a latent solution of a global optimization with decision variable size of D , where $n \in [1, N]$. Mutation is initiated to generate a donor vector $Z_n = [Z_{n,1}, \dots, Z_{n,D}]$ for each target vector $X_n = [X_{n,1}, \dots, X_{n,D}]$ during the evolution process. “DE/rand/1” is the often-used

mutation strategies to compute Z_n , i.e.

$$Z = X_{r1} + F(X_{r2} - X_{r3}) \quad (1)$$

where F is a mutation scale factor changes from $[0,1]$; X_{r1} , X_{r2} and X_{r3} are three stochastically picked solutions and $n \neq r1 \neq r2 \neq r3$.

An offspring vector of $O_n = [O_{n,1}, \dots, O_{n,D}]$ is produced by a crossover process for each n -th candidate according to the relevant vectors of $X_n = [X_{n,1}, \dots, X_{n,D}]$ as well as $Z_n = [Z_{n,1}, \dots, Z_{n,D}]$ as described in Eq. 2, where d acts as the dimensional index and Cr indicates the crossover probability in scope of $[0, 1]$. The $f(X_n)$ and $f(Z_n)$ values are further compared by applying a selection process.

$$O_{n,d} = \begin{cases} Z_{n,d}, & \text{if } rand_d \leq Cr \\ X_{n,d}, & \text{otherwise} \end{cases} \quad (2)$$

2.2 Conventional PSO

Inspired by collective behaviors of fish schooling and bird flocking when searching for food source, PSO was invented in 1995 [8]. Each n -th particle consists of position vector that represented as $X_n = [X_{n,1}, \dots, X_{n,D}]$ and a velocity vector of $V_n = [V_{n,1}, \dots, V_{n,D}]$, where $n \in [1, N]$. Note that each particle has the ability to memorize the so far fittest solution discovered denoted as personal best position of $P_n^{best} = [P_{n,1}^{best}, \dots, P_{n,D}^{best}]$. The global best position $G^{best} = [G_1^{best}, \dots, G_D^{best}]$ is the fittest solution obtained by the whole PSO population. The updated velocity $V_{n,d}^{new}$ and position $X_{n,d}^{new}$ of each n -th particle in each iteration are introduced as follows:

$$V_{n,d}^{new} = \tau V_{n,d} + c_a r_a (P_{n,d}^{best} - X_{n,d}) + c_b r_b (G_d^{best} - X_{n,d}) \quad (3)$$

$$X_{n,d}^{new} = X_{n,d} + V_{n,d} \quad (4)$$

where τ represents inertia weight; c_a and c_b imply the acceleration coefficients; r_a and r_b are two stochastic numbers in scope of $[0, 1]$ that produced from uniform distribution. The objective function value of new X_n , i.e., $f(X_n)$ is competed with $f(P_n^{best})$ for each n -th particle. For minimization problems, the P_n^{best} and G^{best} are replaced if the new X_n is having better value. Moreover, the search procedures of particles are repeated applying Eqs. (3) and (4) until all predefined termination conditions of PSO are met and satisfied.

3 SADE-PSO

A novel hybridization scheme is proposed in SADE-PSO to further evolve the initial population $\mathbf{P}^{Initial} = [X_1, \dots, X_n, \dots, X_N]$ by balancing distinct search features of DE and PSO. DE and PSO are premeditated as primary and secondary optimizers in the hybridization scheme to demonstrate searching of new solutions, respectively. Adoption of various candidate solution generating strategies (CSGSs) was reported to have higher effectiveness during different stages of evolution [16]. Hence, a self-adaptive mutation scheme is designed into the DE phase to enhance the optimization performance of SADE-PSO. Three mutation operators, known as *DE/rand/1*, *DE/best/1* and *DE/current-to-best/1*, are selected to develop the strategy candidate pool, targeting to enhance the population diversity during the mutation operation. Suppose that Z_n refers to the donor vector of candidate solution X_n , the selected mutation operators are described as follows:

$$\text{DE/rand/1: } Z_n = X_{r1} + F(X_{r2} - X_{r3}) \tag{5}$$

$$\text{DE/best/1: } Z_n = X_{best} + F(X_{r1} - X_{r2}) \tag{6}$$

$$\text{DE/current-to-best/1: } Z_n = X_n + F(X_{best} - X_n) + F(X_{r1} - X_{r2}) \tag{7}$$

where F refers to a mutation scale factor; the best solution denoted as X_{best} with the minimum fitness value. At the early stage of optimization process, each mutation operator is assigned with an equal choice probability, where the initial mutation operator choice probability P^{mo} is determined by the reciprocal of mutator operator number I .

During mutation operation, a mutation operator is randomly selected from the strategy candidate pool by roulette wheel selection based on P^{mo} , to generate promising donor vector Z_n of candidate solution n . The framework of the self-adaptive mutation operation scheme is described in Fig. 1. Given the donor vector Z_n , the crossover operation is performed to generate offspring solution O_n based on

Algorithm 1: Self-adaptive mutation operation scheme	
Input:	$X_n, P^{mo}, F, X_{best}, X_{r1}, X_{r2}, X_{r3}$
01:	Produce mutation operation index Ind^{mo} ; /*Roulette wheel selection based on P^{mo} */
02:	switch $Ind^{mo} \in \{1,2,3\}$
03:	case 1 /*DE/rand/1*/
04:	Generate Z_n using Eq. (5);
05:	case 2 /*DE/best/1*/
06:	Generate Z_n using Eq. (6);
07:	case 3 /*DE/current-to-best/1*/
08:	Generate Z_n using Eq. (7);
09:	end switch
Output:	Z_n

Fig. 1 Pseudocode of self-adaptive mutation operation scheme

Eq. (2). The P^{mo} of each mutation operator is updated based on statistical information known as S_{KG} and F_{kg} at the end of each generation of DE. Particularly, S_{KG} and F_{kg} represents the total successful evolutions $nsFlag$ and total failed evolutions $nfFlag$ matrices, respectively. When a new candidate is developed by a mutation operator, the matrix $nsFlag$ is updated if the new candidate is fitter than the current candidate. Otherwise, the matrix $nfFlag$ is updated. Given S_{KG} and F_{kg} , the pre-normalized probability \bar{P}^{mo} and the normalized P^{mo} of each i -th mutation operator is updated as follows, respectively.

$$\bar{P}_i^{mo} = \begin{cases} \sum_{n=1}^N S_{kg} / (\sum_{n=1}^N S_{kg} + \sum_{n=1}^N F_{kg}), & \text{if } \sum_{n=1}^N S_{kg} \neq 0 \\ \sum_{n=1}^N S_{kg} / (\varepsilon + \sum_{n=1}^N F_{kg}), & \text{otherwise} \end{cases} \quad (8)$$

$$P_i^{mo} = \bar{P}_i^{mo} / \sum_{i=1}^I \bar{P}_i^{mo} \quad (9)$$

where ε refers to a small unchanging value used to prevent division by zero; i represents the mutation operator index with number of mutation operator I . In this study, I set as 3. The probabilities of all DE mutation operators are updated for every N solution.

After the offspring solution of all candidate solutions are generated, the greedy selection operation is performed to compare each offspring solution O_n with the candidate solution X_n . The X_n solution is replaced and updated by the O_n solution with better fitness. Otherwise, PSO algorithm is triggered as secondary algorithm to further advance the candidate solution X_n . In contrary to the conventional PSO, the velocity update mechanism premeditates the social component only. When a better solution is found, each candidate solution X_n is only replaced. Thus, it is indispensably equal with the personal best position P_n^{best} . Due to the nullified effect of P_n^{best} and X_n , the self-cognitive component can then be relinquished from Eq. (3). Denote c as the acceleration coefficient, the new velocity component V_n^{new} is formulated as follows.

$$V_n^{new} = \tau V_n + c(G^{best} - X_n) \quad (10)$$

Given the new V_n^{new} produced from Eq. (10), the updated position X_n^{new} of each n -th individual is computed using Eq. (4). After demonstrating the fitness assessment of X_n^{new} , it will be contrasted with X_n and G^{best} . Both X_n and G^{best} will be updated by fitter X_n^{new} . In the contrary, the inferior X_n^{new} will be discarded.

The overall framework of SADE-PSO is highlighted in Fig. 2. Self-adaptive mutation operation scheme is designed to enable various types of mutation operators to be adaptively selected in dealing with distinct types of optimization problems. SADE-PSO is employed with hybrid scheme and anticipated to provide preceding capability in tackling the conflict conditions of high population diversity and fast convergence speed of optimizer by balancing the ascendancies of DE and PSO. The searching procedure of SADE-PSO is duplicated before the counter of fitness assessment δ

```

Algorithm 2: SADE-PSO
Input:  $D, N, X_{min}, X_{max}, Cr, F, \tau, c, \delta, \varphi, I$ 
01: Initialize  $P^{mo} \leftarrow 1/I$ ;
02: Initialize  $\delta \leftarrow 0$  and generate random initial population  $P = \{X_1, \dots, X_n, \dots, X_N\}$ ;
03:  $\delta \leftarrow \delta + N$ ;
04: Sort  $P$  from best to worst based on fitness value;
05:  $G^{best} \leftarrow X_1, f(G^{best}) \leftarrow f(X_1)$ ; //Initialize global best position and fitness
06: while  $\delta \leq \varphi$  do
07:   for each  $n$ -th candidate solution do //Perform searching using DE
08:     Randomly select  $X_{r1}, X_{r2}$  and  $X_{r3}$  from population;
09:     Generate  $Z_n$  using self-adaptive mutation operation scheme (Algorithm 1);
10:     Generate offspring vector  $O_n$  by crossover using Eq. (2);
11:     Ensure  $O_n$  is within search boundaries and then evaluate its fitness;
12:      $\delta \leftarrow \delta + 1$ ;
13:     Update  $nsFlag$  and  $nfFlag$  based on  $O_n$  and  $X_n$ 
14:   end for
15:   Generate new  $P^{mo}$  using Eqs. (8) and (9).
16:   for each  $n$ -th candidate solution do
17:     if  $f(O_n) < f(X_n)$  then /*Perform greedy selection*/
18:        $X_n \leftarrow O_n, f(X_n) \leftarrow f(O_n)$ ;
19:       if  $f(X_n) < f(G^{best})$  then
20:          $G^{best} \leftarrow P_n^{best}, f(G^{best}) \leftarrow f(P_n^{best})$ ;
21:       end if
22:     else /*Perform PSO*/
23:       Update velocity  $V_n$  using Eq. (10);
24:       Calculate new position  $X_n^{new}$  using Eq. (4);
25:        $\delta \leftarrow \delta + 1$ ;
26:       if  $f(X_n^{new}) < f(X_n)$  then
27:          $X_n \leftarrow X_n^{new}, f(X_n) \leftarrow f(X_n^{new})$ ;
28:         if  $f(X_n^{new}) < f(G^{best})$  then
29:            $G^{best} \leftarrow X_n^{new}, f(G^{best}) \leftarrow f(X_n^{new})$ ;
30:         end if
31:       end if
32:     end if
33:   end for
34: end while
Output:  $G^{best}, f(G^{best})$ 

```

Fig. 2 Pseudocode of SADE-PSO framework

surpasses the predefined maximum fitness assessment number φ . The global best solution G^{best} acquired by SADE-PSO is considered as backtracked at the end of searching process.

4 Comparative Studies

Comparative studies of SADE-PSO are evaluated using 30 CEC 2014 benchmark functions [17]. These benchmark functions contain various features in terms of fitness landscapes and complexity. The more thorough information can be acquired from [17]. Performance metrics known as mean fitness F_{mean} and standard deviation SD are employed in this research to estimate the search preciseness and conformity of optimizer, respectively. Optimizer with lower SD is more desirable because it implies

the optimizer has higher conformity in tackling certain test problem with cognate global best fitness.

The demonstrations of SADE-PSO in dealing with all CEC 2014 test functions are competed with five MSAs abbreviated as: multi-swarm PSO without velocity (MPSOWV) [18], opposition based chaotic differential evolution algorithm (OCDE) [19], hybridized firefly and PSO (HFPSO) [20], hybrid PSO with GWO (HPSOGWO) [21] and a conventional hybrid optimizer known as DE-PSO (DE-PSO) [22]. Population size used by all compared competitors is set as $N = 100$ while maximum fitness evaluations numbers is set as $\varphi = 10,000 \times D$. The simulations are conducted using Matlab 2018a with 30 independent runs on a workstation utilized with Intel® Core i5-7300HQ CPU @ 2.5 GHz to deal with all test problems from CEC 2014 at $D = 30$. From Table 1, F_{mean} and SD values generated by all algorithms are introduced, where the best and second-best F_{mean} results are implied in boldface and underlined fonts, respectively. Note that the performance of each competitor is summarized and indicated as $w/t/l$ and $\#BMF$. $w/t/l$ represents that the proposed SADE-PSO performs better in w functions, performs same in t functions, and performs worse in l functions. The quantity of functions that tackled by all algorithms with lowest F_{mean} is summarized as $\#BMF$.

The hybrid SADE-PSO has proposed one best F_{mean} values in tackling the unimodal functions (F1 to F3). DE-PSO optimizer has the best performance in handling the unimodal functions with two best and one second-best F_{mean} values. On the other hand, HFPSO is considered as the worst algorithm in solving unimodal functions by generating three highest F_{mean} values. The most competitive search accuracy is demonstrated by SADE-PSO in order to deal with the simple multimodal functions (F4 to F16) by performing nine best and four F_{mean} results. DE-PSO also considers as second-best optimizers as manage to produce four best and eight second-best F_{mean} results.

In solving hybrid functions of F17 to F22, the novel SADE-PSO remains competitive and delivers promising performance to handle four functions with best F_{mean} values and one functions with second-best F_{mean} values. The optimizers OCDE is reported to be the second-best optimizer as it manages to tackle two hybrid functions with well F_{mean} values. HPSOGWO is presenting well performance to deal with the unimodal and simple multimodal functions, but it performs a degradation in tackling the hybrid functions with mediocre search accuracy as introduced by larger F_{mean} values. This is followed by HFPSO that performs worst with the poorest F_{mean} values in all six hybrid functions. The novel SADE-PSO is managed to attain the highest search accuracies in dealing with the composite functions (F23 to F30). Meanwhile, HFPSO emerges eightworst F_{mean} values in handling the composite functions.

The simulation results introduced in Table 1 also shows that the HPSOGWO has limitations because this peer algorithm only deliver good performance in few test functions. Both OCDE and DE-PSO are managed to tackle CEC 2014 functions with better performance as these optimizers balance the advantages of search features devoted by different algorithms. In particular, the modified self-adaptive mechanism adopted by SADE-PSO in generating population with better quality is the main

Table 1 Performance evaluations of SADE-PSO and five optimizers in solving CEC 2014 benchmark functions ($D = 30$)

Function	Metrics	MPSOWV	OCDE	HFPSO	HPSOGWO	DE-PSO	SADE-PSO
F1	F_{mean}	1.61E+08	8.57E+07	6.20E+09	3.28E+07	<u>2.48E+07</u>	2.46E+06
	SD	7.38E+07	1.21E+08	2.93E+09	3.26E+07	3.39E+06	1.42E+06
F2	F_{mean}	2.12E+10	1.50E+09	9.40E+10	1.58E+09	2.84E-14	4.17E-01
	SD	2.74E+09	2.12E+09	4.68E+10	2.92E+08	0.00E+00	3.18E-01
F3	F_{mean}	8.20E+04	2.78E+04	2.02E+05	3.81E+04	2.69E-03	<u>1.89E+00</u>
	SD	2.44E+04	3.91E+04	1.15E+05	1.75E+04	3.10E-03	3.74E-01
F4	F_{mean}	1.39E+03	3.68E+02	4.57E+04	6.25E+02	<u>1.26E+02</u>	7.16E+01
	SD	3.56E+02	4.24E+02	4.31E+04	6.57E+02	2.10E+00	3.60E+00
F5	F_{mean}	2.09E+01	2.09E+01	2.11E+01	2.09E+01	<u>2.07E+01</u>	2.01E+01
	SD	3.01E-02	6.03E-03	3.93E-02	3.78E-02	3.49E-02	2.67E-02
F6	F_{mean}	3.38E+01	2.12E+01	4.59E+01	1.47E+01	<u>1.22E+01</u>	1.15E+01
	SD	9.46E-01	5.42E-01	8.06E+00	1.47E+00	3.70E+00	5.07E+00
F7	F_{mean}	1.62E+02	7.74E-01	1.54E+03	1.86E+02	2.15E-09	<u>8.62E-03</u>
	SD	1.44E+01	1.09E+00	2.96E+01	2.59E+02	3.03E-09	1.22E-02
F8	F_{mean}	2.68E+02	1.75E+02	4.52E+02	1.51E+02	8.53E+00	<u>6.91E+01</u>
	SD	2.21E+01	5.42E+00	1.75E+02	4.20E+01	1.73E+00	9.15E+00
F9	F_{mean}	3.29E+02	2.23E+02	7.08E+02	2.74E+02	<u>1.43E+02</u>	1.05E+02
	SD	1.25E+01	2.78E+00	1.37E+02	1.00E+02	1.24E+01	1.20E+01
F10	F_{mean}	6.22E+03	4.90E+03	7.01E+03	4.01E+03	2.12E+02	<u>7.97E+02</u>
	SD	2.45E+02	3.24E+02	2.45E+02	1.72E+03	3.31E+01	3.09E+02
F11	F_{mean}	7.24E+03	<u>3.85E+03</u>	7.73E+03	6.66E+03	5.06E+03	3.83E+03
	SD	4.44E+02	3.25E+03	5.47E+02	1.79E+03	8.98E+01	3.95E+02
F12	F_{mean}	2.65E+00	2.17E+00	3.93E+00	2.36E+00	<u>1.08E+00</u>	1.42E-01
	SD	1.63E-01	5.63E-01	2.68E-01	3.81E-01	3.07E-02	1.36E-02
F13	F_{mean}	3.09E+00	8.50E-01	1.18E+01	3.65E+00	<u>3.95E-01</u>	3.29E-01
	SD	2.33E-01	9.16E-02	1.10E+00	1.59E+00	1.52E-02	8.90E-02
F14	F_{mean}	4.48E+01	1.15E+01	5.28E+02	6.24E-01	<u>2.69E-01</u>	2.03E-01
	SD	7.08E+00	2.46E+00	3.47E+02	4.00E-01	9.26E-03	3.23E-02
F15	F_{mean}	9.90E+04	3.22E+02	5.44E+06	3.91E+01	1.44E+01	<u>1.76E+01</u>
	SD	2.26E+04	4.32E+02	2.41E+06	2.54E+01	7.15E-01	2.78E-01
F16	F_{mean}	1.30E+01	1.31E+01	1.43E+01	1.23E+01	<u>1.12E+01</u>	1.06E+01
	SD	1.38E-01	4.88E-02	1.58E-01	6.77E-01	4.41E-01	6.59E-01
F17	F_{mean}	6.01E+06	<u>1.04E+05</u>	3.59E+09	3.25E+06	1.17E+06	3.61E+03
	SD	1.12E+06	1.42E+05	4.66E+09	2.02E+06	7.47E+05	9.44E+01

(continued)

Table 1 (continued)

Function	Metrics	MPSOWV	OCDE	HFPSO	HPSOGWO	DE-PSO	SADE-PSO
F18	F_{mean}	2.58E+08	8.33E+01	1.67E+10	1.03E+08	1.38E+04	6.94E+01
	SD	8.05E+07	1.43E+01	1.48E+10	1.45E+08	1.70E+04	3.31E+00
F19	F_{mean}	1.41E+02	4.34E+00	3.94E+03	1.79E+01	8.13E+00	<u>5.90E+00</u>
	SD	2.38E+01	8.92E-01	2.54E+03	1.98E+00	2.62E-01	3.36E-01
F20	F_{mean}	2.90E+04	<u>3.56E+02</u>	2.31E+09	4.68E+03	2.01E+03	6.50E+01
	SD	9.09E+03	3.21E+02	3.27E+09	4.69E+03	1.65E+03	3.13E+00
F21	F_{mean}	1.56E+06	<u>1.61E+03</u>	1.45E+08	3.46E+04	5.16E+05	1.49E+03
	SD	5.59E+05	6.90E+02	1.77E+08	9.20E+02	4.20E+05	3.82E+00
F22	F_{mean}	9.76E+02	1.12E+02	7.72E+06	7.01E+02	<u>1.92E+02</u>	2.72E+02
	SD	1.96E+02	1.28E+02	5.94E+06	3.32E+02	1.18E+01	5.68E+00
F23	F_{mean}	4.17E+02	<u>3.27E+02</u>	4.82E+03	4.65E+02	3.15E+02	3.15E+02
	SD	1.26E+01	1.69E+01	3.14E+03	3.71E+01	0.00E+00	3.80E-06
F24	F_{mean}	3.20E+02	2.72E+02	4.92E+02	3.12E+02	<u>2.24E+02</u>	2.00E+02
	SD	1.13E+01	1.39E+01	2.76E+01	8.91E+01	4.95E-01	2.48E-05
F25	F_{mean}	2.25E+02	2.13E+02	5.37E+02	2.11E+02	<u>2.09E+02</u>	2.00E+02
	SD	5.87E+00	9.00E+00	3.57E+02	1.06E+00	1.94E+00	0.00E+00
F26	F_{mean}	1.03E+02	<u>1.01E+02</u>	4.98E+02	1.52E+02	1.00E+02	1.00E+02
	SD	5.44E-01	3.05E-02	3.54E+01	6.74E+01	3.79E-02	1.68E-02
F27	F_{mean}	1.14E+03	8.95E+02	3.43E+03	1.14E+03	<u>4.30E+02</u>	4.02E+02
	SD	6.81E+01	9.09E+01	1.78E+03	5.24E+00	1.02E+01	1.86E+00
F28	F_{mean}	1.36E+03	<u>8.62E+02</u>	1.01E+04	1.03E+03	9.43E+02	7.86E+02
	SD	3.69E+02	9.76E+01	5.77E+02	6.06E+01	4.62E+01	6.06E+00
F29	F_{mean}	2.15E+06	3.98E+06	1.89E+09	<u>7.39E+04</u>	2.71E+06	9.75E+02
	SD	4.02E+06	5.62E+06	4.25E+08	3.31E+04	3.83E+06	2.09E+01
F30	F_{mean}	4.04E+04	<u>2.37E+03</u>	1.47E+08	4.11E+04	2.51E+03	1.01E+03
	SD	8.14E+03	1.85E+03	1.97E+08	7.12E+03	4.45E+02	2.16E+02
	#BMF	0	2	0	0	8	22
	w/t/l	30/0/0	28/0/2	30/0/0	30/0/0	21/2/7	-

element leading to the performance differences between SADE-PSO, OCDE and DE-PSO.

Table 2 Pairwise comparison between SADE-PSO with Five optimizers using Wilcoxon signed rank test

SADE-PSO vs	MPSOWV	OCDE	HFPSO	HPSOGWO	DE-PSO
<i>R+</i>	4.65E+02	4.41E+02	4.65E+02	4.65E+02	3.70E+02
<i>R-</i>	0.00E+00	2.40E+01	0.00E+00	0.00E+00	9.55E+01
<i>p</i> -value	1.86E-09	1.42E-06	1.86E-09	1.86E-09	3.89E-03
<i>h</i> -value	+	+	+	+	+

Table 3 Average ranking and associated *p*-values produced by Friedman test (*D* = 30)

Algorithm	SADE-PSO	MPSOWV	OCDE	HFPSO	HPSOGWO	DE-PSO
Ranking	1.3333	4.6833	3.0333	6	3.8167	2.1333
Chi-Square Statistics	124.547619					
<i>p</i>-value	0.00E+00					

Wilcoxon signed rank test is employed as statistical test performance comparison between the novel SADE-PSO and the competitors [23]. The non-parametric statistical test results known as *R+*, *R-*, *p* and *h* values at *D* = 30 are summarized in the Table 2. The significant performance enhancements of SADE-PSO over its competitors are verified as all statistical test results generate the *h*-values of ‘+’. When handling the benchmark problems at *D* = 30, the average rankings of all compared approaches and their related *p*-values originated from chi-square statistic are acquired in Table 3 using Friedman test [24]. According to the Friedman test results, three post-hoc statistical analyses are executed to discover the specific differences with confine optimizer of SADE-PSO. The significant performance improvement of SADE-PSO over MPSOWV, OCDE, HFPSO and HPSOGWO in tackling all test functions, are verified by all post-hoc procedures at significance level of $\alpha = 0.05$ (Table 4).

Table 4 Adjusted *p*-values generated by Bonferroni-Dunn, Holm and Hochberg procedures

SADE-PSO vs.	z	Unadjusted p	Bonferroni-Dunn p	Holm p	Hochberg p
HFPSO	9.66E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
MPSOWV	6.94E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
HPSOGWO	5.14E+00	0.00E+00	1.00E−06	1.00E−06	1.00E−06
OCDE	3.52E+00	4.33E−04	2.16E−03	8.65E−04	8.65E−04
DE-PSO	1.66E+00	9.77E−02	4.88E−01	9.77E−02	9.77E−02

5 Conclusion

A modified hybrid variant of DE and PSO algorithm known as SADE-PSO is introduced in this study to deal with various single objective optimization problems effectively. One of the contributions, self-adaptive mutation operation mechanism is adopted to produce fitter solutions, aiming to advantage SADE-PSO from the prospect of convergence rate and capability in managing premature convergence. Secondly, SADE-PSO is also incorporated with a hybrid scheme to further attain better balancing of exploitation and exploration behaviors of optimizers, by leveraging the search features of both DE and PSO. Extensive comparative analyses were performed to analyze SADE-PSO. The proposed SADE-PSO algorithm is also reported to outperform all peer competitors. The feasibility of the proposed SADE-PSO in training deep learning models such as artificial neural network (ANN) can be further investigated as the future works of current study.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme under Project Proj-FRGS/1/2019/TK04/UCSI/02/1.

References

1. Yao L, Chen Y-Q, Lim WH (2015) Internet of things for electric vehicle: an improved decentralized charging scheme. In: 2015 IEEE international conference on data science and data intensive systems
2. Yao L, Lim WH (2018) Optimal purchase strategy for demand bidding. *IEEE Trans Power Syst* 33(3):2754–2762
3. Yao L, Lim WH, Tiang S, Tan T, Wong C, Pang J (2018) Demand bidding optimization for an aggregator with a genetic algorithm. *Energies* 11(10):2498
4. Choi ZC, Ang KM, Lim, WH, Tiang SS, Ang CK, Solihin MI, Juhari MRM, Chow CE (2021) Hybridized metaheuristic search algorithm with modified initialization scheme for global optimization. In: *Advances in robotics, automation and data analytics: selected papers from iCITES 2020*, pp 172–182. Springer
5. Ang KM, Lim WH, Isa NAM, Tiang SS, Ang CK, Chow CE, Yeap ZS (2022) Modified particle swarm optimization with unique self-cognitive learning for global optimization problems. Springer, Singapore, pp 263–274
6. Koh WS, Lim WH, Ang KM, Isa NAM, Tiang SS, Ang CK, Solihin MI (2022) Multi-objective particle swarm optimization with alternate learning strategies. Springer, Singapore, pp 15–25

7. Price KV, Storn RM, Lampinen JA (2005) The differential evolution algorithm. In: *Differential evolution: a practical approach to global optimization*. Springer, Heidelberg, pp 37–134
8. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *IEEE international conference on neural networks, IV*. IEEE Press, Perth, pp 1942–1948
9. Bilal Pant M, Zaheer H, Garcia-Hernandez L, Abraham A (2020) Differential Evolution: a review of more than two decades of research. *Eng Appl Artif Intell* 90:103479
10. Bonyadi MR, Michalewicz Z (2017) Particle swarm optimization for single objective continuous space problems: a review. *Evol Comput* 25(1):1–54
11. Lim WH, Isa NAM, Tiang SS, Tan TH, Natarajan E, Wong CH, Tang JR (2018) A self-adaptive topologically connected-based particle swarm optimization. *IEEE Access* 6:65347–65366
12. Lim WH, Isa NAM (2015) Particle swarm optimization with dual-level task allocation. *Eng Appl Artif Intell* 38:88–110
13. Das S, Mullick SS, Suganthan PN (2016) Recent advances in differential evolution – an updated survey. *Swarm Evol Comput* 27:1–30
14. Yang X (2020) Metaheuristic optimization: algorithm analysis and open problems, arXiv.org
15. Wolpert D (1997) No free lunch theorem for optimization. *IEEE Trans Evol Comput* 1:467–482
16. Xue Y, Jiang J, Zhao B, Ma T (2017) A self-adaptive artificial bee colony algorithm based on global best for global optimization. *Soft Comput* 22:2935–2952
17. Liang JJ, Qu BY, Suganthan PN (2013) Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization. Zhengzhou University, Zhengzhou China Computational Intelligence Laboratory
18. Ang KM, Lim WH, Isa NAM, Tiang SS, Wong CH (2020) A constrained multi-swarm particle swarm optimization without velocity for constrained optimization problems. *Expert Syst Appl* 140:112882
19. Thangaraj R, Pant M, Chelliah TR, Abraham A (2012) Opposition based Chaotic Differential Evolution algorithm for solving global optimization problems. In: *2012 fourth world congress on nature and biologically inspired computing (NaBIC)*
20. Aydilek İB (2018) A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Appl Soft Comput* 66:232–249
21. Singh N, Singh SB (2017) Hybrid algorithm of particle swarm optimization and grey wolf optimizer for improving convergence performance. *J Appl Math* 2017:1–15
22. Pant M, Thangaraj R, Abraham A (2011) DE-PSO: a new hybrid meta-heuristic for solving global optimization problems. *New Math Nat Comput* 07(03):363–381
23. García S, Molina D, Lozano M, Herrera F (2008) A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 Special Session on Real Parameter Optimization. *J Heuristics* 15(6):617–644
24. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol Comput* 1(1):3–18

Development of Small Scale Machine for Non-Destructive Testing (NDT) to Assess Structural Integrity of Composite



Mohamad Akmal Ridzuan Bin Radzi, Mohammad Nishat Akhtar, Jazmina Binti Ja'afar, Elmi Abu Bakar, S. S. N. Alhady, and Mohamad Nazir Abdullah

Abstract There are many types of the non-destructive test (NDT) method and in this regard ultrasonic testing is the most common method used in composite inspection. The main purpose of the inspection is to detect the defect that occurs in the composite such as cracks, delamination, void, and others. The common Ultrasonic NDT equipment is quite large, expensive and the system is also complex. This research is conducted to develop a small scale of NDT machine with a simple system having an ability to operate immersion ultrasonic testing of the composite and obtain enhanced results compared to conventional product design. The water is used as a coupling medium for the sound beam to travel between the transducer and specimen. The new design of water tank is also developed to make this product become more ergonomic. This small-scale machine has an ability to operate in two axis which is x and y axis and there was a stepper motor used on the axis which was controlled using Arduino Genuino Mega 2560 and Adafruit Motor Shield v2 as a microcontroller and a motor driver respectively. In the inspection process, a transducer of Harisonic Ultrasonic from Olympus corporation with a frequency of 2.25 MHz was used with a pulser-receiver. The result of the specimen such as the location and size of a crack were obtained from the generated signal. The sample of the specimen tested was fiber glass composite laminates (FGCL) with holes as an artificial defect. The suitable stepping mode of the stepper motor with the least vibration produced during operation was analyzed. The structural analysis of the proposed design was done using simulation in Solidwork software. Motion study was used to analyze the basic hand, arm, and body movements of users as they used the machine. It was found that the newly developed NDT machine produced less vibration due to the enhanced mechanical design and stepping motor configuration, and it was found to be more ergonomic compared to previous scanning unit product design.

M. A. R. B. Radzi · M. N. Akhtar · J. B. Ja'afar · E. A. Bakar (✉)
School of Aerospace Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia
e-mail: meelmi@usm.my

S. S. N. Alhady · M. N. Abdullah
School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia

Keywords Non-destructive testing · Composite inspection · Ultrasonic testing · Vibration · Ergonomic

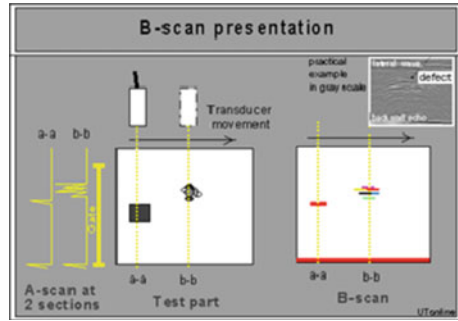
1 Introduction and Background

Non-Destructive Testing (NDT) plays an important role in ensuring the safety and capability of equipment and assets used by people in their daily lives. NDT has been used for the inspection of equipment operating in high pressure environments. This type of test is widely used because it does not affect the specimen being tested. In the aerospace industry, NDT is used for the inspection of materials, especially composite materials. There are various types of NDT methods and ultrasonic testing widely used in the aerospace industry, especially for the inspection of composite materials. The test results are shown as a signal profile. The signal profiles will have different patterns representing different information on each. Commonly used non-destructive testing equipment is quite large, expensive and the system is also complex. This research is done to develop a small-scale non-destructive testing (NDT) machine with a simple system and can carry out ultrasonic composite immersion testing and obtain the same results as the current products. There are several modifications from the previous Multilayer Scanning Array Unit to improve the performance by reducing vibration of the newly developed machine and become more ergonomic.

The scope of work in this manuscript is divided into 4 parts which are designing, fabrication, system development and analysis. The new design of the machine has been made via using Solidwork with added new features. The fabrication process involved some type of processes such as machining and others. To control the mechanism for the inspection process, the stepper motor was programmed, and the data acquisition system was developed in the system development.

NDT refers to a method of evaluation and inspection of materials or additives to characterize or locate defects and deficiencies in an evaluation with some standard without altering the original properties or damaging the specimen being tested. There are other terms to describe non-destructive testing (NDT) which are Non-destructive assessed (NDE) or Non-destructive inspection (NDI). There are various techniques in non-destructive testing (NDT) such as Visual Testing (VT or VI), Ultrasonic Testing (UT), Thermography, Radiographic Testing (RT), Electromagnetic Testing (ET), Acoustic Emission (AE), and Shearography. All these techniques are divided into two categories, namely relational and non -relational methods. The contact method requires contact between the sensor and the specimen surface to obtain good data. Non-contact was developed to speed up the data collection process. Each type of technique has its respective advantages and disadvantages as per Gholizadeh [1]. The proposed work demonstrates a small-scale NDT machine to operate Immersion Ultrasonic Testing (IUT). This type of test uses high -frequency sound energy to perform inspections and make measurements. Ultrasonic inspection can be used for detection, measurement of dimensions, materials and characterization. A typical ultrasonic inspection (UT) system consists of several functional units, such as a

Fig. 1 Example of B-scan data presentation



pulsar/receiver, a transducer, and a display device as demonstrated by Zhang and Richardson [2]. There were several techniques demonstrated by them such as pulse-echo, through-transmission, backscattering, acousto-ultrasonics and ultrasonic spectroscopy. Nonetheless, IUT was found to be promising as it was able to maintain a continuous coupling fluid column between the transducer and the preserved specimen when the probe distance of the specimen changed significantly [2, 3].

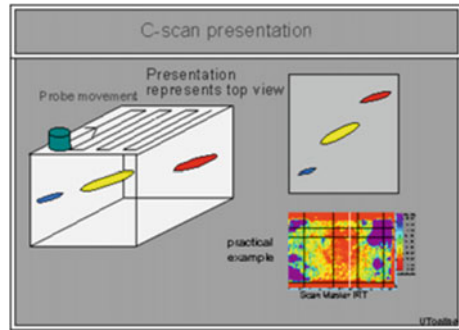
A receiver is an electronic device that produces high voltage electrical pulses. The transducer generates high frequency ultrasonic power driven by a pulser. Some of the energy is reflected back from the defective surface such as cracks when there is a discontinuity in the waveband. The transducer converts the reflected wave signal into an electrical signal and displays it on the screen. The strength of the reflected signal is displayed relative to the time of the signal generated when the echo is received. The results of ultrasonic testing can be shown using several types of data presentation. There are three most common ways to interpret data, namely A-scan, B-scan, and C-scan as evident from the articles by Gandhi et al. [4] and Chen et al. [5]. All types of data presentation have different ways and information. The ability to collect all types of data presentations depends on modern computerized ultrasonic systems. Figure 1 shows the principles of functional ultrasonic testing.

1.1 Working Principle of Ultrasonic Testing

The cross-sectional view of the specimen can be presented in the B-scan (Fig. 1). While in B-scan, the vertical axis represents travel time of the sound energy and the horizontal axis represent the linear position of the transducer. The B-scan can be generated by establishing a trigger gate on the A-scan with a great enough signal intensity.

As per Fig. 2, C-scan is generated with an automated data acquisition system such as a computer-controlled immersion scanning system [2].

Fig. 2 Example of C-scan data presentation



1.2 Role of Motion Study and Vibration Testing

Motion studies are the analysis of the basic movements of the hands, arms, and body of workers as they perform work. Examples of movement were listed as: releasing objects, grasping objects, moving objects, walking and eye movements.

Vibration is defined as the oscillating motion of an object about a reference position. The accelerometer is mounted on the structure to be tested. There are two types of accelerometers, namely high impedance charge output accelerometer and low impedance output accelerometer. The results of the vibration test can be formed in the signal graph, which is the vertical axis of the displacement, and the horizontal axis is the frequency. There are several aspects of the environment that can affect test results such as transverse vibration, acoustic noise, and corrosive materials [6, 7]. Section 2 elaborates on the proposed methodology whereas Sect. 3 provides with the results and discussion followed by Sect. 4 which gives the conclusion.

2 Methodology

There were multiple stages in the methodology of this project. The first stage was designing, as the project started with the research on the problem that has on the current product in term of its mechanism and design. After the problems were identified, the preliminary design of the new product was done using Solidworks and the design was finalized according to the requirement. The second stage started with the fabrication process which involved machining the housing of the stepper motor and linear shaft holder, purchasing the electronic part and assembly. Assembly and finishing process was done after all of the parts were ready and this was the final step in this stage. The third stage was the system development, which was essential to determine the performance of the inspection process. This stage initiated with the programming of the stepper motor controller that was involved in the assembly process. The data acquisition system was also developed in this stage that makes the transducer to collect the data. Then the final stage was related to the analyzation part

i.e., the analyzation of the performance of the product was conducted. The suitable stepping mode of the stepper motor which produced the least vibration during operation was analyzed. The structural analysis was done using Solidwork. In addition, the motion study was used to analyze the basic hand, arm and body movements of users as they used the newly developed machine.

2.1 Designing

The design of this machine is a redesign of the previous scanning unit which is the Multilayer Scanning Array Unit as shown in the Fig. 3. The concept design of a new machine was done using Solidworks. The new design of the developed machine is as shown in Fig. 4. The size of the scanning unit is also the same with the smaller dimensions of $360 \times 280 \times 280$ mm. The new design is redesigned from the previous product by changing some design paradigm, adding new features and using components that are more suitable to suit its operation. The parts that have been redesigned from the scanning unit were linear shaft holder, transducer holder, and stepper motor holder. All of the parts were redesigned in order to make it easy for installation and the structure to become more stable. The new features that have been added to the new machine was a water tank. The user of the scanning unit had much trouble in handling the water tank. Therefore, the new design of water tank for this machine was created to make it more ergonomic. There were also some changes in the component used because it was not suitable for the inspection process. In this

Fig. 3 Multilayer Scanning Array Unit

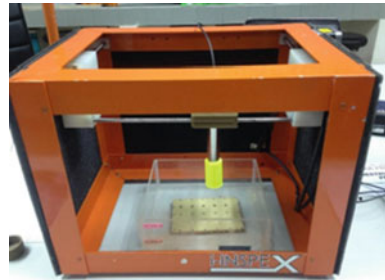
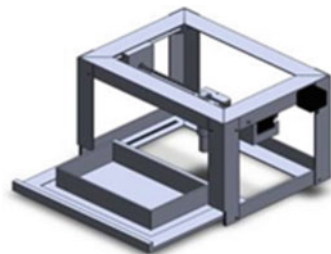


Fig. 4 New design in solidwork



regard, threaded rod was used in the new machine compared to the belting system used in scanning unit because it had many advantages in terms of precision, and the maintenance involved was also less.

2.2 Frame Development

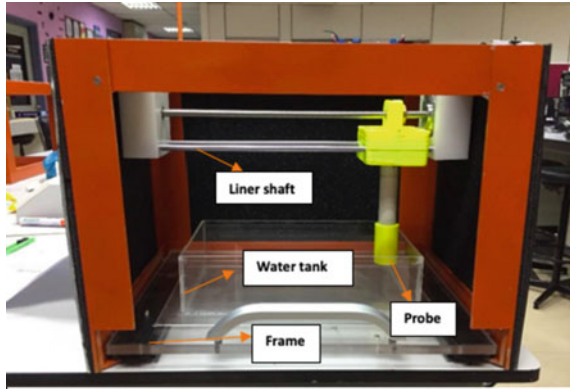
The fabrication started with the frame development. The edges of the frame were joined by using the fastener which formed a strong frame of the machine. The holes for the fasteners at each edge of the frame was made by a drilling process. The linear shaft holder was also fastened at each edge of the body frame.

2.3 Axis and Holder Installation and Finishing

Linear shaft was used as the motion path to move the transducer and helped by the stepper motor. There two types of holder have been made in this project which is stepper motor holder and transducer holder. Stepper motor holder was made from a block of Teflon and machining process was required to get the shape needed by using computer numerical control (CNC) milling machine. The transducer holder was designed and made by using the 3D printer. Linear shaft was installed on each axis which is x and y-axis. The fabrication of x and y-axis have the same function in holding and moving the transducer during the inspection process. On the x-axis, the linear shaft was attached with stepper motor holder and transducer holder which will travel along with it. On the y-axis, the linear shaft was attached to its holder at each edge of the body frame with the stepper motor holder. The new design of water tank was setup after all other parts got assembled. The water tank was made up of Perspex and glued together to get a container shape. The water tank was tested to avoid any leakage during the inspection process.

The finishing process was required in order to increase durability and improve decorative appeal (Fig. 5). The process of finishing was painting and adding perspex as the outer layer of the machine. The color that has been selected for painting is orange to give more attraction to the machine. Before the perspex was added to the machine, the acoustic soundproof foam was glued on it to reduce the noise produced by the environment.

Fig. 5 Small scale machine of immersion ultrasonic testing



3 Results and Discussion

3.1 Structural Analysis

Structural analysis was done to measure the capability of the body frame to support all the load that is mounted on it. There are two parameters measured in this analysis which is the displacement and stress. This analysis is done by using simulation in Solidwork software and the type of analysis is considered static. The geometry for the analysis is taken from the process of designing. Then, the variable spacing mesh was defined all over the geometry. For the meshing, the element size used in this analysis was 10 to attain better accuracy and the shape is tetrahedral as shown in the Fig. 6. The load was applied to the center of the linear shaft structure and an encastre boundary condition was applied on the bottom of the structure. Then the deformed result is plotted to find the value of stress and displacement. The maximum stress concentrated at the below hole of stepper motor on y-axis with a magnitude of 4.39 MPa, whereas the minimum stress occurred at the bottom of the structure with a magnitude of 0.03445 Pa as evident from Fig. 6. Stress concentrations are considered

Fig. 6 Von-misses stress result

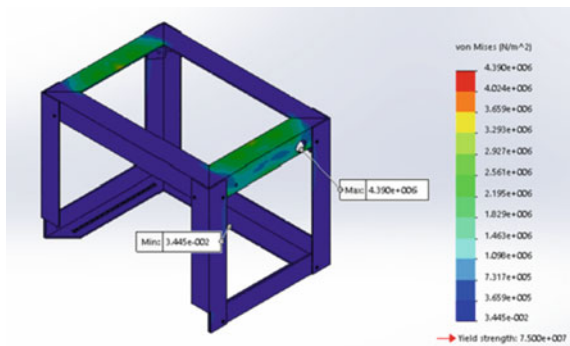
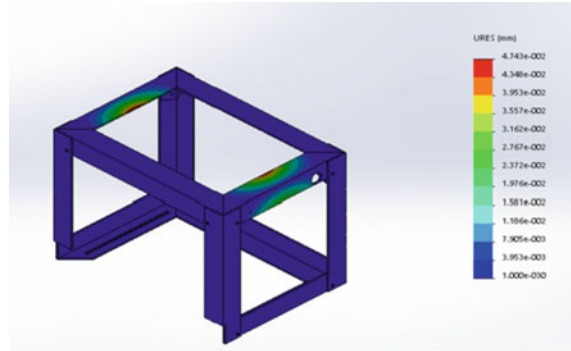


Fig. 7 Displacement result

to arise at the segment where there is a sudden change in the cross-sectional area. The stress concentration will be high if there are severe changes the cross-sectional segment. The value of maximum displacement was 0.04743 mm and the value of minimum displacement was 0.001 mm as evident from Fig. 7. The displacement occurred at the top of the structure. However, this value is too small and will not give any impact to the body frame.

3.2 *Vibration Test*

The movement of the discrete motor can be controlled by selecting the stepping mode. There are 4 stepping modes on the stepper motor which are full step (1 phase), full step (2 phase), half step (interleave) and microstep. Each mode produces different values of vibration, current, and torque output and thus will affect the performance of the machine. Tests were performed to analyze the vibrations generated by the stepper motor. The accelerometer was mounted on the frame structure of the machine body. The data obtained by the accelerometer sensor was interpreted by the Scadas Mobile LMS and displayed on the desktop. The waveforms from the test are plotted in FFT for different modes as shown in Figs. 8, 9, 10 and 11. From the waveform of the vibration, the full step (double phase) mode shows the amplitude is always higher along with the sampling rate. This is a synonym to the theory which indicates that it produces high vibration and also produce higher torque. While the microstep mode shows the amplitude is the lowest along with the sampling rate compared to other modes. This represents that microstep mode is the best mode for stepper motor as it produces low vibration. This mode also generates enough torque for the inspection process.

Fig. 8 Full step(double) mode result

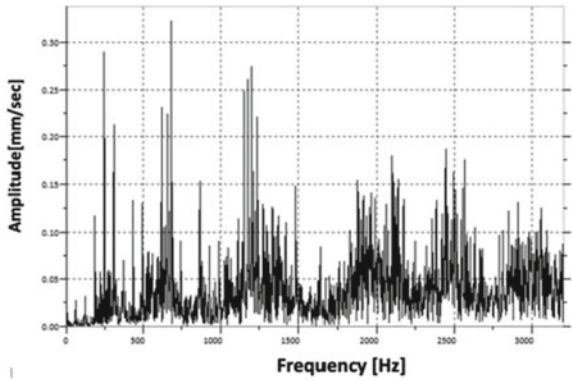


Fig. 9 Full step(single) mode result

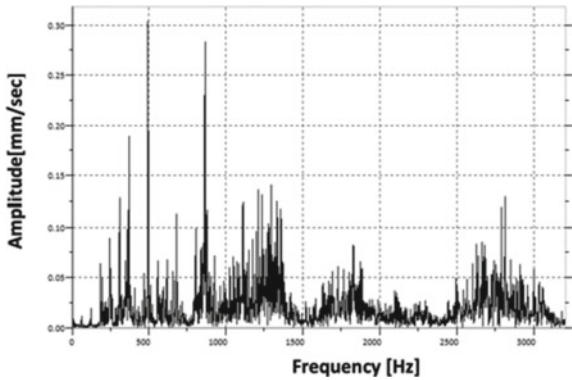


Fig. 10 Interleave mode result

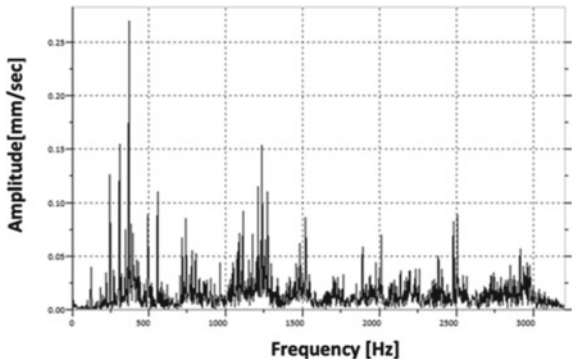
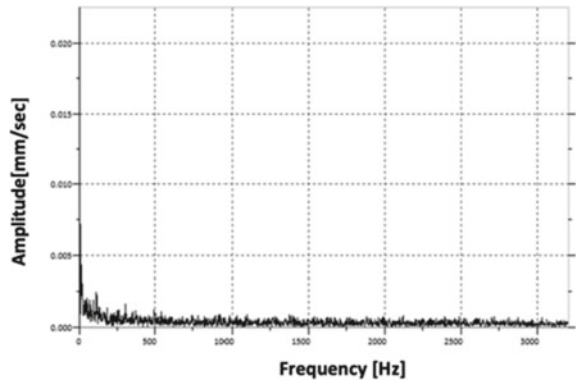


Fig. 11 Microstep mode result



3.3 Motion Study

Motion study is the analysis of the basic hand, arm and body movements of workers as they used the machine. Example of motion is reaching for an object, grasping the object, moving the object, walking and eye movement. This analysis was conducted to show there is some improvement in term of ergonomics on the newly developed machine. A comparison test between the previous and the newly designed machine was conducted whereby the test was conducted 15 times. The total time is taken for the user to operate this machine was recorded. The list of symbols and its associated description were as; *TE*-Reach of an object; *G*-Grasp an object; *TL*-Move an object with hand and arm; *RL*-Release control of an object. Table 1 shows the result of motion study in handling the newly designed products compared with the previous product. Based on the result, the previous product consumed a longer time than the new machine which is 17 s compared to new product design which bears 6 s respectively. The cycle time reduced by 64.70% after making a new improvement to the design. In addition, non-value added activities such as lift up and push down motions were removed. Moreover, the speed of the movement also increased to make

Table 1 Handling the new developed machine

Product type	Activity	Right hand	Left hand	Time (s)
Previous product	Lift the body frame	TE, G, TL	TE, G, TL	7.00
	Put the body frame over the watertank	RL	RL	2.00
	Lift the body frame out of the water tank	TE, G, TL	TE, G, TL	6.00
	Put the body frame on the table	RL	RL	2.00
New product	Pull out the platform of the water tank	TE, G, TL	–	3.00
	Push the platform of the water tank	TE, G, TL, RL	–	3.00

the work faster as the new machine no longer required pull and push from one place to other. Hence, the new machine only uses one hand compared to the previous product.

In this experiment, the results of the inspection process using a new machine were compared with the results obtained using the previous product. A sample composite of 80×120 mm with 15 holes were considered as artificial defects. There were 4 different types of surface on the sample which had no hole, a through hole and two different depth of the hole. The two different depth of holes were about 2 and 4 mm respectively. An experiment was conducted to test the functionality of the mechanism of machine and inspection results on a sample composite. In this experiment, the result from the inspection process by using the new machine was compared with the result obtained using the previously designed product. The result of the inspection was in the form of signal which got displayed on the computer screen. In the interface of the signal, a gate was attached to detect any amplitude that exceeds the limit of requirement needed. The system is well eligible in detecting a hole if there any discontinuity between the peak of the front wall and back wall of the specimen. Figures 12, 13, 14 and 15 shows the results of the composite examination of the sample without holes (Fig. 12), 2 mm depth (Fig. 13), 4 mm depth (Fig. 14) and

Fig. 12 No hole result

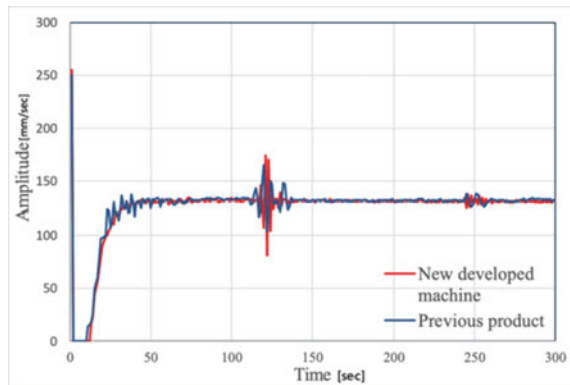


Fig. 13. 2 mm depth result

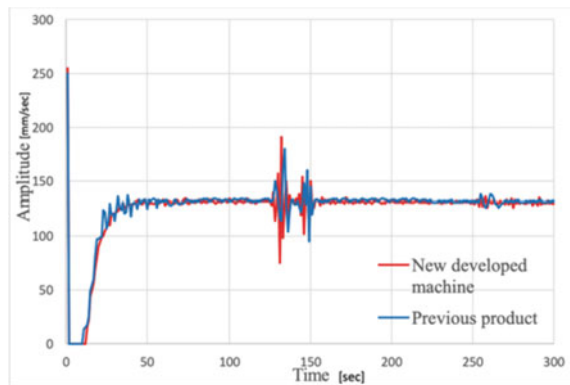


Fig. 14. 4 mm depth result

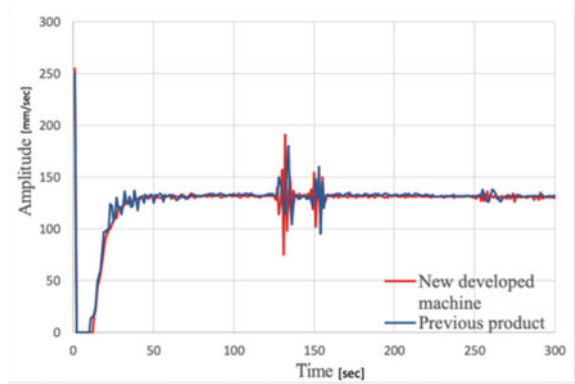
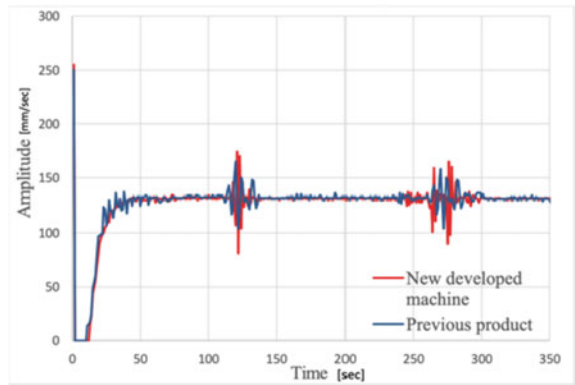


Fig. 15 Through hole result



through holes (Fig. 15). The vertical and horizontal graph axes show the amplitude and time travel of the respective ultrasonic waves. The time travel of the ultrasonic wave also indicates the depth of the composite panel. The graph also shows that the peak for 2 mm depth is nearer to the peak of front wall compared to the peak for 4 mm depth. This shows that the depth of the 2 mm hole has a shorter distance from the surface of sample composite as the ultrasonic wave travel time is shorter. From the graph of through hole result, there are extra peaks at the end of the signal. This is due to the fact that the transducer detects the surface of the water tank floor. It was also evident from the 4 graphs that there was a formation of signal scattering at the initial stage i.e., between 0–100 s with respect to previous product. This was due to the belting system design paradigm of the previous product.

4 Conclusion

A small-scale machine of Non-Destructive Testing (NDT) for composite inspection has been developed and proven working by being tested with the inspection of sample composite with artificial defects. The performance of the newly developed small-scale machine has been increased by using a suitable design and components. The selection of the equipment such as threaded rod instead of the belting system can make the mechanism of inspection process become more precise. The vibration produced by a stepper motor in the operation also has been reduced by choosing a suitable stepping mode during the vibration analysis. The stepping mode was chosen with low amplitude during the sampling rate on the FFT diagram which was evident from the result of the vibration test. It is worth to be noted that due to the possibility of fatigue failure in composite at microlevel, a dynamic response analysis will be incorporated in the proposed design as future work perspective. The newly developed small-scale machine is found to be more ergonomic with added new features. This is proven in the motion study analysis by comparing the results of the newly developed small-scale machine and the previous product which is a scanning unit.

Acknowledgements The authors would like to acknowledge the RU-Top-Down grant 1001.PAERO.870052 by RCMO, USM.

References

1. Gholizadeh S (2016) A review of non-destructive testing methods of composite materials. *Proc Struct Integr* 1:50–57. <https://doi.org/10.1016/j.prostr.2016.02.008>
2. Zhang Z, Richardson M (2011) Nondestructive testing of composite materials. *Handb Multiph Polym Syst* 1:777–796. <https://doi.org/10.1002/9781119972020.ch20>
3. Wang B, Zhong S, Lee T-L, Fancey KS, Mi J (2020) Non-destructive testing and evaluation of composite materials/structures: a state-of-the-art review. *Adv Mech Eng* 12(4):1–28
4. Gandhi N, Rose R, Croxford A, Ward C (2020) Developing a high-fidelity knowledge base for improvements in the non-destructive testing of advanced composite material products. *Proc Manuf* 51:345–352
5. Chen H, Zhou M, Gan S, Nie X, Xu B, Mo YL (2021) Review of wave method-based non-destructive testing for steel-concrete composite structures: multiscale simulation and multi-physics coupling analysis. *Construc Build Mater* 302:123832
6. Berde NN, Sanap SB, Thorat SG (2021) Study of impact and fatigue on 3D printed composites. *Mater Today Proc* 47:1–3
7. Yilmaz B, Asokkumar A, Jasiūnienė E, Kažys RJ (2020) Air-coupled, contact, and immersion ultrasonic non-destructive testing: comparison for bonding quality evaluation. *Appl Sci* 10(19):6757

Hands and Fingers Tracking for Tactile Graphics Reading Assistive Device



Muhammad Ikmal Hakim Shamsul Bahrin, Hazlina Md Yusof,
and Shahrul Na'im Sidek

Abstract Awareness on the difficulties faced by the blind and visually impaired (BVI) people to read tactile graphics has caught the attention of many researchers to develop assistive devices using machine vision approach. At the moment, most techniques that have been used to detect and track hands and fingers are insufficient to support the active and complex behavior of the tactile graphics reading and exploration. This paper presents a system to track both hands and multiple fingers using MediaPipe Hands solution. A benchmark experiment has shown the accuracy of the system to detect and track for single and both hands with 93.1 and 99.9% respectively at an average speed of 20 FPS. Besides that, a preliminary design of tactile graphics reading assistive device has been introduced in this paper. By using the xy -coordinates that have been extracted from the hand landmarks detection, a specific hand gesture and conditions have been configured to allow reader to interact with the tactile graphics. The works presented have shown a promising result to be further explored for future development of tactile graphics reading assistive device which support two hands and multiple fingers exploration.

Keywords Visual impairment · Blindness · Tactile graphics · Tactile exploration · Hand and finger tracking system · MediaPipe

1 Introduction

In this information-driven era, accessing print media that consist of pictures and diagrams is a challenge for blind and visually impaired (BVI) people. It is a critical issue to be acknowledged because graphical information is very important for

M. I. H. Shamsul Bahrin (✉) · H. Md Yusof (✉) · S. Na'im Sidek
Department of Mechatronics Engineering, Kulliyah of Engineering International Islamic
University Malaysia, Kuala Lumpur, Malaysia
e-mail: muhammad.ikmalhakim@gmail.com

H. Md Yusof
e-mail: myhazlina@iium.edu.my

education and navigation. A conventional method for BVI people to access graphical information is by using tactile graphics (refer Fig. 1). It is a two-dimensional format of print media that is made using a variety of techniques such as crafting, swelling printing and embossed drawing [3]. Instead of describing graphical information in braille format, tactile graphics became supplementary materials to help BVI people to visualize pictures and diagrams in textbooks such as mathematics, science, and geography. The challenges of reading tactile graphics using touch sensory is time consuming and cognitive overload [8]. Without any assistance from teachers, instructors, or a sighted person, BVI people will feel lost and confuse when reading them.

Over the past decade, there is a trend of developing assistive devices in machine vision field to overcome the issues of tactile graphics reading. Since it is a touch-based activity, researchers have focused on detecting and tracking hands and fingers to develop interactive and assistive features of the device. The example techniques that have been implemented are skin color detection & background segmentation [2, 4], markers detection [6], color-based foreground and depth segmentation [7] and hand keypoint detection using deep neural network [5]. Table 1 gives a summary of the related works that have been identified.

Nevertheless, there are several limitations of previous approaches that need to be highlighted. Firstly, the techniques used could not support the active and complex nature of tactile graphics reading and exploration. For example, Bardot et al. [1]

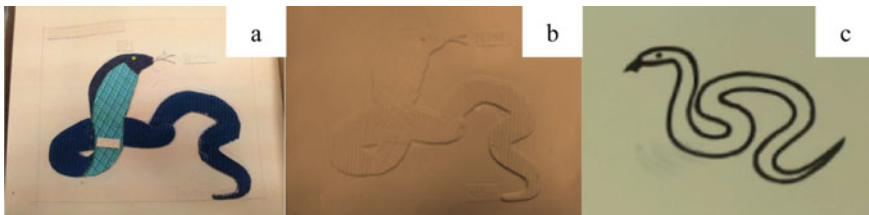


Fig. 1 Example of two-dimensional tactile graphics. **a** and **b** were taken during a visit to KL Braille Resources (27 October 2016). **c** is an example of tactile graphics made using Zytech swell paper

Table 1 Related works

References	Techniques	Outcomes
[4]	1. Background subtraction 2. Color-based skin detection	Single fingertip detection
[2]	1. Background subtraction 2. Distance transformation to binary image	Five fingertips detection
[6]	ArUCO marker detection	Two index fingertips detection
[7]	1. Color-based foreground segmentation 2. Depth segmentation	Five fingertips detection
[5]	Hand keypoint detection using deep neural network	Five fingertips detection

reported that, BVI people tend to perform bimanual exploration which means both are moving simultaneously when reading tactile graphics. Secondly, more improvements need to be addressed to detect and track multiple fingertips during unfavorable and unpredictable conditions such as overlapped or hidden fingers from camera's view, poor lighting condition and positions of hand is half-out from the camera's view. Thus, researchers need to be aware that, there is a variety of behaviour when BVI people read and explore tactile graphics and the techniques that will be used have to be robust enough to overcome the related issues.

The focus of this current study is to propose a method that supports both hands and multiple fingers detection and tracking using MediaPipe platform provided by Google. Currently, the system has been run on a personal laptop with a webcam in real-time. An experimental setup has been developed to provide a benchmark for the system in tracking and detecting both hands and multiple fingers during tactile graphics exploration. Finally, this paper presents an example of hand gesture for the future assistive device that works as an interaction feature for tactile graphics reading and exploration. The proposed method has the potential to be integrated with the assistive device that uses machine vision technique for helping BVI people to read and explore tactile graphics. Not only that, it also can be used for researchers who are interested to investigate the haptic exploratory behaviour of BVI people towards tactile graphics.

2 Methods

2.1 *MediaPipe for Hands and Fingers Tracking*

MediaPipe¹ is a cross-platform framework that applies machine learning pipelines. It is an open-source platform that has been developed by Google which offers solutions for face detection, multi-hand tracking, object detection, etc. This current work has used MediaPipe Hands² solution to detect and track multiple hands and fingers. The MediaPipe Hands solution is capable to infer 21 3D landmarks of a hand after it has been trained with ~30 K real-world images. The hand landmark model can be seen in Fig. 2.

The Python script has been used to build the system and the MediaPipe's hand landmark model has been imported using OpenCV. The setup is shown in Fig. 3. In term of system's configuration, it has been programmed to detect two hands with minimum detection confidence and minimum tracking confidence value to 0.5 (refer to MediaPipe Hands documentation (See footnote 2)). The purpose of having two hands detection is to ensure that both single and bimanual hand exploration can be

¹ <https://opensource.google/projects/mediapipe>.

² <https://google.github.io/mediapipe/solutions/hands>.

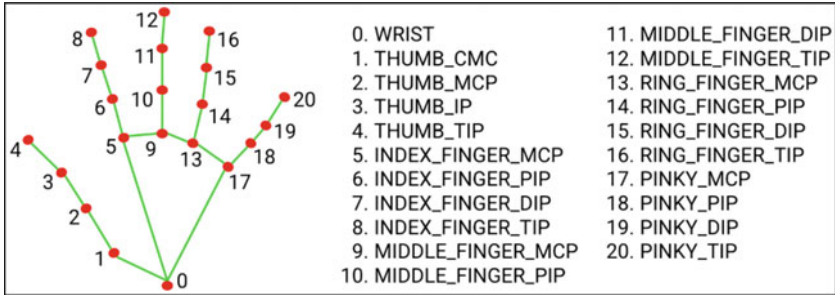


Fig. 2 The 21 hand landmarks model of MediaPipe Hands solution (taken from <https://mediapipe.dev/>)

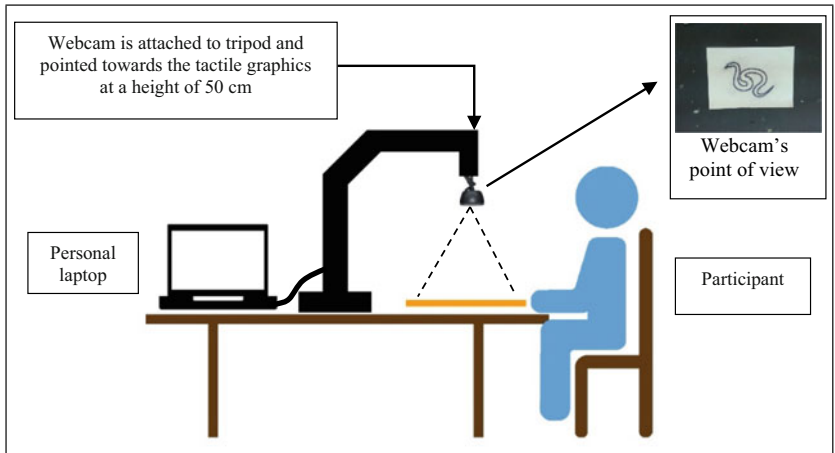


Fig. 3 The equipments setup

detected. Figure 4 shows some examples of hand detection using the MediaPipe Hands solution.

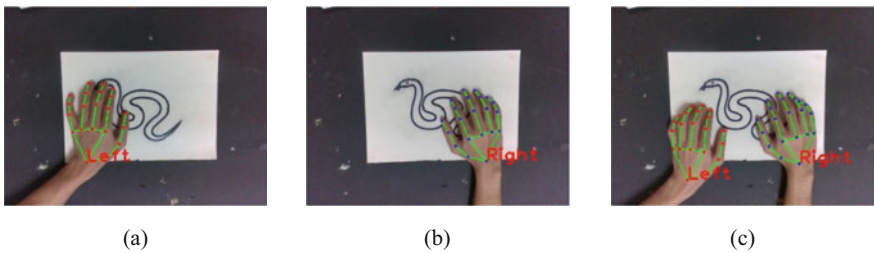


Fig. 4 The example hand detection. **a** shows the left hand detection. **b** shows the right hand detection. **c** shows both hands detection

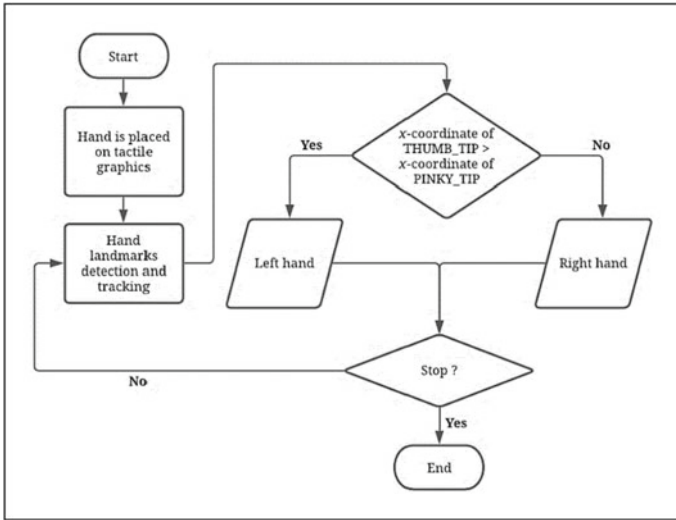


Fig. 5 The system flowchart

Once the hand landmarks coordinates have been attained, the system will differentiate between the left and right hand with respect to the position of thumb fingertip (refer as THUMB_TIP) and pinky fingertip (refer as PINKY_TIP) in x -coordinates. If the value of x -coordinate of THUMB_TIP is greater than the x -coordinates of PINKY_TIP, it will be decided as the left hand, and vice versa. Figure 5 shows the flowchart of the system’s algorithm. Finally, the output data of the system contains a number of hands that are detected and x and y -coordinates of the 21 hand landmarks corresponding to each hand.

2.2 Benchmark Experiment


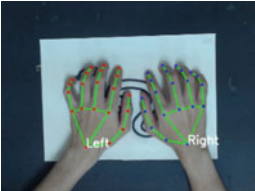
The attention of benchmark experiment is to determine the system’s performance to track hands and fingers during exploration of tactile graphics in real-time. The collected results were, system processing speed (FPS), number of hands and fingers detected and classification of hands. Currently, the system was not tested with any subject due to Covid-19 situation. There were two conditions of the experiment that have been conducted which were single hand and two hands detection and each condition was conducted with five trials. During the experiment, the hands were in tactile reading posture and being moved around on the tactile graphic. It was a situation to imitate the common behaviour of BVI people when reading and exploring tactile graphics.

3 Results

The reported experiment was acquired and processed on an Intel Core i5 laptop computer running MS Windows at 2.4 GHz with 8 GB RAM. The system was able to detect and track hands and fingers with an average speed of 20 fps. To evaluate the performance of the system, 1000 frames have been captured in real-time. Then, each frame has been inspected if the hands and fingers landmarks were correctly detected and tracked for every condition. For every correct detection and tracking, it was scored as 1 otherwise 0. Finally, the percentage of correctly detected and tracked of hands and fingers landmarks has been calculated using formula (1). The results of correct rate are shown in Table 2 which are 93.1% for single hand and 99.9% for two hands detection and tracking. In sum, the correct rate represents the accuracy of the system.

The position error of fingertips landmark between consecutive frames has also been evaluated by placing a single hand in a static position until 1000 frames have been captured. The average position error for each fingertip's landmark was very low approximately 0.0027 to 0.0164 pixels. Throughout the experiment, there has been some misdetection of hands and fingers which are shown in Fig. 6. When a single hand was used, the system sometimes detected two hands as shown in Fig. 6(a). It can also be seen that, when the hand was half out the camera view, the system could not detect the fingers as shown in Fig. 6(b) and Fig. 6(c). However, the percentage of misdetection occurrence is very low for single and both hands detection which are 6.9 and 0.1% respectively.

Table 2 Results of correct hands and fingers detection and tracking

Condition	Correct rate (%)					Average correct rate (%)
	1 st trial	2 nd trial	3 rd trial	4 th trial	5 th trial	
 Single hand	97.6	94.9	92.8	90.0	87.2	93.1
 Two hands	100	100	99.8	99.7	99.8	99.9

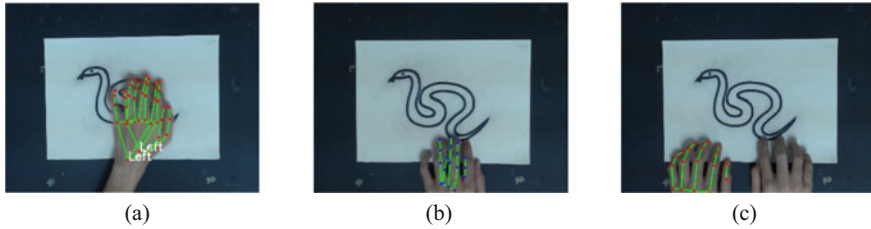


Fig. 6 Example of false detection and tracking. **a** shows that two left hands were detected. **b** shows the imperfection of fingers detection when the hand was half out of the camera view. **c** also shows imperfection to detect both hands when they were half out of the camera view

$$\text{Correct rate (\%)} = \frac{\text{Total count of label '1' frames}}{\text{Total number of frames captured}} \times 100\% \quad (1)$$

4 Preliminary Design of Tactile Graphics Reading Assistive Device

This section presents the preliminary design of an assistive device for tactile graphics reading. Besides MediaPipe Hand solution, the assistive device comprises ArUCO marker detection module.³ In brief, an ArUCO marker has been used to locate the position of tactile graphics material within the webcam's view. It also consists of interactive areas that has been predefined and will appear once the ArUCO marker has been detected. Each interactive area (red transparent box) as shown in Fig. 7(a) can be triggered upon two conditions which can be referred to Table 3. Once both conditions have been met, the interactive area will change into green transparent box as shown in Fig. 7(b) and Fig. 7(c). Audio will be played to convey the information of the area of interest of the tactile graphics to the reader.

5 Discussion

In Sect. 3, the results show that, the accuracy for two hands detection is better than single hand detection. The accuracy for single hand detection was lesser due to detection error as shown in Fig. 6(a) which appeared several times throughout the trials. The detection error might occur because of fast hand movements between consecutive frames. Next, the position of hands and fingers also affected the detection accuracy. If the hands are half out of the camera view, the system cannot detect and track the hands and fingers perfectly as shown in Fig. 6(b) and Fig. 6(c). Based on

³ https://docs.opencv.org/4.5.2/d5/dae/tutorial_aruco_detection.html.

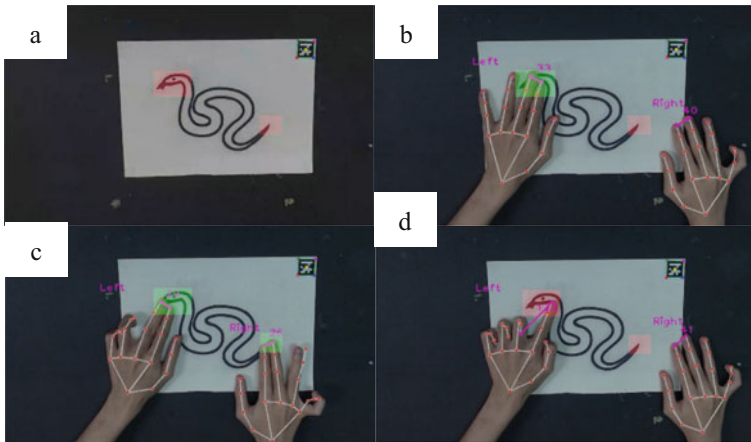
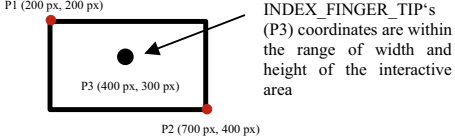


Fig. 7 Preliminary design of tactile graphics reading assistive device. **a** shows that ArUCO marker has been detected and two interactive areas (indicated by red transparent boxes) appeared. **b** and **c** show that left hand and both hands moved toward the interactive area respectively. When both conditions have been met, the red transparent box changed into green color and audio is played. The audio for snake's head is "This is the head of a snake" and the audio for the other area is "This is the tail of a snake". The Fig **d** shows that both conditions are not fulfilled, hence the interactive area will remain the same.

the architecture of MediaPipe Hand solution [9], the localization of hand landmarks model can be performed once the palm has been detected. Thus, the system can only work best when the hand palm is within the camera view. During the experiment, the system was able to maintain the speed of detection and tracking with an average of 20 FPS. It is assumed that the speed might vary if different physical computing platform is being used. Other than that, the position error of each fingertip landmark is very small which shows the effectiveness of the MediaPipe Hand solution in detection and tracking.

In Sect. 4, a preliminary design of tactile graphics assistive device has been demonstrated by using ArUCO detection module and MediaPipe Hand solution. To interact with the specified interactive areas of the tactile graphics, two conditions need to be fulfilled. First, the index fingertip needs to be inside the interactive area. Second, a type of hand gesture which is closely together index and middle fingertips has been used to play the audio of the interactive area. The implementation has been achieved by using the xy -coordinates of the index and middle fingertips from the hand landmark detection. Currently, the assistive device is limited to only one type of hand gesture detection. To include more hand gestures for the interactive features, a further investigation with BVI people and teachers in special education school for the blind need to be conducted to obtain more insight on their reading and exploration behaviour with tactile graphics.

Table 3 Conditions to trigger the interactive area

No	Conditions	Description
1	Index fingertip must be inside the interactive area	<p>To determine if the position of index fingertip is inside the interactive area or not:</p> <ul style="list-style-type: none"> • Firstly, the xy-coordinates of index fingertip has been extracted from the hand landmark (refer as INDEX_FINGER_TIP). • The position of x and y-coordinates must within the range of width and height of the interactive are. The range are defined as follows: <ul style="list-style-type: none"> ○ Width: P1 (x-coordinate) to P2 (x-coordinates) ○ Height: P1 (y-coordinate) to P2 (y-coordinates) • For example: <div style="text-align: center;">  </div>
2	Index and middle fingertip must be kept close together	<p>To determine if the index and middle fingertips are kept close together:</p> <ul style="list-style-type: none"> • Firstly, the xy-coordinates of index fingertip (refer as INDEX_FINGER_TIP) and middle fingertip (refer as MIDDLE_FINGER_TIP) have been extracted from the hand landmark. • The Python <code>math.hypot()</code> function has been used to calculate the Euclidean distance between and middle fingertip in pixel unit. • The Euclidean distance must be less than 50 pixel which indicate that both fingertips are close together.

6 Conclusion

In this paper, a system to detect and track both hands and multiple fingers during tactile graphics reading and exploration has been developed. The system developed using the MediaPipe Hand solution has shown a precise detection of 21 hand landmarks localization in real-time. Left and right hand can also be classified by the

system after the position of thumb and pinky fingertip of each hand have been differentiated based on the proposed method. Based on the benchmark experiment, the system works best when hands are within the camera field of view. If the hand palm is half out of camera field of view, the detection and tracking accuracy will drop. Therefore, it is important to place the camera at suitable height to ensure the hands can be well detected and tracked on tactile graphics materials.

Besides that, a preliminary design of an assistive device has been demonstrated to highlight how the MediaPipe Hand solution can be implemented for developing interactive assistive device for tactile graphic reading and exploration. Such assistive device can benefit the BVI people to gain more access to graphical information in tactile format. However, further investigation needs to be conducted with BVI people validate the practicality of the assistive device. Moreover, this method can also be used to further investigate tactile graphics reading behavior of the BVI people. This will lead towards a design of gestural interaction which can be included for the future assistive device.

Acknowledgements We would like to thank KL Braille Resources and Malaysian Association for the Blind (MAB) for all the supports and guidance given to conduct this research. This work was granted by FRGS/1/2018/TK04/UIAM/02/19.

References

1. Bardot S, Serrano M, Oriola B, Jouffrais C (2017) Identifying how visually impaired people explore raised-line diagrams to improve the design of touch interfaces. In: Proceedings of the 2017 CHI conference on human factors in computing systems, pp 550–555
2. Fusco G, Morash VS (2015) The tactile graphics helper: providing audio clarification for tactile graphics using machine vision. In: Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility, pp 97–106
3. Gupta R, Balakrishnan M, Rao P (2017) Tactile diagrams for the visually impaired. *IEEE Potentials* 36:14–18
4. Kane SK, Frey B, Wobbrock JO (2013) Access lens: a gesture-based screen reader for real-world documents. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 347–350
5. Miwa T, Hosokawa Y, Hashimoto Y, Lisi G (2020) TARS mobile app with deep fingertip detector for the visually impaired. In: International conference on intelligent human systems integration, pp 301–306. Springer, Cham
6. Pandey M, Subramonyam H, Sasia B, Oney S, O'Modhrain S (2020) Explore, create, annotate: designing digital drawing tools with visually impaired people. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp 1–12
7. Reichinger A, Carrizosa HG, Wood J, Schröder S, Löw C, Luidolt LR, Schimkowitsch M, Fuhrmann A, Maierhofer S, Purgathofer W (2018) Pictures in your mind: using interactive gesture-controlled reliefs to explore art. *ACM Trans Access Comput (TACCESS)* 11(1):1–39
8. Sebastian S (2005) The intelligibility of tactile graphics as perceived by blind students. *Jurnal Pendidikan* 99–113
9. Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang CL, Grundmann M (2020) Mediapipe hands: on-device real-time hand tracking. arXiv preprint [arXiv:2006.10214](https://arxiv.org/abs/2006.10214)

Assembly Sequence Optimization Using the Bees Algorithm



Shafie Kamaruddin , Nabilah Azmi, and Nor Aiman Sukindar 

Abstract The determination of the assembly sequence is an important decision in assembly planning. Optimum sequence selection is challenging because of several reasons such as optimization criteria and precedence constraints. Furthermore, a product can be assembled in many different alternatives in accordance with different sequences, thereby making the optimization of assembly sequences a multi-modal solution optimization problem. To allow the process planner to decide, unique optimum solutions are required to be developed as much as possible. In this study, the assembly sequence of a product was optimized by applying an algorithm known as the Bees Algorithm. To assess the performance of this algorithm, the results are compared with results found by other algorithms. It is shown that, the Bees Algorithm obtained similar optimum fitness value with other algorithms but with the greatest number of optimal assembly sequences. As a result, the Bees Algorithm outperforms other algorithms in dealing with the multi-modal optimization problem of assembly sequence optimization.

Keywords Bees Algorithm · Assembly sequence planning · Optimization

1 Introduction

In the past few years, the production rate in the industry is increasing due to the increase in demand from customers. Thus, the assembly process of a product needs to be more effective and efficient to reduce cost and manufacturing time. Assembly process planning is a process of deciding the best sequence during product assembly [1]. Moreover, finding multiple solutions for assembly sequence is necessary to obtain the optimal solution. The criterion of optimal solution includes less assembly time, labour cost, energy consumption, and waste material. Nowadays, researchers are working on finding and studying the most effective way to solve complex problems

S. Kamaruddin (✉) · N. Azmi · N. A. Sukindar

Department of Manufacturing and Materials Engineering, Faculty of Engineering, International Islamic University Malaysia, 53100 Kuala Lumpur, Selangor, Malaysia

e-mail: shafie@iium.edu.my

in the industry. Then, the researchers finally came up with different types of optimization methods that able to solve optimization problems [2]. The most common method to solve these assembly process problems is using optimization algorithms. There are many types of optimization algorithms which are Evolutionary Algorithm (EA), Flower Pollination Algorithm (FPA), Genetics Algorithm (GA), Ant Colony Optimization (ACO) Algorithm, Bees Algorithm (BA), Artificial Bee Colony (ABC) algorithm and Firefly Algorithm (FA). The Bees Algorithm is one of the well-known optimization algorithms where it has been used to solve many problems. It is a population-based search, and it mimics the food foraging behaviour of the bee's colony in nature [3]. In this study, a problem known as motor drive assembly problem was selected to find the optimal assembly sequence. For this problem, there are precedence constraints in finding the optimal assembly sequence of the motor drive [4]. Therefore, components that need to be the predecessor and successor for the assembly process must be considered so that feasible assembly sequences can be obtained.

The remaining of this paper is as follows. Next section describes an overview of optimization assembly sequence planning, the bees in nature, the Bees Algorithm and description of motor assembly problem. Then, the methodology section describes approach adopted for this study. Section results and discussion describes the results of comparison of the Bees Algorithm with other algorithms. Finally, the last section concludes and presents scope for future work.

2 Literature Review

2.1 *Optimization of Assembly Sequence Planning*

This section describes different optimization algorithms that have been used for assembly sequence planning. Genetic Algorithm (GA) is one of the earliest methods used to find optimal solution. It was reported that the GA found a valid assembly plan, but it does not ensure optimal plan and the search progress was slow [4]. Other algorithms for assembly sequence planning are Particle Swarm Optimization (PSO) algorithm, Firefly Algorithm (FA) and Artificial Bee Colony (ABC) algorithm. The ABC algorithm is an algorithm that is almost similar with the Bees Algorithm because both are inspired by food foraging behavior of bees in nature. However, the Bees Algorithm and the ABC algorithm are different in term of work mechanism. Previously, several ABC algorithm based approaches had been reported in solving assembly sequence planning [5, 6]. The finding of these studies show it is efficient and effective for this problem. To date, there are several studies using ABC algorithm but there is still lack of study using the Bees Algorithm. Thus, these studies highlight the need to implement the Bees Algorithm for assembly sequence planning.

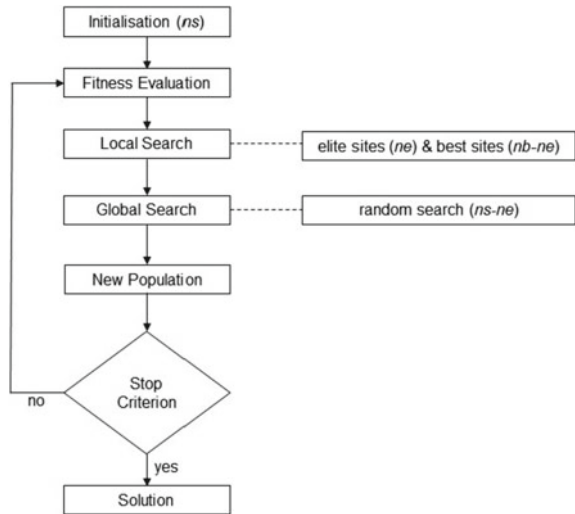
2.2 *Bees in Nature*

During the harvesting season, a part of the honeybees in the hive called scout bees will be sent to find promising flower patches around the field near the hive. The flower patches are their source of food and contain nectar. The scout bees will visit the nectar that is high in volume, easy to get, and abundant in sugar content and the bees will move from one patch to another while gathering information [3]. After returning to the hive, the scout bees who found high-quality flower patches will go to the ‘dance floor’ in front of the hive and do the ‘waggle dance’ [7]. This dance is a way the bees communicate with each other, and all the information gathered will be delivered here. There are three important of information delivered, which are the direction to the flower patches, the distance between the flower patches to the hive, and the quality rating of the nectar [3]. This dance enables the rest of the bees to evaluate the quality of the nectar and the amount of energy needed to gain it. After the ‘waggle dance’ is done, the scout bees will go back to the flower patches with the follower bees (the other bees in the hive). The greater number of bees being sent to the flower patches, the more efficient the process. Hence, the bees will be able to collect the food quickly. Next, the bees will monitor their food level, if the patch is good to be the food resource, they will recruit more bees to go to the source [3].

2.3 *The Bees Algorithm*

The Bees Algorithm is a nature-inspired algorithm inspired by the natural behaviour of honeybees to find the best solution in obtaining food sources. The flow chart of the Bees Algorithm is shown in Fig. 1 [8]. The algorithm begins with dispatching several scout bees (*ns*) randomly in the search space. The process then continues with the fitness evaluation, where the sites visited by the scout bees (*ns*) will be evaluated and sorted based on fitness values. The best sites based on fitness value are named as the selected sites (*m*). The *m* sites are classified into two subsites which are the best sites (*nb-ne*) and elite sites (*ne*). Then, at the local search stage of the algorithm, more bees are recruited to search in the best sites (*nb-ne*) and the elite sites (*ne*). The parameters for recruited bees for the best sites and elite sites are *nrb* and *nre* respectively. The number of recruited bees for elite sites (*nre*) are larger than the number of recruited bees for best sites (*nre*). After that, the process proceeds with the global search which is a random search process at the non-best sites (*ns-ne*). Finally, the best recruited bee from each site form new population with the global search and sorted according to fitness value. While stopping criteria is not been met, the process will be repeated until the global optimum solution is found [9]. The stopping criterion is either the completion of a predefined number of evolution cycles or the location of a solution of fitness above a predefined threshold. In addition, the probability of getting the optimum solution in the Bees Algorithm will be high when more bees are recruited for searching in the elite sites (*ne*) [10].

Fig. 1 Flow chart of the Bees Algorithm



Based on the results, the Bees Algorithm yielded more optimal assembly sequences for the motor drive assembly problem. The fitness value obtained for both 500 iterations and 2000 iterations are 0.2. This fitness value is similar as found by Mishra and Deb (2019), where the fitness value is also 0.2.

2.4 Motor Drive Assembly

This section describes the details of the motor drive assembly problem [4]. The motor drive consists of 12 components (Fig. 2). Table 1 shows the list of grippers and tools used in the assembly adapted from Mishra and Deb (2019). Part No. 1 has been chosen as the base of the components. Table 2 shows the precedence relationships between the parts in the motor drive assembly [4]. The precedence matrix shows the detailed set of predecessor and successor parts. The first column of the table shows that Part No. 1 is the predecessor of all components. The assembly directions in the Table 2 and the tool name in the Table 3 were used to calculate the Fitness Function (fitness value).

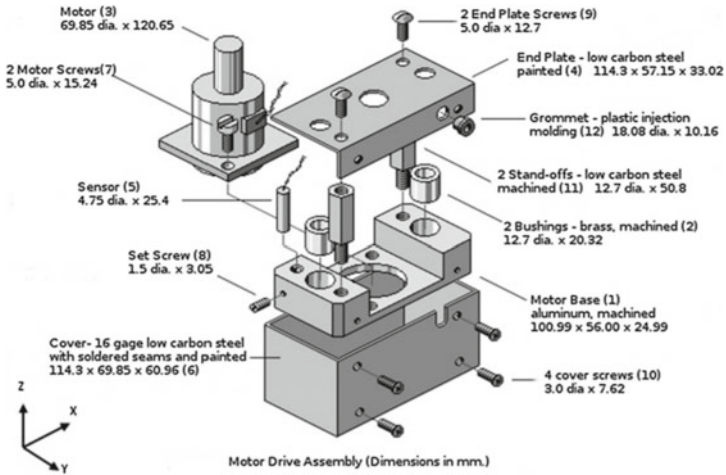


Fig. 2 Motor drive assembly [4]

Table 1 List of tool and grippers used in motor drive assembly [4]

Part no.	Part name	Tool/gripper name
1	Motor base	Two-finger parallel gripper
2	Bushing_1	Two-finger parallel gripper
3	Motor	Two-finger parallel gripper
4	Motor	Two-finger adaptive gripper
5	End plate	Two-finger adaptive gripper
6	Sensor	Two-finger parallel gripper
7	Cover	Slotted screwdriver no. 1
8	Motor screw_1	Allen key
9	End plate screw_1	Slotted screwdriver no. 2
10	Cover screw_1	Philips screwdriver
11	Stand off_1	Open end wrench
12	Grommet	Hammer

Table 2 Precedence matrix for motor drive assembly [4]

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	1	0	1	1	0	0	0	0
5	1	0	0	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	0	1	1	1	0	1	1
7	1	0	1	0	0	0	0	0	0	0	0	0
8	1	0	0	0	1	0	0	0	0	0	0	0
9	1	1	1	1	1	0	1	1	0	0	1	0
10	1	1	1	1	1	1	1	1	1	0	1	1
11	1	0	0	0	0	0	0	0	0	0	0	0
12	1	1	1	1	1	0	1	1	1	0	1	0
Assembly directions	-z	-z	-z	-z	-z	+z	-z	+x	-z	-y	-z	-y

Table 3 Parameters of the Bees Algorithm

Set	<i>ns</i>	<i>ne</i>	<i>m</i>	<i>nre</i>	<i>nrb</i>
1	10	3	5	7	3
2	10	2	5	6	4
3	10	1	5	9	1
4	10	4	5	7	3
5	10	3	6	6	4
6	10	3	6	7	3
7	10	2	6	6	4
8	10	1	7	8	2
9	10	5	6	8	2
10	10	3	7	7	3

3 Methodology

The Bees Algorithm was run using a software known as **R** software. A computer with processor Intel(R) Core (TM) i5-7200U CPU @ 2.50 GHz, 8 GB RAM was used to run the **R** software. Then, the Bees Algorithm was iterated for 100 times for each set of parameters. The parameters used by the Bees Algorithm for this study is shown in Table 3. To find the optimum sequence, the fitness function of the sequences was calculated based on the number of direction changes and the number tool changes in the sequences as described in the next section. A sequence with lower value of fitness function is considered better sequence. The parameters of the Bees Algorithm were

Table 4 Optimal assembly sequence after 100 iterations

No.	1	2	3	4	5	6	7	8	9	10	11	12	Fitness function
1	1	3	2	8	5	7	11	4	9	12	6	10	0.2000
2	1	2	3	5	8	7	11	4	9	12	6	10	0.2000
3	1	3	2	11	7	5	8	4	9	12	6	10	0.2000
4	1	3	2	5	11	7	8	4	9	12	6	10	0.2000
5	1	3	2	5	8	11	7	4	9	12	6	10	0.2000
6	1	2	3	5	8	11	7	4	9	12	6	10	0.2000
7	1	2	3	8	11	5	7	4	9	12	6	10	0.2000
8	1	2	3	11	5	7	8	4	9	12	6	10	0.2000
9	1	2	3	5	7	8	11	4	9	12	6	10	0.2000
10	1	2	3	11	8	5	7	4	9	12	6	10	0.2000
11	1	2	3	11	7	5	8	4	9	12	6	10	0.2000

set based on several general rules. The number of scout bees (*ns*) are same for all set of parameters. The elite sites (*ne*) are the subsets of the best sites (*m*). Thus, elite sites (*ne*) should be less than the best sites (*nb-ne*). The number of recruited bees (*nre*) for the elite sites should be more than the number of recruited bees (*nrb*) for best sites. All parameters were run at 100 iterations and the optimal sequences are recorded in Table 4. The parameters which yielded the highest number of optimal assembly sequence was selected for comparing with other algorithms. Finally, the selected parameters were used to run the Bees Algorithm where the population size is set to 20 and iterated for 500 and 2000 times.

3.1 Calculating Fitness Function

The main objective of this study is to find the optimal assembly sequence of motor assembly problem. This optimal assembly sequence is based on two important factors which are total number of direction changes and total number of tool changes during the assembly process. An assembly process with less tool changes and less direction changes reduces manufacturing cost and time [11–14]. Thus, to evaluate the assembly sequence, a fitness function (*FF*) has been established as in Eq. 1.

$$FF = \frac{1}{(w_x * \text{Total no. of direction changes} + (w_y * \text{Total no. of tool changes} - \text{feasibility index})} \tag{1}$$

The total number of direction changes and total number of tool changes are as stated Eq. 2 and Eq. 4 respectively.

$$\text{Total number of direction changes} = \sum_{i=1}^{n-1} (dir_change_{i,i+1}) \tag{2}$$

$$(dir_change_{i,i+1}) = \begin{cases} 0, & \text{if } assembly_dir_i = \text{if } assembly_dir_{i+1} \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

$$\text{Total number of tool changes} = \sum_{i=1}^{n-1} (tool_change_{i,i+1}) \quad (4)$$

$$(tool_change_{i,i+1}) = \begin{cases} 0, & \text{if } tool_number_i = \text{if } tool_number_{i+1} \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

where, n is the number of components in the assembly, $dir_change_{i,i+1}$ indicates the change in direction of the assembly for two consecutive assembly operations, $assembly_dir_{i+1}$ indicates the assembly direction of component number i , $tool_change_{i,i+1}$ indicates the change in assembly tool for two consecutive assembly operations and $tool_number_i$ indicates the number of tools required for handling/insertion of component number i .

$$\text{feasibility index} = \begin{cases} 0, & \text{if the assembly sequence is feasible} \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

4 Result and Discussion

Based on the result of sequences yielded by each set of parameters, the set of parameter number 10 obtained the greatest number of optimal assembly sequence which are 11 sequences as shown in the Table 4. Then, this set of parameters was used as the parameters to run the Bees Algorithm at 500 iterations and 2000 iterations for comparing the result with other algorithms. Table 5 shows the results of the experiment after 500 iterations and the overall fitness value comparing with other algorithms. The maximum number of optimal assembly sequences obtained are 13 sequences. For the next experiment, the same set of parameters was used with population size of 20 and 2000 iterations. The result of this experiment is recorded in Table 6.

Table 5 Result after 500 iterations comparing with other algorithms

Part no.	GA	ACO	IHS	FPA	BA
Optimum fitness value	0.2000	0.2000	0.2000	0.2000	0.2000
Optimum number of direction changes	9	9	9	9	9
Optimum number of tool changes	5	5	5	5	5
Average fitness value	0.2000	0.1991	0.1891	0.1909	0.2055
Max. number of optimal assembly sequence	4	9	8	9	13

Based on the results, the Bees Algorithm yielded more optimal assembly sequences for the motor drive assembly problem. The fitness value obtained for both 500 iterations and 2000 iterations are 0.2. This fitness value is similar as found by Mishra and Deb (2019), where the fitness value is also 0.2. Thus, this demonstrates the ability of the Bees Algorithm to find optimal solution for this type of problem. Moreover, based on the fitness value, it shows that the optimal assembly sequences provided the least direction changes and tool changes which means the time consumed to assemble the motor drive becomes less. After 500 iterations with the population size of 20, the Bees Algorithm yielded 13 optimal assembly sequences compared to other algorithms as shown in the Table 5. For 2000 iterations, the Bees Algorithm found 18 optimal assembly sequences compared to other algorithms (Table 6). Despite all algorithms yielded similar fitness value but the Bees Algorithm found more optimal assembly sequences.

Figure 3 shows the comparison of optimal assembly sequences of the Bees Algorithm with other algorithms [4]. Based on this comparison, GA and ACO found similar optimal sequences at different iterations. Meanwhile, IHS, FPA and the Bees Algorithm found a greater number of optimal sequences at higher number of iterations. At 2000 iterations with the population size of 20, the Bees Algorithm obtained the largest number of optimal assembly sequence for the motor drive assembly problem. The number of sequences is 18, with fitness value of 0.2. In general, it seems that at higher number of iterations there is a possibility for the Bees Algorithm to yield more optimal assembly sequences. In addition, there is a significant difference in the number of optimal assembly sequences obtained by the Bees Algorithm compared to other algorithms. A possible explanation for this might be the ability of the Bees Algorithm to do local search and global search, thus yielded more possible sequences with the optimal fitness value of 0.2. The global search provided by the Bees Algorithm caused the sequences found at the non-best sites with the fitness value of 0.2222 to rearrange and form the new optimal sequences with the fitness value of 0.2. Moreover, better sequences with lower fitness value could be obtained if there is no restriction in optimizing the assembly sequence. The reason is because the restrictions in solving the problem will restrict the ability of the Bees Algorithm to do local search and global search. Thus, the Bees Algorithm has a bigger potential in solving infinite optimization problem.

Table 6 Result after 2000 iterations comparing with other algorithms

Part no.	GA	ACO	IHS	FPA	BA
Optimum fitness value	0.2000	0.2000	0.2000	0.2000	0.2000
Optimum number of direction changes	9	9	9	9	9
Optimum number of tool changes	5	5	5	5	5
Average fitness value	0.2000	0.1952	0.1894	0.1894	0.2022
Max. number of optimal assembly sequence	4	9	12	15	18

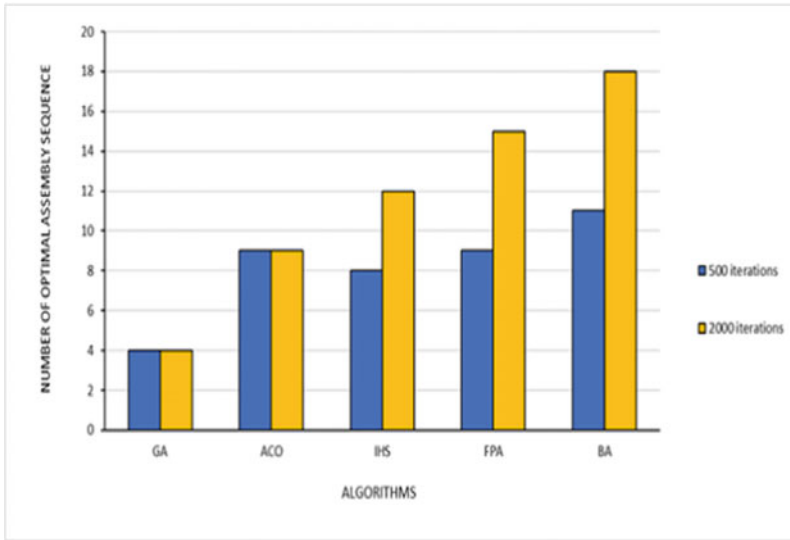


Fig. 3 Comparison of optimal assembly sequences

5 Conclusion

This study has shown that the Bees Algorithm has been successfully applied to this motor drive assembly problem. The optimal fitness value and optimal assembly sequences have been obtained. Then, the result obtained by the Bees Algorithm has been compared with the result obtained by other algorithms such as Genetic Algorithm (GA), Ant Colony Optimization Algorithm (ACO), Improved Harmony Search Algorithm (IHS), and Flower Pollination Algorithm (FPA) which reported by Mishra and Deb (2019). In the comparison, it shows that Bees Algorithm found similar fitness value as found by other algorithms which is 0.2. Another significant finding to emerge from this study is that the Bees Algorithm found higher number of optimal assembly sequences at 500 iterations and 2000 iterations. This new finding should help to improve the decision making in the industry for this type of problem. Using the Bees Algorithm, the assembly sequence of a product can be optimized and give benefits in reducing energy, time, and cost of assembling a product. Finally, further research should be undertaken to explore the performance of the Bees Algorithm on solving more complex problem such as adding larger number of parts and reduce the precedence constraints.

Acknowledgements This study was supported by FRGS grant No. FRGS/1/2019/TK03/UIAM/02/3 from Ministry of Higher Education Malaysia (MOHE). Authors also grateful to the International Islamic University of Malaysia (IIUM) which made this study possible.

References

1. De Fazio TL, Whitney DE (1987) Simplified generation of all mechanical assembly sequences. *IEEE J Robot Autom* 3(6):640–658
2. Dong T, Tong R, Zhang L, Dong J (2007) A knowledge-based approach to assembly sequence planning. *Int J Adv Manuf Technol* 32(11–12):1232–1244
3. Pham DT, Ghanbarzadeh A, Koç E, et al (2006) The Bees Algorithm – a novel tool for complex optimisation problems. In: *Intelligent production machines and systems - 2nd I*PROMS virtual international conference 2006*, pp 454–459
4. Mishra A, Deb S (2019) Assembly sequence optimization using a flower pollination algorithm-based approach. *J Intell Manuf* 30(2):461–482
5. Özmen Ö, Batbat T, Özen T, Sinanoğlu C, Güven A (2018) Optimum assembly sequence planning system using discrete artificial bee colony algorithm. *Math Probl Eng* 2018. Article ID 3407646, 14 pages
6. Yuan W, Chang L, Zhu M, Gu T (2016) Assembly sequence planning based on hybrid artificial bee colony algorithm. In: Shi Z, Vadera S, Li G (eds) *Intelligent information processing VIII, IIP 2016. IFIP advances in information and communication technology*, vol 486. Springer, Cham
7. Seeley TD (1995) *The wisdom of the hive: the social physiology of the honey bee colonies*. Harvard University Press, Cambridge
8. Pham DT, Castellani M (2009) The bees algorithm: modelling foraging behaviour to solve continuous optimization problems. *Proc Inst Mech Eng Part C J Mech Eng Sci* 223(12):2919–2938
9. Yusof NJ, Kamaruddin S (2021) Optimal design of step – cone pulley problem using the bees algorithm. In: Bahari MS, Harun A, Zainal Abidin Z, Hamidon R, Zakaria S (eds) *Intelligent manufacturing and mechatronics. Lecture notes in mechanical engineering*. Springer, Singapore
10. Zhou Z, Xie Y, Pham D, Kamsani S, Castellani M (2016) Bees Algorithm for multimodal function optimisation. *Proc Inst Mech Eng C J Mech Eng Sci* 230(5):867–884
11. Bala Murali G et al (2018) Modified BAT algorithm for optimum assembly sequence planning. *IOP Conf Ser Mater Sci Eng* 377:012091
12. Gao L, Qian WR, Li XY, Wang JF (2010) Application of memetic algorithm in assembly sequence planning. *Int J Adv Manuf Technol* 49:1175–1184
13. Zhou W, Zheng JR, Yan JJ, Wang JF (2011) A novel hybrid algorithm for assembly sequence planning combining bacterial chemotaxis with genetic algorithm. *Int J Adv Manuf Technol* 52:715–724
14. De Fazio T, Whitney D (1987) Simplified generation of all mechanical assembly sequences. *IEEE J Robot Autom* 3(6):640–658

The Application of Conjugate Gradient Method to Motion Control of Robotic Manipulators



Ibrahim M. Sulaiman, Maulana Malik, Wed Giyarti, Mustafa Mamat, Mohd Asrul Hery Ibrahim, and Muhammad Zaini Ahmad

Abstract Many industrial and engineering problems are transformed into optimization problems and solved using various numerical based methods. One of the frequently used method is the Steepest descent algorithm which converge to the solution in only one iteration, given the current point and provided the quadratic function is positive definite. However, this method is not suitable for large scale functions because of lack of gradient information and high computational cost. This study aims to suggest a new conjugate gradient algorithm for motion control of robotic manipulators and unconstrained optimization models. The convergence result of the new algorithm would be discussed under some suitable conditions. Computational simulations are carried out on the discrete-time kinematics equation of a two-joint planar robot manipulator to illustrates the efficiency of the algorithm. The algorithm was further extended to unconstrained optimization problems in addition to motion control of robotic manipulators. Preliminary results prove that the new algorithm is efficient compared to the existing CG algorithm. The comparisons are made using the set of 50 standard benchmark functions including number of iterations and CPU time.

I. M. Sulaiman (✉)

Institute of Strategic Industrial Decision Modelling (ISIDM), School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
e-mail: i.mohammed.sulaiman@uum.edu.my

M. Mamat

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 21300 Kuala Terengganu, Terengganu, Malaysia

M. Malik

Department of Mathematics, Universitas Indonesia, Depok, Indonesia

W. Giyarti

Mathematics Education Study Program, Universitas Islam Negeri Sunan Kalijaga, Depok, Indonesia

M. A. H. Ibrahim

Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Kota Bhar, Malaysia

M. Z. Ahmad

Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kangar, Malaysia

Keywords Unconstrained optimization · Robotic motion control · Line search techniques · Convergence analysis

1 Introduction

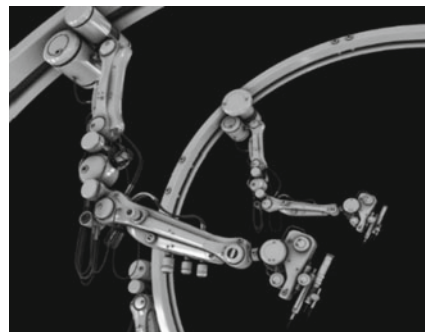
Robotic manipulators (see, Fig. 1) are machines whose grasping device is used in manipulating materials without the operator having any direct physical contact with the machine.

These devices are often found in manufacturing procedures for example; painting and spot welding, and can execute monotonous tasks effectively and at a rate which is faster when compared to human operators [1]. To achieve an efficient result, the velocities and positions of the manipulators' end-effector are digitally controlled [2]. A separate position control system is often used to position each motion of the manipulator. Due to its important industrial applications, numerous literatures consider designing simple and efficient practical controllers that can give optimal controlled performance [3–5]. For instance, [6] considered Jacobian matrix kinematic characteristics and Lagrangian equation to develop an algorithm that would formulate the joint space dynamic model for predicting motion of considered manipulator hand. Other literatures also considered other approaches for predicting the motion of end effector (see [7–9]). However, most of these models are based on Kane, Lagrange, Newton Euler, or other approaches whose dynamics and kinematics of robot arms (multiple-link) are obtained and transform as follows:

$$H(q)\ddot{q} + C(q, \dot{q}) + G(q) + K(q)^T M = T \quad (1)$$

where H represent the inertia matrix, q is robot joint angles vector; C denote the matrix stipulating the Coriolis and centrifugal effects. Also, G represent the vector that specify the effects of gravity, and K is the Jacobean matrix relating the joint torques to M [10]. Equation (1) does not entail the much-needed connections with the suppose drive system and thus, cannot be regarded as sufficient when representing the dominant dynamics of the plant. Based on the matrix-based approach presented

Fig. 1 Robotic manipulator [1]



in the existing literatures, many researchers employ other gradient-type method to investigate the motion control for various robotic manipulators. Recently, several studies considered the CG method for problems of joint planar robotic manipulator. Awwal et al. [11] applied a projection derivative-free iterative methods to solve problem of motion control with regards to two planar robot and [12] considered a derivative-free CG approach for motion control of robot manipulator. Also, [13] and [14] studied the CG technique for unconstrained optimization with application to problem of motion control.

The CG method is an iterative process generating the sequence $\{x_k\} \subset R^n$ via the formula given below:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, 2, \dots \tag{2}$$

with the step size denoted as α_k is computed along the search direction d_k [15]. This direction reduces to the classical steepest descent method if $k = 0$, i.e., $d_k = -g_k$. However, subsequent directions are computation as follows; $k \geq 0$,

$$d_k = -g_k + \beta_k d_{k-1}. \tag{3}$$

Here, $g(x) = \nabla f(x)$ represent the gradient and β_k denote the conjugate gradient parameter that differentiate the CG algorithms. Among the classical CG method includes Polak-Ribière (PRP) [16] and Rivaie et al. [17] with formula as follows

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}}, \beta_k^{RMIL} = \frac{g_k^T y_{k-1}}{d_{k-1}^T d_{k-1}}.$$

where $y_{k-1} = g_k - g_{k-1}$ The method possesses the restart properties and their convergence analysis have been studied by several researchers. For recent studies on this topic (see; [18–27]).

Motivated by the literature discussed above, this paper tends to present a modified CG algorithm for solving problems of motion control for robotic manipulators. The proposed method would further be extended to solve unconstrained optimization benchmark problems. In the subsequent section, the proposed CG algorithm would be defined and its convergence analysis established. Section 3 present the solution for problems of robotic motion control in addition to results based on some benchmark optimization test problems. Finally, the conclusion in given in Sect. 4.

2 New Method and Algorithm

Recently, [28] extended the PRP [16] method to define a new direction of search for the CG algorithm with formula as follows.

$$d_k = -g_k + \beta_k^{PRP} d_{k-1} - \theta_k y_{k-1}, k = 0, 1, 2, \dots \tag{4}$$

where $\theta_k = \frac{g_k^T g_{k-1}}{\|g_{k-1}\|^2}$. The convergence result of this method was established under some suitable Armijo-type line search conditions. Result of numerical experiment demonstrate the robustness of the method. Based on the idea of [17, 28], we define a search direction for the CG algorithm as follows:

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \frac{\|g_k\|^2}{d_{k-1}^T d_{k-1}} \cdot d_{k-1} + \frac{g_k^T g_{k-1}}{d_{k-1}^T d_{k-1}} \cdot d_{k-1}, & k \geq 1 \end{cases} \tag{5}$$

Next, we present algorithm for our proposed search direction below.

Algorithm 1. Algorithm for new search direction

Step 1. Starting with an initial guess $x_0 \in R^n$, set $k = 0$.

Step 2. Compute for α_k using the exact line search with formula as follows:

$$f(x_k + \alpha_k d_k) = \min_{k \geq 0} f(x_k + \alpha d_k). \tag{6}$$

Step 3. Update next iterative point via (2).

Step 4. Update d_k by (5).

Step 5. If $\|g_k\| \leq 10^{-6}$, terminate the process. Else, return to the second step by applying the next k .

2.1 Convergence Analysis

For convergence analysis of the CG method, we consider the assumption defined below.

Assumption A.

- I. The level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded.
- II. The gradient $g(x)$ is Lipchitz continuous in some neighborhood N of Ω , in the sense that, a constant $L > 0$ exist that satisfy;

$$\|g(x) - g(y)\| \leq L\|x - y\| \forall x, y \in N \tag{7}$$

Remark 2.1 Since exact line search was considered in this study, then, from (5), it is obvious that

$$g_k^T d_k = -\|g_k\|^2 \text{ and } \|g_k\| \leq \|d_k\|. \tag{8}$$

This implies the proposed algorithm (5) possess the descent property under the exact minimization rule.

3 Numerical Experiments

This part presents the experimental performance of the new algorithm on problem of motion control for robotic manipulators and unconstrained optimization benchmark models.

To demonstrate the proficiency of the new algorithm, the performance was compared with that of Wei-Yao-Liu (WYL) [29] CG coefficient under exact minimization condition. All algorithms are coded in MATLAB R2018a subroutine program, and run on an Intel Corei7 MacBook pro computer with RAM 4.00 GM.

3.1 Problem of Motion Control for Robotic Manipulator

One indicator of a good method is that it is applicable to real problems. The new algorithm was employed to examine the motion control problem. As describe in [13, 30], the two-joint planar robot manipulators' discrete-time kinematics equation at the position level is expressed as:

$$f(\vartheta_k) = q_k, \tag{9}$$

with $q_k \in \mathbb{R}^2$ representing the end effector position vector, $\vartheta_k \in \mathbb{R}^2$ represent the joint angle vector, while the kinematics function is defined by $f(\cdot)$ and formulated as follows:

$$f(\vartheta) = \begin{bmatrix} l_1 \cos(\vartheta_1) + l_2 \cos(\vartheta_1 + \vartheta_2) \\ l_1 \sin(\vartheta_1) + l_2 \sin(\vartheta_1 + \vartheta_2) \end{bmatrix}, \tag{10}$$

where l_1 is the first rod length and l_2 is the second rod length. The problem of robotic motion control can be defined as nonlinear least square problem:

$$\min_{q_k \in \mathbb{R}^2} \frac{1}{2} \|q_{dk} - q_k\|^2, \tag{11}$$

where q_{dk} control the end-effector when tracking a Lissajous curve, and represented as follows:

$$q_{dk} = \begin{bmatrix} 0.2 \sin\left(\frac{\pi t_k}{5}\right) + 1.5 \\ 0.2 \sin\left(\frac{2\pi t_k}{5} + \frac{\pi}{3}\right) + \frac{\sqrt{3}}{2} \end{bmatrix}. \tag{12}$$

The initial point of the joint angle vector is chosen as $\vartheta_0 = [0, \frac{\pi}{3}]^T$. For the length of rod, we set $l_1 = l_2 = 1$, and the time duration in $[0, 10]$ is divided into 200 parts. MATLAB R2018a software is used to obtain the numerical results of the problem. The experimental results are presented in Figs. 2(a), (b), 3(a) and (b). Figure 2(a)

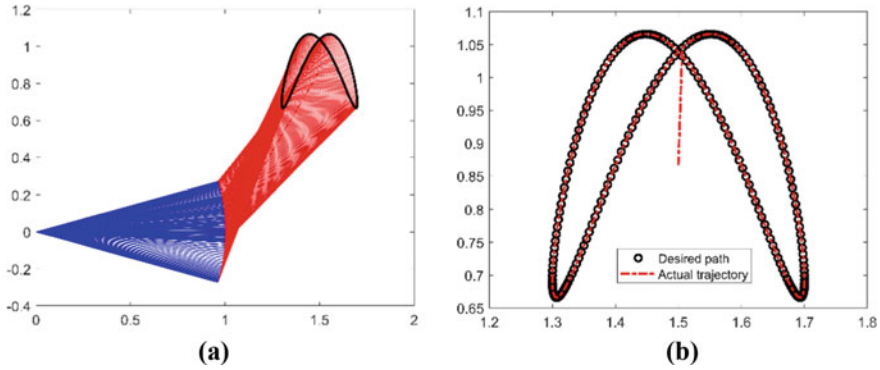


Fig. 2 a Robot trajectories b End effector trajectory and desired path

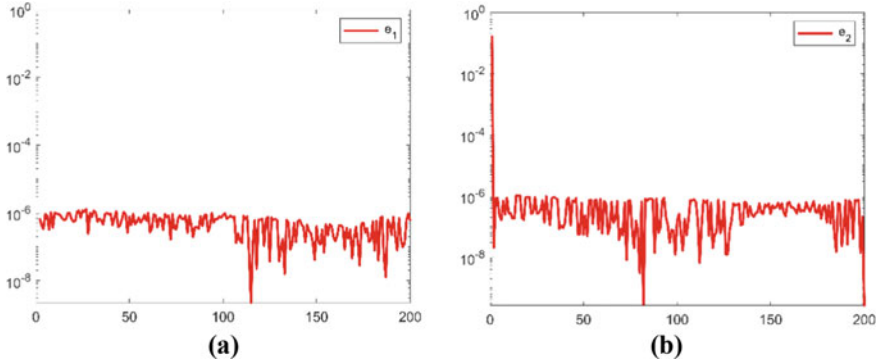


Fig. 3 a Residual error on the horizontal x-axis b Residual error on the vertical y-axis

shows the robot trajectories synthesized by our new algorithm, while Fig. 2(b) depicts end effector trajectory and desired path. Figure 3 demonstrates the trailing error in x and y axes. Based on the Fig. 2, we can see that the new method can be solving the robotic motion control problem and from Fig. 3(a) present the residual error on the horizontal x-axis while Fig. 3(b) depicts the residual error on the vertical y-axis whose trailing error are below 10^{-6} . Hence, we can conclude by saying the proposed technique is effective for solving robotic motion control problem.

3.2 Unconstrained Optimization Problems

The proposed method was applied to solve different unconstrained optimization problems with stopping criteria set as $\|g_k\| < 10^{-6}$. The benchmark problems considered

for this study as presented in Table 1 are from Andrei [31] and Jamil-Yang [32]. The obtained results are demonstrated in the term iteration number and CPU time.

The performance profile of the new method and WYL method are presented in Figs. 4 and 5. Figure 4 present the performance profile with respect to iteration number and Fig. 5 with respect to CPU time. The tool was suggested by Dolan and More [33].

Suppose there exist n_p problems and n_s solvers, then, for every s and p , the study defined $\tau_{p,s}$ as the required time (based on iteration number and CPU time) to compute needed to apply solver s to solve problem p . Based on the algorithms considered in the study, the figures illustrate the part P of the problems such that the algorithm lies in neighbourhood of the factor of τ of the fastest time. In general, the

Table 1 Test functions

No	Functions	No	Functions
1	DENSCHNA	26	Quartic
2	Extended Rosenbrock	27	Matyas
3	ENGVAL8	28	Colville
4	Extended Beale	29	El-Attar-Vidyasagar-Dutta
5	Raydan 1	30	Sphere
6	Extended Tridiagonal 1	31	Sum Squares
7	DENSCHNF	32	Extended White & Holst
8	Extended Himmelblau	33	DENSCHNC
9	FLETCHCR	34	Diagonal 4
10	NONSCOMP	35	Zirilli or Aluffi-Pentini's
11	Extended DENSCHNB	36	Staircase S2
12	Extended Penalty	37	Hager
13	Staircase S3	38	Extended Block-Diagonal BD1
14	Extended Maratos	39	Extended quadratic penalty QP1
15	Generalized Tridiagonal 2	40	Tridiagonal White & Holst
16	Trecanni	41	ENGVAL1
17	Wayburn Seader 1	42	Extended Freudenstein & Roth
18	Shallow	43	Linear Perturbed
18	Generalized Quartic	44	QUARTICM
20	Quadratic QF2	45	Brent
21	Generalized Tridiagonal 1	46	Deckkers-Aarts
22	Booth	47	Dixon and Price
23	Quadratic QF1	48	Rotated Ellipse 2
24	Extended quadratic penalty QP2	49	Zettl
25	HIMMELBH	50	Staircase S1

Fig. 4 Performance profile based on number of iterations

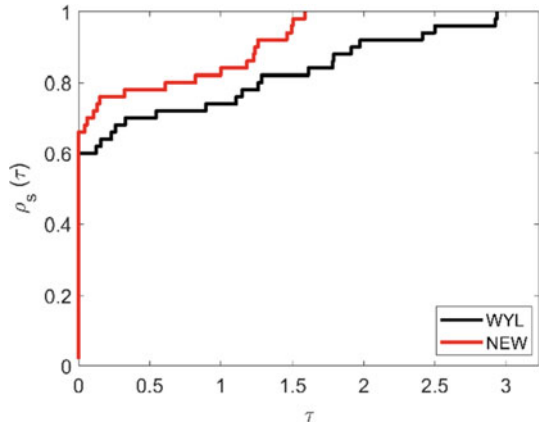
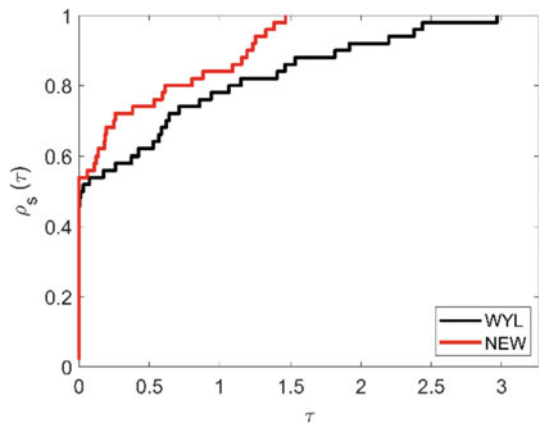


Fig. 5 Performance profile based on CPU time



method at the top is the best method. The top curve shows the algorithm that solves majority of the problems at the best iteration and CPU time.

From Figs. 4 and 5, we can see the profile of the new algorithm at the top from the WYL curve. Hence, we can say the new algorithm perform better than WYL method based on number of iterations and CPU time under the test functions given.

4 Conclusion

In this study, a modified CG parameter is suggested and its performance on motion control of robotic manipulators and benchmark problems is investigated. The convergence analysis of the new algorithm was established under suitable exact minimization condition. The algorithm was employed to solve the problem of robotic

motion control, and further generalized to unconstrained optimization benchmark problems. The result from numerical simulation shows the robot trajectories synthesized, desired path and effector trajectory for the manipulators. Also, the results based on the test functions illustrates that the proposed algorithm is very efficient and competitive compared to the WYL CG method in terms of iteration number and CPU time. In the future studies, researchers can investigate the performance of this new approach using the inexact line search with application to image recovery problems.

References

1. An Introduction to Robotic Manipulators (2014) AZoRobotics.com. <https://www.azorobotics.com/Article.aspx?ArticleID=181>. Accessed 30 July 2021
2. Fadali MS, Visioli A (2020) Chapter 1 - introduction to digital control. In: Fadali MS, Visioli A (eds) *Digital control engineering*, 3rd edn. Academic Press, pp 1–8 (2020). <https://doi.org/10.1016/B978-0-12-814433-6.00001-6>
3. Azar AT, Zhu Q, Khamis A, Zhao D (2017) Control design approaches for parallel robot manipulators: a review. *Int J Model Identif Control* 28(3):199. <https://doi.org/10.1504/IJMIC.2017.086563>
4. Vaidyanathan S, Azar AT (2015) Anti-synchronization of identical chaotic systems using sliding mode control and an application to Vaidyanathan–Madhavan chaotic systems. In: Azar AT, Zhu Q (eds) *Advances and applications in sliding mode control systems*. Springer International Publishing, Cham, pp 527–547. https://doi.org/10.1007/978-3-319-11173-5_19
5. Kathuria T, Kumar V, Rana KPS, Azar AT (2018) Chapter 16 - control of a three-link manipulator using fractional-order PID controller. In: Azar AT, Radwan AG, Vaidyanathan S (eds) *Fractional order systems* Academic Press, pp 477–510. <https://doi.org/10.1016/B978-0-12-816152-4.00016-9>
6. Khatib O (1987) A unified approach for motion and force control of robot manipulators: the operational space formulation. *IEEE J Robot Autom* 3(1):43–53. <https://doi.org/10.1109/JRA.1987.1087068>
7. Slotine J-JE, Li W (1987) On the adaptive control of robot manipulators. <https://journals.sagepub.com/doi/10.1177/027836498700600303>. Accessed 30 July 2021
8. Kim J, Lee AS, Chang K, Schwarz B, Gadsden SA, AlShabi M (2017) Dynamic modeling and motion control of a three-link robotic manipulator. *Proc Int Conf Artif Life Robot* 22:380–383. <https://doi.org/10.5954/ICAROB.2017.GS8-3>
9. Arimoto S, Miyazaki F, Kawamura S (1988) Motion control of robotic manipulator based on motor program learning. *IFAC Proc* 21(16):169–176. [https://doi.org/10.1016/S1474-6670\(17\)54605-4](https://doi.org/10.1016/S1474-6670(17)54605-4)
10. Sweet L, Good M (1985) Redefinition of the robot motion-control problem. *IEEE Control Syst Mag* 5(3):18–25. <https://doi.org/10.1109/MCS.1985.1104955>
11. Awwal AM, Kumam P, Wang L, Huang S, Kumam W (2020) Inertial-based derivative-free method for system of monotone nonlinear equations and application. *IEEE Access* 8:226921–226930. <https://doi.org/10.1109/ACCESS.2020.3045493>
12. Sun M, Liu J, Wang Y (2020) Two improved conjugate gradient methods with application in compressive sensing and motion control. *Math Probl Eng* 2020:e9175496. <https://doi.org/10.1155/2020/9175496>
13. Awwal AM, Sulaiman IM, Malik M, Mamat M, Kumam P, Sithithakerngkiet K (2021) A spectral RMIL+ conjugate gradient method for unconstrained optimization with applications in portfolio selection and motion control. *IEEE Access* 9:75398–75414. <https://doi.org/10.1109/ACCESS.2021.3081570>

14. Abubakar AB, Kumam P, Malik M, Ibrahim AH (2021) A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems. *Math Comput Simul.* <https://doi.org/10.1016/j.matcom.2021.05.038>
15. Sulaiman IM, Mamat M, Waziri MY, Yakubu UA, Malik M (2021) The performance analysis of a new modification of conjugate gradient parameter for unconstrained optimization models. *Math Stat* 9(1):16–23. <https://doi.org/10.13189/ms.2021.090103>
16. Polak E, Ribiere G (1969) Note sur la convergence de méthodes de directions conjuguées. *ESAIM Math Model Numer Anal - Modélisation Mathématique Anal Numér* 3:(R1):35–43 (1969)
17. Rivaie M, Mamat M, June LW, Mohd I (2012) A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Appl Math Comput* 218(22):11323–11332. <https://doi.org/10.1016/j.amc.2012.05.030>
18. Abbas U, Ibrahim S, Mamat M, Ghazali P, Khalid K (2020) The global convergence properties of a descent conjugate gradient method. *J Adv Res Dyn Control Syst* 12:1012–1016. <https://doi.org/10.5373/JARDCS/V12I2/S20201128>
19. Audu Omesa U et al (2020) New hybrid conjugate gradient method for solving fuzzy nonlinear equations. *J Adv Res Dyn Control Syst* 12:585–590. <https://doi.org/10.5373/JARDCS/V12I2/S20201081>
20. Mamat M, Ibrahim S, Malik M, Firman S, Zakaria Z (2020) An efficient spectral conjugate gradient parameter with descent condition for unconstrained optimization. *J Adv Res Dyn Control Syst* 12:2487–2493. <https://doi.org/10.5373/JARDCS/V12I2/S20201296>
21. Liu JK, Feng YM, Zou LM (2018) Some three-term conjugate gradient methods with the inexact line search condition. *Calcolo* 55(2):16. <https://doi.org/10.1007/s10092-018-0258-3>
22. Fang M, Wang M, Sun M, Chen R (2021) A modified hybrid conjugate gradient method for unconstrained optimization. *J Math* 2021:1–9. <https://doi.org/10.1155/2021/5597863>
23. Malik M, Mamat M, Abas S, Ibrahim S, Firman S (2020) A new coefficient of the conjugate gradient method with the sufficient descent condition and global convergence properties. *Eng Lett* 28:1–11
24. Malik M, Mamat M, Abas SS, Sulaiman IM (2020) A new spectral conjugate gradient method with descent condition and global convergence property for unconstrained optimization. *J Math Comput Sci* 10(5) (2020). Art no 5
25. Ibrahim SM, Bakar NA, Mamat M, Hassan BA, Malik M, Ahmed AM (2021) A new hybrid conjugate gradient algorithm for optimization models and its application to regression analysis. *Indones J Electr Eng Comput Sci* 23(2). <https://doi.org/10.11591/ijeecs.v23.i2>. Art. no. 2
26. Dawahdeh M, Ibrahim S, Rivaie M, Mamat M (2020) A new spectral conjugate gradient method with strong Wolfe–Powell line search, vol 8, pp 391–397
27. Kamfa K, Waziri MY, Sulaiman IM, Ibrahim MAH, Mamat M, Abas SS (2020) An efficient hybrid BFGS-CG search direction for solving unconstrained optimization problems. *J Adv Res Dyn Control Syst* 12(02):1035–1041. <https://doi.org/10.5373/JARDCS/V12SP2/SP20201161>. Special Issue
28. Zhang L, Zhou W, Li D-H (2006) A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. *IMA J Numer Anal* 26(4):629–640. <https://doi.org/10.1093/imanum/drl016>
29. Wei Z, Yao S, Liu L (2006) The convergence properties of some new conjugate gradient methods. *Appl Math Comput* 183(2):1341–1350. <https://doi.org/10.1016/j.amc.2006.05.150>
30. Zhang Y, He L, Hu C, Guo J, Li J, Shi Y (2019) General four-step discrete-time zeroing and derivative dynamics applied to time-varying nonlinear optimization. *J Comput Appl Math* 347:314–329. <https://doi.org/10.1016/j.cam.2018.08.017>

31. Andrei N (2020) Nonlinear conjugate gradient methods for unconstrained optimization. Springer International Publishing. <https://doi.org/10.1007/978-3-030-42950-8>
32. Jamil M, Yang X-S (2013) A literature survey of benchmark functions for global optimization problems. *Int J Math Model Numer Optim* 4(2):150. <https://doi.org/10.1504/IJMMNO.2013.055204>
33. Dolan ED, Moré JJ (2002) Benchmarking optimization software with performance profiles. *Math Program* 91(2):201–213. <https://doi.org/10.1007/s101070100263>

Analysis of Multiple Prediction Techniques of Received Signal Strength to Reduce Surveying Effort in Indoor Positioning



Mohd Amiruddin Abd Rahman, Caceja Elyca Anak Bundak,
and Muhammad Khalis Abdul Karim

Abstract Received Signal Strength is the measure of attenuation of electromagnetic signals emitted by the access point, reaching the receiver after traveling some distance. This work used the attenuation of Wireless Local Area Network signals propagated through the air for the purpose of indoor positioning. Previous research had shown some problems such as indoor mapping requires human effort and are time-consuming. Furthermore, received signal strength for different indoor conditions may vary such that constant calibration and new acquisition for unknown indoor locations is required. An approach to reduce manual acquisition is by employing prediction algorithms. In this work, an analysis on prediction techniques used predict the RSS is analyzed in the context on indoor positioning. First, to determine the optimum training size for the models, the models are given different training size. Then the models are evaluated based on the similarity of signal pattern predicted and the error between the predicted signal and real signal. In conclusion, the random function model showed best estimation for signal for most of the tested signal received at certain distances from the transmitter. The optimum training size found for all the prediction models are 1100 out of 1200 data. It is also found that for a very noisy data set, the minimum training size for best result are at 900 out of 1200. Bayesian Support Vector Regression outperforms other models in terms of root mean square error.

Keywords Indoor positioning · Fingerprint · Received Signal Strength · WLAN

M. A. Abd Rahman (✉) · C. E. Anak Bundak · M. K. Abdul Karim
Faculty of Science, Universiti Putra Malaysia (UPM), 43400 UPM Serdang, Serdang, Malaysia
e-mail: mohdamir@upm.edu.my

C. E. Anak Bundak
e-mail: gs52905@student.upm.edu.my

M. K. Abdul Karim
e-mail: mkhalis@upm.edu.my

1 Introduction

Today, the emergence of smart phones application that utilizes mobile applications to access the positioning technology is available. The currently available positioning utilizes satellite technology. Examples of satellite navigations includes Global Positioning System (GPS) [1], Global Navigation Satellite System (GLONASS) and many more. The positioning system is widely used in many mobile applications to provide location services such as Google Maps and Waze. However, the usage of GPS in an indoor environment is inaccurate due to poor signal received due to multipath reflection and signal blockage from buildings [2]. Therefore, other technologies such as magnetic field, Wireless Local Area Network (WLAN), and Bluetooth is used for indoor positioning system.

This paper is focused on the technique of indoor localization using WLAN. The technology of WLAN positioning is based on matching the Received Signal Strength (RSS) that is collected during the training phase and with the one that is seen by the user during the testing phase. This technique is called as fingerprinting. However, the RSS for WLANs indoor is affected by many factors especially in complex indoor environments. The infrastructural disturbance causes significant multipath effect for radio signal propagation. Moreover, random error that are induced by mobile device user such as attenuation of the human body, receiver hardware orientation reduces the indoor positioning accuracy. When RSS is employed in a real indoor environment, the RSS will have severe variation which subjects to the real time environment condition and infrastructure within the indoor environment. Multipath effect due to those infrastructural disturbances and user induced attenuations are significant to the result obtained. Hence a general approach is to consider the errors as noise and mean signal filter is used to reduce the noise. This method has limited the performance of an indoor positioning system as RSS recorded offline might have a high chance of mismatching to the online RSS due to the real time environment when the data is taken. Since WLAN RSS is very different for every indoor environment, mapping and constant calibration of RSS is needed for high quality indoor positioning system.

To map the RSS of an indoor positioning, the RSS of multiple access point at multiple location need to be measured and recorded. This causes low efficiency as it is both time and power consuming. Hence, we propose multiple prediction techniques to automate or to reduce the human power in collecting the raw data of WLAN RSS. This paper makes the following contributions. i) To investigate the pattern of predicted signals by 4 different prediction models at different distance away from the access point, ii) To study the performance of all the prediction models at different train size, iii) To evaluate the performance of the prediction models at different distance away from the access point.

2 Signal Prediction Algorithms

After the RSS dataset is collected during the training phase, it is segmented to be used as validation set and training sets. The following sections explains 4 models used in approach to predict the signal pattern.

2.1 Mean Function

The Mean function is assumed as the representation of the controlled set of prediction, which the function calculates the mean of the training set and project the value as prediction for the testing signals, hence the output pattern plot will be a straight line. For the mean model, by setting the mean function to calculate the mean for training set size Q . The total training set is 1 to Q . The predicted signal is then compared to the signal from $(Q + 1)$ th to the last signal for performance. The mean function could be written as:

$$xMean = \text{mean}(T(1 : Q)); \quad (1)$$

where $xMean$ is the output variable containing the mean for 1 to N where N is representing the training size required and T is the table containing the filtered raw data. The root mean square of the mean function is then calculated by using:

$$RMSEMean = \sqrt{\frac{1}{N} \sum_i^N ((xMean - \text{realSignal}(i))^2)} \quad (2)$$

where i is first number of testing data, and N is the 1200.

2.2 Mode Function

The comparison of mode function compared to the mean function is that the mean function may output a value which is not whole numbers, while mode function generates whole number outcomes which in turn reduces the error. The Mode function that calculate the mode of the training set, outputs a whole numbered straight-line plot. For the mode model, by setting the mode function to calculate the mode for training set size N the training set is 1st signal to the N th signal, the predicted signal is then compared to the signal from $(N + 1)$ th to the last signal for performance. The mode function could be described as:

$$xMode = \text{mode}(\text{realSignal}(1 : N)); \quad (3)$$

where $xMode$ is the output variable containing the mean for 1 to N where N represents the training size required and $realSignal$ is the filtered raw data. The root mean square of the mode function is then can be calculated by using:

$$RMSEMode = \sqrt{\frac{1}{N} \sum_i^N ((xMode - realSignal(i))^2)} \quad (4)$$

where i is first number of testing data, and N is the 1200.

2.3 Random Function

The Random Function is aimed to be able to describe the fluctuation of the signals. The Random Function is designed to replicate real time fluctuation of received signal strength. To generate the Random Function model, the raw data is analyzed using an algorithm which outputs the total recorded occurrence of each signal over time t , then from the occurrence, the probability of occurrence of each signal can be obtained. Algorithm 1 describes the random function algorithm.

Algorithm 1: Random function

Input: filtered data from FilterScript

Output: structure $psignal$ containing information of signals and relative probability.

```

1: initialize  $d=1$ ;
2: for  $signal= -90$  to  $-20$  do{
3: for  $a=1:N$  do{
4: if  $realSignal(a)==signal$  do{
5:  $freq=freq+1$ ; } /*checks how many occurrence*/
6: end for statement
7:  $psig(d,1).signal= signal$ ;
8:  $psig(d,1).freq=freq$ ;
9:  $d=d+1$ ;
10: end for statement
```

The algorithm obtains real signal of 1 to N and groups it according to each single signal value from -90 to -20 dB according to observed occurrence frequency of each signal. The probability of occurrence of each signal is calculated as in Algorithm 2.

Algorithm 2: Probability of signal occurrence

Input: filtered data from FilterScript, mode of the selected data xm.

Output: structure p_{sig} containing information of signals and relative probability.

```

1: initialize d=1;
2: for signal=( xm-2 : xm+2) do {
3: for a=1:N do {
4: if realSignal(a)==signal do {
5: freq=freq+1;}} /*checks how many occurrence*/
6: end for statement
7: psig(d,1).signal= signal;
8: psig(d,1).freq=freq;
9:d=d+1;
10:end for statement

```

Then, a range of ± 2 dBm range is selected for the random function to generate a new set of data using the probability of individual signals from the raw data. The adjusted random function is described in Algorithm 3.

Algorithm 3: Adjusted Random function

Input: structure containing signals with corresponding probability respective to signal.

Output: Generated random signal with M dimension.

```

1 :a=row of signals
2: b=row of corresponding probabilities
3: c=[a;b];
4: gx=randsrc(M,1,c);

```

In Algorithm 3, M is the needed size for the generated random signal used to calculate the root mean square error or simply $M = 1200 - N$, the algorithm used to calculate the root mean square error described as follows.

$$RMSE_{Random} = \sqrt{\frac{1}{N} \sum_i^N ((RandomSignal(i) - realsignal(i))^2)} \quad (5)$$

2.4 Bayesian Support Vector Regression

The Support Vector Regression is used to perform a time series prediction. In this study, the BSVR algorithm is obtained and modified in the training size segment from the algorithm used in [3, 4]. However, rather than using multiple descriptors to correlate with the target, this work only uses one descriptor which is the prior value of RSS as training. Inside BSVR algorithm, the hyperparameters are optimized based on Bayesian function. The iteration is set to 100 runs to find the best hyperparameter

values. Algorithm 4 shows the developed Bayesian optimization function of the algorithm to tune the hyperparameters by minimizing the objective function [5].

Algorithm 4: Bayesian optimization to find the best objective function

Input: Input space D_0 ; GP prior μ_0, σ^2_t

- 1: for $t = 1, 2, 3, \dots$ do;
- 2: Select next point to evaluate x_{t+1} through the optimization of the acquisition function α
 $X_{t+1} = \text{argmax } \alpha(x; D_t)$
 $x \in D_t$
- 3: Evaluate the objective function to obtain y_{n+1}
- 4: Augment the data observed $D_{t+1} = \{D_t, (x_{t+1}, y_{t+1})\}$
- 5: update the Gaussian process model using μ_{t+1}
- 6: end

Result: Optimize the Gaussian process mean to find an optimized solution.

3 Experiment

The experiments were conducted at the instrumentation lab at Faculty of Science, Universiti Putra Malaysia. The raw RSS data is collected using Vistumblr software for 20 min such that the data is sufficiently big enough to be segmented into training and testing data. To take the raw RSS of an access point without the effect of user presence or any unwanted disturbance, the device is left inside the lab while the user is outside the lab and away from the device. First the RSS is taken first at distance 1 m away from the access point, then the device is moved into the pre-located 2 m point away from the access point and the process is repeated for distance of 3, 4, 5, 6 m. Figure 1 shows the experimental setup. The maximum distance that could be set up for the measurement is 6 m due to constrain of the lab space.

Then the raw data is then filtered and segmented into specified training and testing sizes. There are $N_{\text{total}} = 1200$ data collected for 6 datasets which refers to every

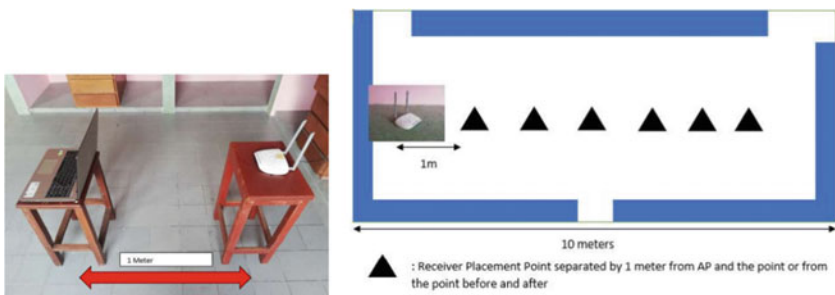
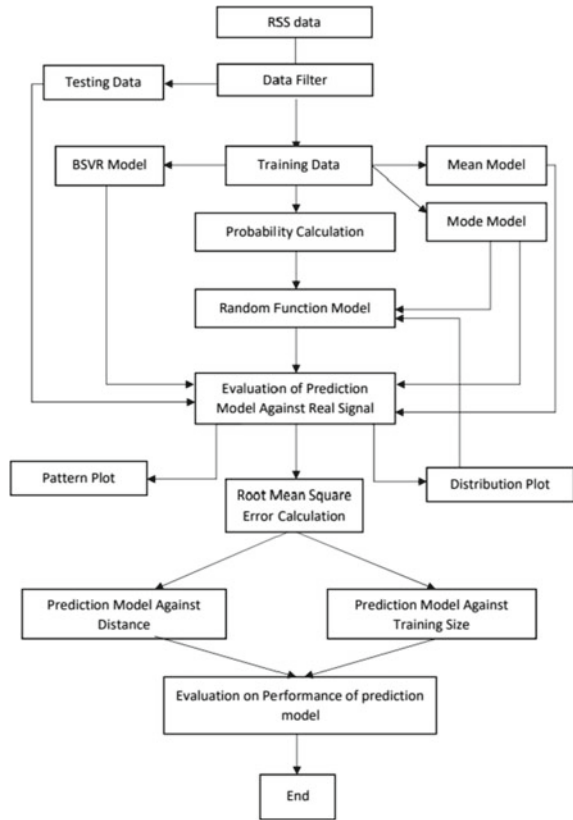


Fig. 1 The experiment setup

Fig. 2 Flowchart representing the analysis of data



tested location distance. This means a total of $1200 \times 6 = 7200$ data were collected. The filtered data is segmented into training, N_{train} and testing, N_{test} data. The ratio of training to testing data is made as $N_{train} : N_{test}$ where $N_{test} = N_{total} - N_{train}$. the training data is set to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 and 1100 with respecting testing data. The training data then is used to train the BSVR algorithm as well as the mode, mean and probability of the data is calculated. Then the probability and the mode are used by Random Function Model to predict the signal and all 4 prediction models are evaluated against the real signal. The computation of the algorithm is made using MATLAB software. An overall flowchart procedure is described as in Fig. 2.

4 Results and Discussion

The patterns plot of the four model is plotted and recorded. The objective is to observe the pattern of the predicted signal and the real signal. Figure 3(a) shows the signal

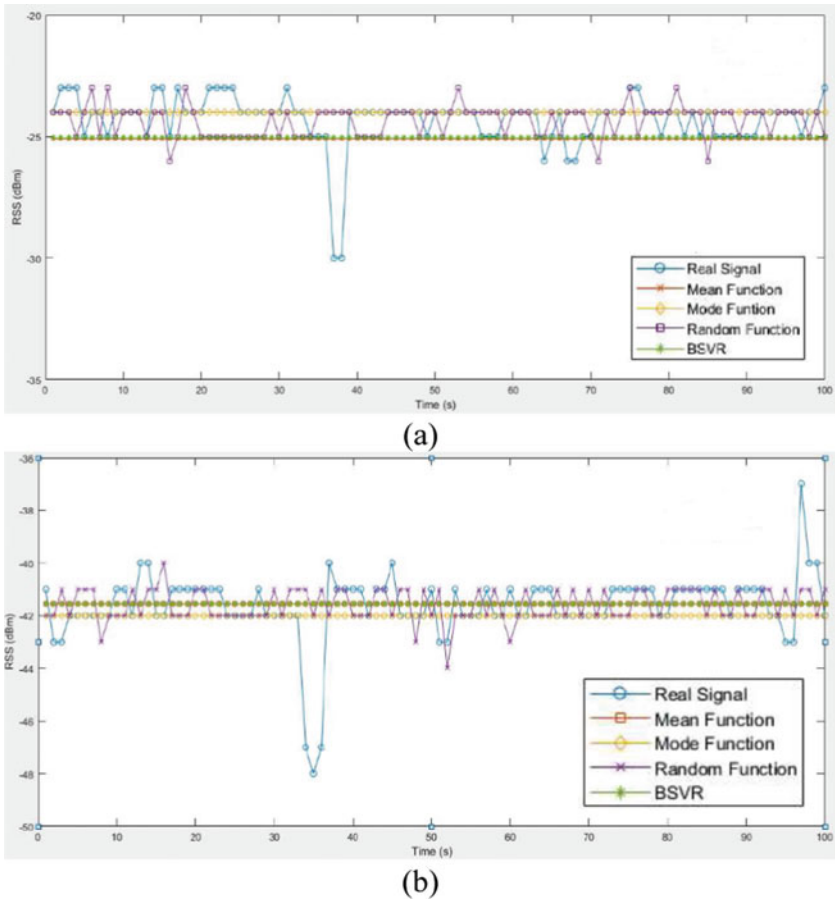


Fig. 3 The predicted signals by 4 models against real signal at distance **a** 1 m and **b** 5 m

pattern predicted by 4 models against real signal at distance of 1 m. From the figure, it could be observed that the random function model best represents the signals while other models generate a straight-line signal prediction with the maximum training size 1100. The random function predicted signal is more similar than other models to the actual signal. From Fig. 3(b), it is observed that for distance 5 m, mean, mode and BSVR function all had predicted a straight-line signal at training size of 1100, except for random function which predicts a fluctuating RSS against time. The random function could predicted signal closest to the actual signal compared to other techniques.

Table 1 and 2 shows the mean errors for each of the models in different train size and distance. Table 1 shows the performance of the models in train size 1100 and with different distance signals of 1, 2, 3, 4, 5, and 6 m. From the table, it can be observed that the lowest mean error of 0.973 m at signal distance 2 m when using the

Table 1 The RMSE of all 4 models at distance 1–6 with validation size 100

Signal distance	Validation size 100, Test size = 1100					
	1 m	2 m	3 m	4 m	5 m	6 m
Mean	3.813	0.988	1.034	1.033	1.190	2.324
Mode	3.799	1.001	1.096	1.060	1.269	2.535
Random function	3.906	1.118	1.459	1.395	1.313	2.421
BSVR	3.803	0.973	1.096	1.150	1.166	1.830

Table 2 The RMSE of all 4 models at distance 1–6 with validation size 1100

Signal distance	Validation size 1100, Test size = 100					
	1 m	2 m	3 m	4 m	5 m	6 m
Mean	1.347	0.523	0.822	0.945	1.303	2.410
Mode	1.105	0.510	1.054	1.005	1.349	2.476
Random function	1.241	0.775	1.049	1.204	1.432	2.567
BSVR	3.042	0.581	0.894	1.195	1.375	1.388

BSVR algorithm. As the validation size increases, the mean error for all the models gets lower as shown in Table 2. In Table 2, the lowest mean error is when using model mode which is 0.510 m at signal distance within 2 m.

Additionally, it is observed that at further distance increment, the BSVR model performs better for all the train size. The BSVR model has a lowest root mean square error compared to all other models at distance at 6 m. This could be due to that the BSVR performs better at predicting signals with more fluctuations. Prediction from mean and mode model does had representation of fluctuation of RSS. The random function model only represents the fluctuation by ± 2 dBm compared to BSVR. However, at distance of 1 m, BSVR has the worst performance compared to other models. It may be due to the two sharp spikes presented in the signals that suddenly appeared within the signal timeframe. The BSVR was much more sensitive to noise compared to the mean and mode models which under large train size less sensitive to signal noises. The random function signal only fluctuates between ± 2 dBm, which could mitigate signal noise since it does not consider those signals outside the range of ± 2 dBm.

Figure 4(a) and (b) shows the e performance of the models against training size separated into each graph by distance away from access point. From Fig. 4(a), it could be concluded that for signal of 1 m between transmitter and the receiver, the mean, mode, and random function performs better with increasing train size N, while BSVR model performance is slightly worse than the other model by at most 2 dBm. The figure also shows that the mode function has the least RMSE which gives the best performance amongst the other models. The mean function performs slightly worse than the random function model, and the BSVR model has the highest root mean square error.

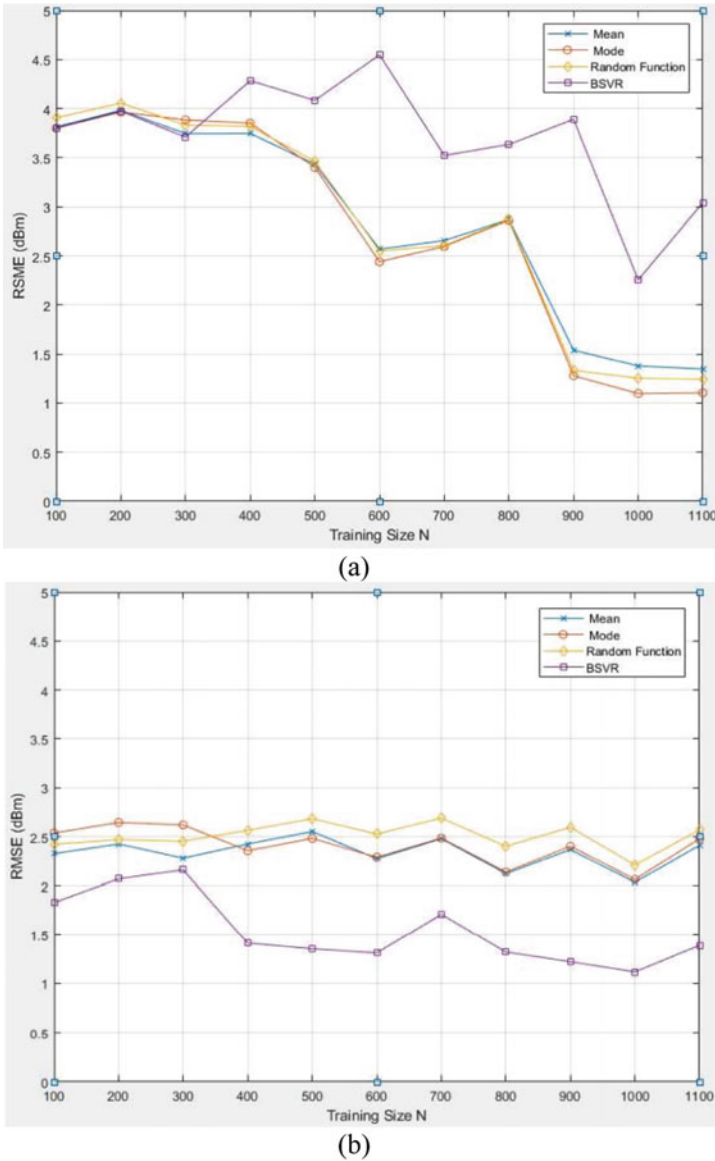


Fig. 4 RMSE plot of various model in different train size at distance a 1 m and b 5 m

At the signal distance of 6 m as shown in Fig. 4(b), the BSVR has a very clear advantage compared to the other algorithms. The mean model and the mode model had large root mean square error when approaching larger training size, while the random function model also performs worse. From all the figures above, it can be concluded that for all distance, BSVR had the overall best performance on signal,

especially when the distance increases. This is due to the reason that BSVR can handle fluctuation of signal better which happens when distance away of the receiver from AP increases.

5 Conclusion

The main objective of this paper is to provide ways of approach to explore the potential of pattern predicting algorithms on improving indoor positioning. By prediction of RSS, the training phase of fingerprinting can be automated as well as greatly reduced. The objective is achieved by implementing 4 prediction models, and the predicted signals are analyzed according to the pattern similarity and the difference with the real signal. For the pattern, the real signal could be replicated the best by the Random Function model if the fluctuation is within ± 2 dBm. On the other hand, the BSVR model could replicate the signal better at 2 m and at 6 m. The BSVR model could only produce a straight-line prediction at 1, 3, 4, 5, meter which deviates from the real signal. At distance further away from the access point, the BSVR had the best performance at predicting the signals as it has the least root mean square error.

Acknowledgements This work is financially supported by Universiti Putra Malaysia and the Ministry of Higher Education, Malaysia under Fundamental Research Grant Scheme (FRGS) with reference number of FRGS/1/2017/TK04/UPM/02/5.

References

1. Sarkisian L, Mickley H, Schakow H, Gerke O, Jørgensen G, Larsen ML, Henriksen FL (2020) Global positioning system alerted volunteer first responders arrive before emergency medical services in more than four out of five emergency calls. *Resuscitation* 152:170–176. <https://doi.org/10.1016/j.resuscitation.2019.12.010>
2. Walton AJ, Black RJ (1999) The global positioning system. *Phys Educ* 34:37–42. <https://doi.org/10.1088/0031-9120/34/1/019>
3. Alade IO, Abd Rahman MA, Abbas Z, Yaakob Y, Saleh TA (2020) Application of support vector regression and artificial neural network for prediction of specific heat capacity of aqueous nanofluids of copper oxide. *Sol Energy* 197:485–490
4. Alade IO, Rahman MAA, Hassan A, Saleh TA (2020) Modeling the viscosity of nanofluids using artificial neural network and Bayesian support vector regression. *J Appl Phys* 128:85306. <https://doi.org/10.1063/5.0008977>
5. Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst* 4:2951–2959

An Approach to Neutral Steering of a 4WIS Vehicle with Yaw Moment Control



Li Maoqi, M. I. Ishak, P. M. Heerwan, and Muhammad Aizzat Zakaria

Abstract There are various Advanced Driver Assistance Systems (ADAS) available in the automotive industry which are developed to improve the safety of vehicles while driving. Vehicle stability system such as yaw moment control is one of ADAS that is common only on luxury vehicles with oversteer (OS) characteristics. However, the majority of vehicles in the market are designed with understeer (US) characteristics. The yaw rate of the vehicle increases gradually at low speed and can easily maneuver during cornering. The vehicle becomes uncontrollable especially for novice drivers when the yaw rate saturates at a certain level during high-speed cornering. Differential drive method as yaw moment control system has low precision as it shares the same function for the longitudinal stability control system. In this paper, we proposed four-wheel independent steering (4WIS) to improve the yaw moment of an understeer test vehicle. A steady-state cornering (SSC) simulation was performed to obtain the steering characteristic of a test vehicle. The yaw rate for the test vehicle is linearized to create a state-space linear model in the simulation. Then, SSC simulations were repeated with the 4WIS input is obtained by a PID control and the state-space as the reference value. The results show that the 4WIS system was able to improve the maneuverability of the understeer vehicle, especially at high speed by shifting the yaw rate from US to a neutral steer characteristic.

Keywords ADAS · 4WIS · PID control system · Yaw moment control · Neutral steering

L. Maoqi · M. I. Ishak (✉) · P. M. Heerwan

Department of Mechanical Engineering, Collage of Engineering, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia

e-mail: mizhar@ump.edu.my

M. I. Ishak · P. M. Heerwan

Automotive Engineering Center, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

M. A. Zakaria

IMAMS Laboratory, Faculty of Manufacturing and Mechatronics Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

1 Introduction

The steering characteristics are particularly important for vehicles. There are three steering characteristics of vehicles: understeer (US), neutral steer (NS), and oversteer (OS). Except for some expensive vehicles that have OS features, most vehicles on the market have US features [1]. The steering characteristics of NS is between US and OS. From the Table 1 we can observe the advantages and disadvantages for US and OS. In general, the characteristics of the US can make the vehicle more stable, and the characteristics of the OS can make the vehicle more sensitive. Because different drivers have different driving habits, and different vehicles have different roles. Therefore, it is difficult to determine which steering characteristic is better. When the steering characteristic of the vehicle approaches NS, the advantages and disadvantages will be reduced, but from the perspective of safety, NS is the optimal solution.

Advanced driver assistance system (ADAS) as a system to improve vehicle performance and improve the driver’s experience [2–4]. It contains many systems, such as an anti-lock braking system (ABS), adaptive cruise control (ACC) system, and so on [5]. These systems are based on the vehicle’s feedback information to control the vehicle. For example, the ACC system, after setting the speed, the ACC system can be separated from the driver to carry out automatic driving at a constant speed and keep a safe distance from the front and rear vehicles [6]. However, the system relies too much on the sensors and other equipment on the vehicle [7]. Once there is a problem with the relevant equipment, it is easy to threaten the safety of the driver [8]. In addition, the lateral stability system is a type of ADAS and is essential to the driver’s driving experience and safety. Therefore, an ADAS controlled by the driver and assisted by an automatic system is the direction of development. Four-wheel independent steering (4WIS) system is a system that used steer-by-wire technology and it’ can be a novel method in ADAS.

4WIS system is a steering system that all the wheels of the vehicle can be rotated independently [9]. 4WIS has higher flexibility comparing with two-wheel steering (2WS), and it can make the vehicle easier to control when turning. The 4WIS system is used to improve steering response, increase vehicle stability at high speeds, or reduce turning radius at low speeds [10]. Its working principle is to increase or decrease the yaw rate of the vehicle through the rotation of the four wheels, thereby

Table 1 The advantages and disadvantages of US and OS [1]

Items	Advantages	Disadvantages
Understeer	Steadier	Less responsive
	Easy on tires	Slower turn in
	Better in rain	Front loses grip
Oversteer	More responsive	Less stable
	Faster turn in	Harsh on tire
	faster corner speeds	Rear loses grip

changing the direction of the vehicle [11, 12]. The 2WS model is used in this article to test the steering characteristics of the vehicle in order to provide comparative data on the yaw rate for the 4WIS in the following. As a kind of ADAS, the 4WIS system is characterized in that when the vehicle is running normally, the system will be in a standby state. When the yaw rate of the vehicle is different from the normal yaw rate of the vehicle in the current state, the system will increase or decrease the yaw rate of the vehicle by increasing the steering angle of the rear wheels automatically, so that the vehicle can drive normally [13, 14]. This study assumes that the steering angle of the rear wheels is positive to the left and negative to the right. When the rear wheels are not turning, the steering angle is zero. Whether it is turning left or right, the steering angle is substantially increased. When the vehicle turns to the left, it is assumed that the rear wheel steering angle is positive and that it turns to the right is negative. The front wheel steering angle is turning to the left during simulation.

In this study, the PID control system was used to study the control performance of the 4WIS system on the yaw rate during steady-state cornering (SSC) by control the rear wheels of the vehicle. PID control system is widely used because of its simple principle and strong system stability [15]. The results of this research can further improve the turning characteristics of the vehicle and thus the safety performance of the vehicle when turning at high speed. The author hopes that the application of the 4WIS system can further improve ADAS and give a solid example of an automatic control system for the 4WIS system so that future researchers to undertake the in-depth study.

2 Methodology

2.1 Block Diagram for the Simulation

Since this research is a simulation, the block diagram of the simulation is shown in Fig. 1. The leftmost part is the state space, which is used to provide the reference yaw rate required by the PID control system. It is used to receive the results of the vehicle

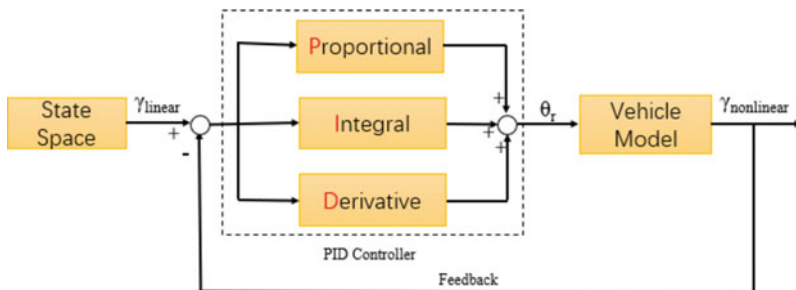


Fig. 1 Block diagram for the simulation

model and state space, and to control the vehicle model according to the yaw rate provided by the state space. The part in the dashed frame is the PID control system. The rightmost part is a vehicle model based on nonlinear dynamic equations. It is used to receive the steering angle of the rear wheels provided by the PID controller and simulate the yaw rate of the vehicle model [16]. The feedback system will feed back the simulated yaw rate to the PID controller, so that the simulated yaw rate is the same as the reference yaw rate.

Vehicle Model

The vehicle model was created by nonlinear dynamics equation of motion in MATLAB software 2013a version which the parameter from the test car and the type of test car is proton persona. For the parameter of the test car which used in all the following equations, the author was mentioned in another paper [14]. The nonlinear dynamics equation of motion shown as bellow:

$$m \left(\frac{du}{dt} - v\gamma \right) = (X_{FR} + X_{FL}) \cos \theta_F + (X_{RR} + X_{RL}) \cos \theta_R - (Y_{FR} + Y_{FL}) \sin \theta_F - (Y_{RR} + Y_{RL}) \sin \theta_R \quad (1)$$

$$m \left(\frac{dv}{dt} + u\gamma \right) = (X_{FR} + X_{FL}) \sin \theta_F + (X_{RR} + X_{RL}) \sin \theta_R + (Y_{FR} + Y_{FL}) \cos \theta_F + (Y_{RR} + Y_{RL}) \cos \theta_R \quad (2)$$

$$I \frac{d\gamma}{dt} = I_F [(X_{FR} + X_{FL}) \sin \theta_F + (Y_{FR} + Y_{FL}) \cos \theta_F] + I_R [(X_{RR} + X_{RL}) \sin \theta_R + (Y_{RR} + Y_{RL}) \cos \theta_R] + \frac{d_F}{2} [(X_{FR} + X_{FL}) \cos \theta_R + (Y_{FR} + Y_{FL}) \sin \theta_F] + \frac{d_R}{2} [(X_{RR} + X_{RL}) \cos \theta_R + (Y_{RR} + Y_{RL}) \sin \theta_R] \quad (3)$$

The State Space and the PID Control

Figure 2 is the reference yaw rate of the PID control system. The state space model has a neutral steering characteristic produced by linearizing the SSC yaw rate of the test vehicle. It can be seen from the Figure that as the speed increases, the value of the yaw rate gradually increases. This is a manifestation of neutral steering. The linear dynamics equation of motion is shown as:

$$mV \frac{d\beta}{dt} + 2(K_f + K_r)\beta + \left\{ mV + \frac{2}{V} (l_f K_f - l_r K_r) \right\} \gamma = 2K_f \theta \quad (4)$$

$$mV \frac{d\beta}{dt} + 2(K_f + K_r)\beta + \left\{ mV + \frac{2}{V} (l_f K_f - l_r K_r) \right\} \gamma = 2K_f \theta \quad (5)$$

If the vehicle steering characteristic is neutral steering, it is also means that $l_f k_f - l_r k_r$ is equals to zero [17]. Since the vehicle's center of gravity to the length of the

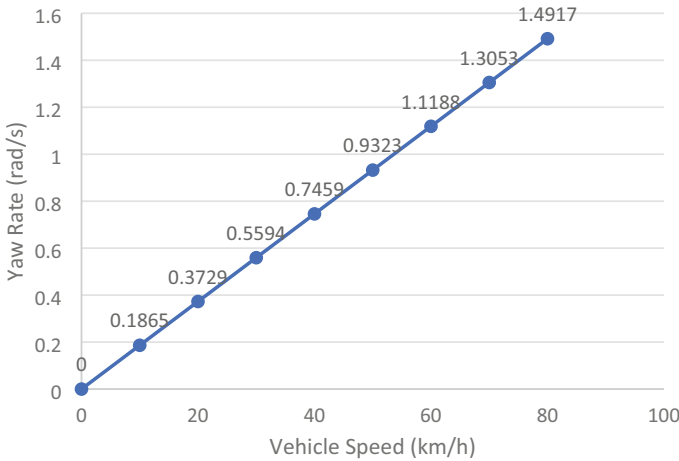


Fig. 2 The relationship between the reference yaw rate and the vehicle speed

front and rear wheels is already a fixed value, and only the cornering stiffness can be changed. The method used by the author here: first fix the cornering stiffness of the front wheel, and then deform the above formula to find the value of the rear wheel cornering stiffness that can make the formula hold.

Figure 2 is the relationship between speed and reference yaw rate based on the final constant value of the yaw rate at each speed. These yaw rate comes from the linear dynamics equation by changes the cornering stiffness K of the rear wheels. Figure 2 shows intuitively that the yaw rate increases regularly with the speed increase. And the graphs in the figure present a linear straight line. The straight line represents the yaw rate of the vehicle in this state as neutral steering which also can call the neutral steering as a good SSC [17]. Although the yaw rate in Fig. 2 seems to be no problem, it can still be found through careful observation that the value of the reference yaw rate is very large. In fact, due to the yaw rate only represents the ratio of the vehicle's yaw to time, and it has nothing to do with whether the control system can control the vehicle to achieve a constant yaw rate. Therefore, the magnitude of the yaw rate does not affect this study. If the PID control system can control the vehicle to achieve the yaw rate for the reference value of the abnormal yaw rate, it can be proved that the PID control system can be used on a vehicle with a 4WIS system to deal with US and OS phenomenon. Therefore, although the yaw rate is 1.4917 rad/s when the vehicle velocity is 80 km/h, it is still can be used as the reference yaw rate for the PID control system.

2.2 Simulation Procedural

This article first uses vehicle dynamics equations to build a 2WS vehicle model in MATLAB Simulink software. Perform SSC with constant velocity from 10 to 80 km/h in the 10 km/h increment. The front steer angle is set to a constant angle of 10 degree and the rear angle is 0°. At each SSC simulation, the yaw rate is recorded in MATLAB software and export as excel file. By simulating the steering of the vehicle, the steering characteristic of the vehicle model is determined to be US. After that, based on the result of the yaw rate in the previous SSC, we linearize the yaw rate to determine the NS parameter in the state space. the 4WIS system controlled by the PID control system was applied to the vehicle model. The 4WIS system is implemented to improve the steering performance of the vehicle. Its main performance is to compare the yaw rate before and after the application of the PID control system to determine that the vehicle controlled by the PID control system has SSC in a high-speed turning state.

3 Results and Discussion

3.1 Steering Characteristic of the Simulation Model by Steady State Cornering Test

Figure 3 shows the simulation result of the yaw rate during the 2WS SSC test at various constant speeds from 10 to 80 km/h with incremental of 10 km/h and a

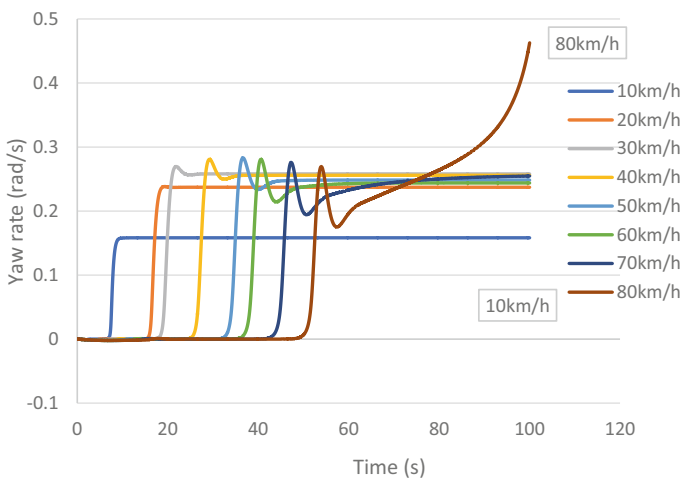


Fig. 3 The yaw rate change of the 2WS vehicle from 10 to 80 km/h

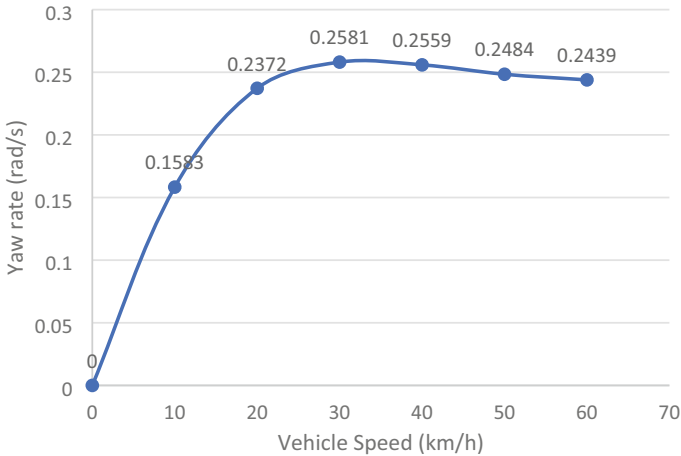


Fig. 4 The intuitive displayed the yaw rate for 2WS vehicle

constant front steer angle of 10°. When the speed is from 0 to 60 km/h, the yaw rate of the vehicle can reach a constant value. This phenomenon means that the vehicle can perform SSC at this speed. However, the value of the yaw rate is at the lowest when the speed is 10 km/h. When the vehicle speed is 70 km/h, the yaw rate cannot reach a constant value and maintain a slow increase. When the vehicle speed is 80 km/h, the yaw rate continues to rise and at the end, there is a rapid rise to infinity. This means the vehicle is spinning rapidly in this condition, and the speed of rotation is getting faster and faster.

Figure 4 shows the relationship between the yaw rate to the constant velocity during the SSC simulations. The yaw rate is increasing rapidly from 0 until 20 km/h. Beyond that velocity, the yaw rate began to saturate approximately at 0.25 rad/s. And the last constant velocity that the vehicle can achieved steady state cornering is at 60 km/h. Based on the profile of the yaw rate from the steady-state cornering simulation, we can conclude that the actual test car has an understeer characteristic. If a vehicle has a neutral steer characteristic, the yaw rate will increase proportional to the increase of constant velocity during cornering. However, if the vehicle has an oversteer characteristic, the yaw rate line increases sharply with vehicle velocity and becomes infinity.

3.2 Yaw Rate of 4WIS Vehicle After Implemented PID Control System

Figure 5 shows the value of the yaw rate controlled by the PID control system at different vehicle speeds. Even if the vehicle speed reaches 80 km/h, the PID control system can still effectively control the vehicle and have a constant yaw rate. Although

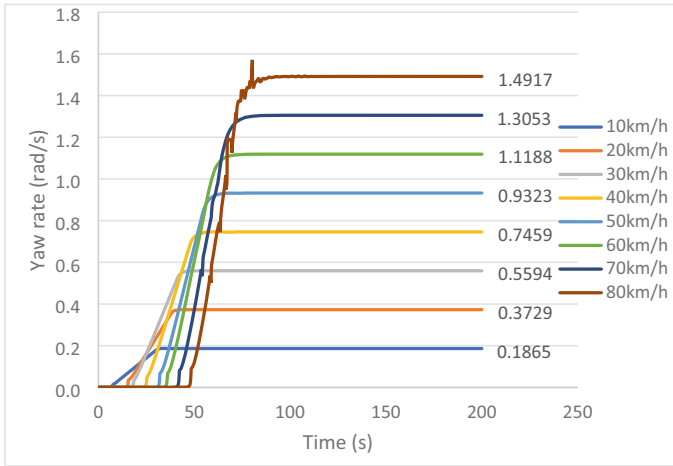


Fig. 5 The yaw rate after implement the PID control system

the yaw rate fluctuates greatly when the speed is 80 km/h and the value of the yaw rate fluctuates slightly after the steering is completed, a constant value of the yaw rate is finally reached. We could observe that both yaw results are same by compared the results between the reference yaw rate and the yaw rate after implemented the PID control system. Therefore, we can conclude that the PID control system can make the vehicle have a steady-state cornering when the PID control can control the 4WS system vehicle efficiently.

3.3 Rear-wheel Steering Angle of 4WIS Vehicle After Implement PID Control System

Figure 6 shows the steering angle of the rear wheels controlled by the PID controller. From the figure, we can observe that when the vehicle speed is less than 60 km/h, the steering angle of the rear wheels is negative, and the minimum value is negative 5.41. When the speed is greater than 60 km/h, the steering angle of the rear wheels is positive, and the maximum value is positive 4.35. When the vehicle speed is from 0 to 60 km/h, the value of the steering angle first decreases and then increases. When the vehicle speed is 70 to 80 km/h, the degree of steering angle is relatively similar. These data show that when the turning speed of the vehicle is less than or equal to 60 km/h, the direction of rotation of the front and rear wheels is opposite. When the turning speed of the vehicle is greater than 60 km/h, the direction of rotation of the front and rear wheels is the same.

Figure 7 shows the yaw rate at a different speed between the different conditions of the 2WS vehicle model without a control system which is “non-control yaw rate”

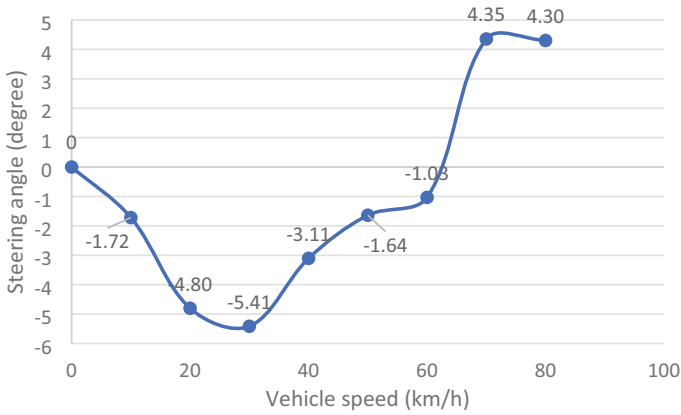


Fig. 6 The rear-wheel steering angle after implement the PID control

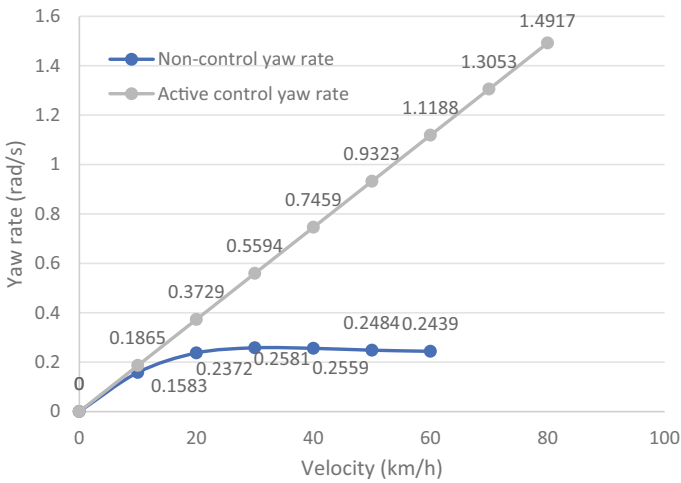


Fig. 7 Comparing the change of the yaw rate at various speeds between different control system conditions

in the Fig. 7, and the 4WIS vehicle model with an active control system. The yaw rate increased gradually by implementing the active control system from 0 to 80 km/h. For the vehicle model without a control system, the yaw rate increase at the velocity from 0 to 30 km/h. When the velocity higher than 30 km/h, the yaw was decreased with the velocity decrease. By compare the yaw rate between the vehicle with the active control and non-control system, the active control system made the vehicle had steady-state cornering at higher speed condition.

4 Conclusion

As a summary, this research first uses nonlinear dynamics equation of motion to establish a vehicle model in MATLAB software and determine the characteristics of the vehicle through simulation. Second, use the result of the linear dynamics equation of motion as the reference yaw of the PID control system. Finally, by establishing a PID controller, the vehicle model becomes a 4WIS system vehicle model with a PID control system. The simulation results show that when the vehicle is under-steered or over-steered when turning, the 4WIS system can be added to make the vehicle have an SSC. The results of this research can be used as a new type of ADAS to further improve the turning characteristics of the vehicle and thus improve the safety performance of the vehicle when turning at high speeds.

Acknowledgements This research was supported by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme FRGS/1/2019/TK08/UMP/02/5. Special thanks to Automotive Engineering Center, Universiti Malaysia Pahang for providing laboratory facilities and technical support.

References

1. Oversteer vs understeer: which is better? - FLOW RACERS. <https://flowracers.com/blog/oversteer-vs-understeer/>. Accessed 02 Sept 2021
2. Orlovska J, Novakazi F, Lars-Ola B, Karlsson M, Wickman C, Söderberg R (2020) Effects of the driving context on the usage of Automated Driver Assistance Systems (ADAS)-Naturalistic Driving Study for ADAS evaluation. *Transp Res Interdiscip Perspect* 4:100093
3. Solmaz S, Holzinger F, Mischinger M, Rudigier M, Reckenzaun J (2021) Novel hybrid-testing paradigms for automated vehicle and ADAS function development. In: *Towards connected and autonomous vehicle highways*. Springer, pp 193–228
4. Han S et al (2021) ADAS-viewer: web-based application for integrative analysis of multi-omics data in Alzheimer's disease. *npj Syst Biol Appl* 7(1):1–10
5. Zulhilmi IM, Peeie MH, Asyraf SM, Sollehudin IM, Ishak IM (2020) Experimental study on the effect of emergency braking without anti-lock braking system to vehicle dynamics behaviour. *Int J Automot Mech Eng* 17(2):7832–7841
6. Perrier MJR, Louw TL, Carsten O (2021) User-centred design evaluation of symbols for adaptive cruise control (ACC) and lane-keeping assistance (LKA). *Cogn Technol Work*, pp 1–19
7. Ziebinski A, Cupek R, Erdogan H, Waechter S (2015) A survey of ADAS technologies for the future perspective of sensor fusion. In: *International conference on computational collective intelligence*, pp 135–146
8. Zhang S, Shi J, Guo K, Wang Y (2020) Virtual validation method for automated driving vehicles based on traffic accident. *CICTP 2020*:365–375
9. Liu Y, Zong C, Zhang D, Zheng H, Han X, Sun M (2021) Fault-tolerant control approach based on constraint control allocation for 4WIS/4WID vehicles. *Proc Inst Mech Eng Part D J Automob Eng* 235(8):2281–2295
10. Ishak MI, Ogino H, Yamamoto Y (2015) Numerical simulation analysis of an oversteer in-wheel small electric vehicle integrated with four-wheel drive and independent steering. *Int J Veh Technol* 2016

11. Song YT, Shu HY, Chen XB (2020) Chassis integrated control for 4WIS distributed drive EVs with model predictive control based on the UKF observer. *Sci China Technol Sci* 63(3):397–409
12. Yim S (2021) Integrated chassis control with four-wheel independent steering under constraint on front slip angles. *IEEE Access* 9:10338–10347
13. Maoqi L, Ishak MI, Heerwan PM (2021) Investigation on linear and nonlinear dynamic equation for vehicle model in numerical simulation. *IOP Conf Ser Mater Sci Eng* 1078(1):012010
14. Maoqi L, Ishak MI, Heerwan PM (2019) The effect of parallel steering of a four-wheel drive and four-wheel steer electric vehicle during spinning condition: a numerical simulation. *IOP Conf Ser Mater Sci Eng* 469(1)
15. Liu Y, Fan K, Ouyang Q (2021) Intelligent traction control method based on model predictive fuzzy PID control and online optimization for permanent magnetic Maglev trains. *IEEE Access* 9:29032–29046
16. Yu W (2018) Chapter 2 - stable PID control and systematic tuning of PID gains. In: Yu W (ed) *PID control with intelligent compensation for exoskeleton robots*. Academic Press, pp 13–33
17. Abe M (2015) *Vehicle handling dynamics: theory and application*. Butterworth-Heinemann

Investigation of Electroencephalogram (EEG) Sensor Position for Brain-Controlled Home Automation



Norizam Sulaiman, Ailis Aimyilia Hasim, Md Nahidul Islam, Mahfuzah Mustafa, and Mohd Shawal Jadin

Abstract Electroencephalogram (EEG) signals are widely employed in Brain-Computer Interface (BCI) or Human–Machine Interface (HMI) technique to provide assistive technology that can be used by paralyzed or disabled people. BCI or HMI technique will let paralyzed people to operate home automation such as lamp, fan, television and other home appliances by using their brainwaves. Since BCI or HMI is constructed from various sensors for various measurement position, it is vital to know which EEG sensors are really contributed to the control of the device. Thus, this study is conducted to investigate the EEG sensors position in controlling device by analyzing public EEG datasets. The scopes of the study include the construction of Graphical User Interface (GUI) in MATLAB for each selected sensor position and the classification of the selected EEG features from each sensor position. To implement the investigation process, first, the public EEG signals from the specified sensor location are filtered using preprocessing technique to remove the noise or artifacts such eyes movement and power line noise. Next, the filtered EEG signals are split to Alpha and Beta power spectrum using Fast Fourier Transform (FFT) technique. Next, the unique features from EEG Alpha and Beta Power are extracted for classification process in term of average (mean), standard deviation and spectral centroid. Finally, the prototype model using Arduino microcontroller is developed to implement home appliance. The results of the study show that the selected EEG features from the EEG signal produced by EEG position at Frontal side of brain lobes (F7 and F8) and Parietal side of brain lobes (P3 and P4) able to be classified with 83% classification accuracy. The selected EEG features from the selected sensors position can be converted to machine code to control the home appliance successfully.

N. Sulaiman (✉) · A. A. Hasim · M. N. Islam · M. Mustafa · M. S. Jadin
Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia
e-mail: norizam@ump.edu.my

M. Mustafa
e-mail: mahfuzah@ump.edu.my

M. S. Jadin
e-mail: mohdshawal@ump.edu.my

Keywords EEG · BCI · HMI · GUI · Power spectrum · Features · Classification

1 Introduction

Nowadays, the Brain-Computer Interface (BCI) or Human–Machine Interface (HMI) technique is successful in providing excellent assistive technology to assist disabled or paralyzed people. BCI or HMI can be described as a direct communication system that enables the human brain to communicate with the digital computer to monitor external devices. To monitor this activity, most BCIs rely on electrical measures of brain activity and sensors located on human’s head. To record electrical activity generated from the change in human cognitive states, Electroencephalogram (EEG) is widely used. It is a device to measurement human neural activities [1–3]. Basically there 4 major components in BCI or HMI system which are signal acquisition, signal pro-processing, feature extraction and classification as shown in Fig. 1. Next, translation algorithm will be constructed to change the classified features into machine code which can control a device. A thorough analysis of EEG signals are required in order to provide unique EEG features that might related to the change in human cognitive state due to the some stimulation on the brain. The selection of correct EEG features with good classification accuracy will enable a device to be controlled [4–6].

The number and position of scalp electrodes or sensors are vital since the large number of scalp electrodes will enables the neural activities from whole brain lobe can be measured and analyzed. Here, human brain consist of several lobes such as Frontal lobe, Parietal lobe, Temporal lobe and Occipital lobe where each lobe has function cognitive function as shown in Fig. 2.

Based on Fig. 2, it is definitely a minimum of 4 EEG electrodes are required to measure the EEG signals from the whole brain region by placing 1 EEG electrode at each brain lobe. However, based on the international 10–20 system of electrode placement or montage, more than 1 EEG electrodes are required to be placed at each brain lobes in order to provide more precise electrical activities generated by neural movement at each brain lobes. However, the use of minimum number of EEG electrodes might reduce the system power consumption and complexity of

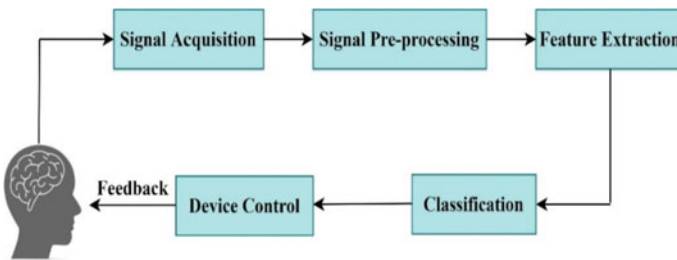


Fig. 1 Complete BCI system [1–3]

Fig. 2 Brain lobes and their functions [3, 4]

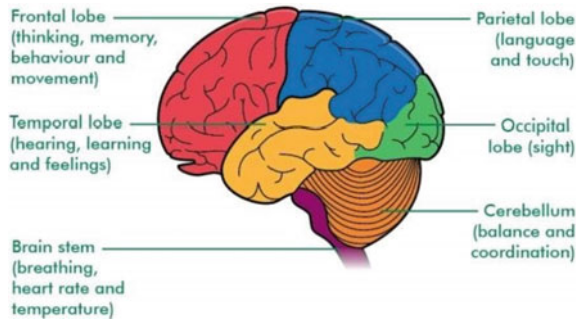


Table 1 EEG frequency bands and neural activities

EEG frequency bands	Frequency range (Hz)	Cognitive state
Delta	0.5–4	Deep sleep
Theta	4–8	Light Sleep
Alpha	8–13	Relax, closed eye
Beta	13–30	Alert, thinking

the computational algorithm in analyzing EEG features [7, 8]. Even though BCI processes and components are eminent as shown in Fig. 1, researchers had yet to identify which brain lobes and EEG sensor position that really contribute to the control of the device. Thus, this study is conducted to figure out which EEG sensor position in BCI system will provide precise EEG signals and EEG features to control device by analyzing public EEG datasets. Since EEG signals consists of several frequency rhythms, this study just focus on the EEG Alpha and Beta frequency band which represent human cognitive state in relax and alert (thinking) activities as shown in Table 1 [9, 10]. In BCI application, human must be in conscious and alert condition in order to control device.

This paper is organized by describing the objective of the study in Sect. 1, followed by methodology in Sect. 2 and results and discussion in Sect. 3. Finally, Sect. 4 will provide a conclusion of the study.

2 Methodology

This section describes the overall process of the study which include the process flow-chart and block diagram. Here, the source of the EEG datasets and analysis of the EEG datasets are explained in details.

Block diagram shown in Fig. 3 illustrates the methods to obtain and analyze EEG signals and then, the classified EEG features will be translated and converted to machine code to control home appliance. The process started by obtaining public

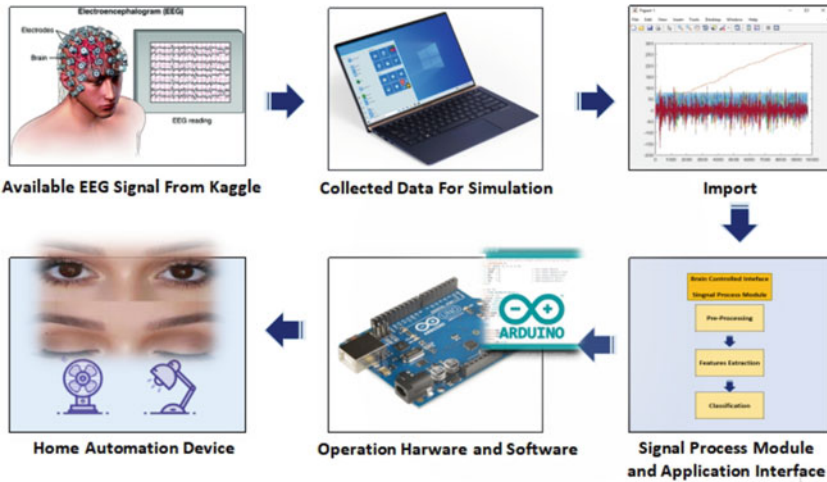


Fig. 3 Block diagram of the research

EEG datasets which was taken from *Kaggle's* website [13]. The raw EEG datasets that were taken from *Kaggle's* website then will be analyzed in computer using MATLAB R2018a software to obtain the required EEG features. The selected EEG features then will be extracted and classified to find the best EEG features that can be converted to machine code to control device using microcontroller. Here, the devices to be controlled using EEG signals are LED light and fan.

In this study, mean, standard deviation and spectral centroids are the selected EEG features and the features are then classified by k-NN classifier to search for the best features to be employed in controlling device.

2.1 Analyses of Public EEG Datasets

The overall process flow of the study is elucidated by Fig. 4. Here, EEG signals can be analyzed by using 2 types of EEG data; public EEG data or EEG data captured from the experimental work. Public EEG data is selected since nowadays due to Covid-19 pandemic, it is very hard to select the real subject and to conduct experimental work. Thus this study focuses on the available EEG datasets that related to BCI taken from *Kaggle's* website [13]. However, since there are various EEG datasets available online, the online EEG datasets are carefully investigated where the EEG datasets come from Frontal and Parietal of both brain hemispheres (right and left) are taken for analysis. EEG datasets that are captured from these regions are suitable to be analyzed in searching for suitable EEG features to control device. Here, EEG signals measured by EEG sensors or electrodes located at Pre-frontal lobe (Fp1 & Fp2), Frontal lobe (F7 & F8) and Parietal lobe (P3 & P4) are taken for analyses using MATLAB R2018a

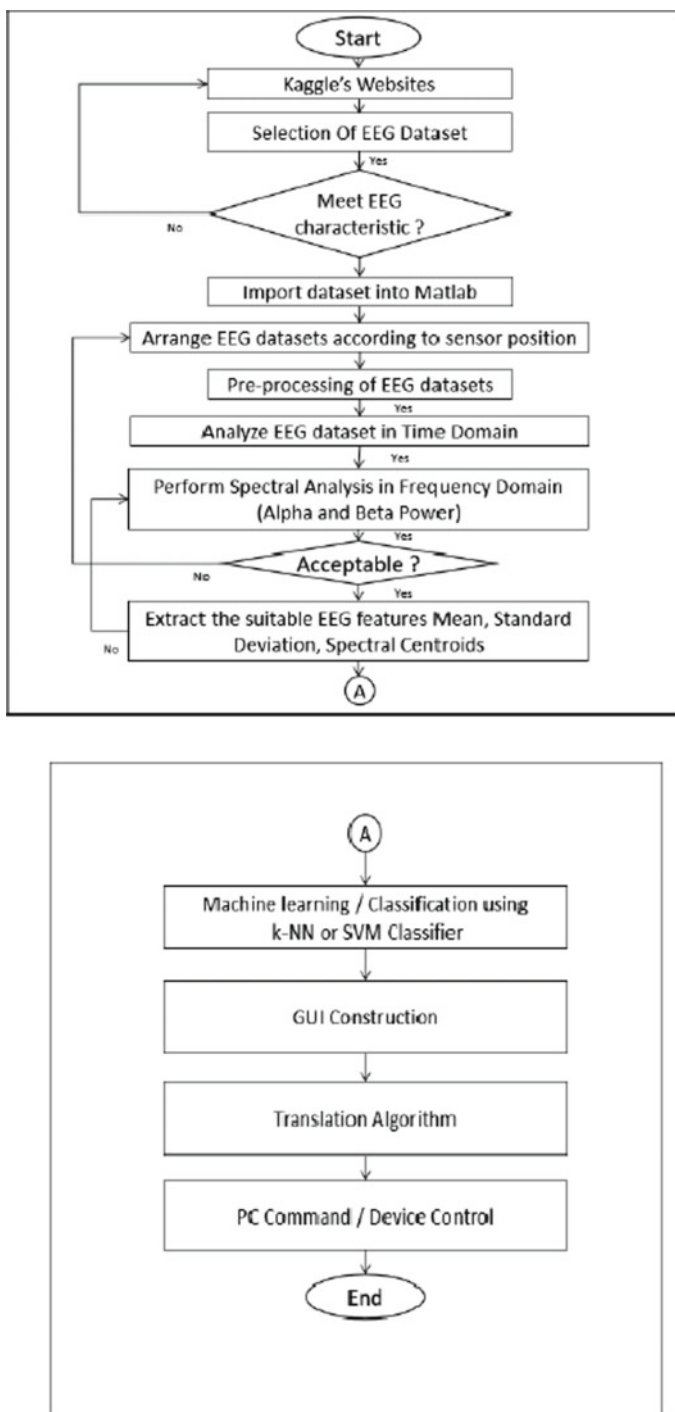


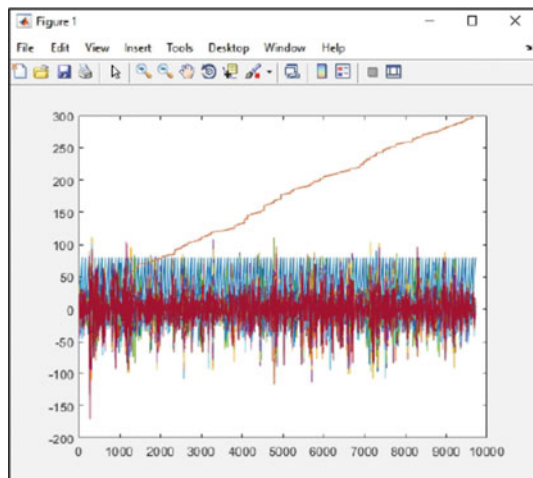
Fig. 4 Process flow chart of the research

software. The captured EEG datasets are inspected by plotting the EEG datasets in Excel format. Here, the EEG signals are supposed to be fluctuating between -100 microvolt (μV) and 100 microvolt (μV) [11, 12]. If the EEG specifications are not met, need to find another type of EEG datasets. The Raw EEG signals then will be filtered from artifacts or noises before converting them to power spectrum using Fast Fourier Transform (FFT) technique. Moreover, the EEG power spectrum then will be split to Frequency bands such as Alpha and Beta band to represent the brain cognitive state either in relax or alert mode. If the EEG power spectrum is not meet the EEG Power Spectral Density characteristic, EEG datasets from another sensor position will be taken for analysis. EEG datasets in term of EEG power spectrum then will be extracted to several features such as mean, standard deviation and spectral centroids. Next, the extracted EEG features are fed to k-NN classifiers. The classified extracted EEG features will be converted to class and machine code using translation algorithm in order to control the function of lights and fan. Graphical User Interface (GUI) is also constructed to show the overall process of analyzing EEG signals in term of EEG raw data, and mean, standard deviation, spectral centroids, of EEG power spectrum. The GUI also will indicate either the selected EEG features from the EEG sensor location contribute to the control of the device or not.

2.2 Example of Public EEG Datasets

The example of the plot of the public EEG datasets is shown in Fig. 5 where x-axis indicates the number of the EEG data (based on the duration of measurement). Meanwhile, y-axis indicates the Amplitude of EEG data. Here, the plot shows that EEG characteristics are met.

Fig. 5 Public EEG datasets



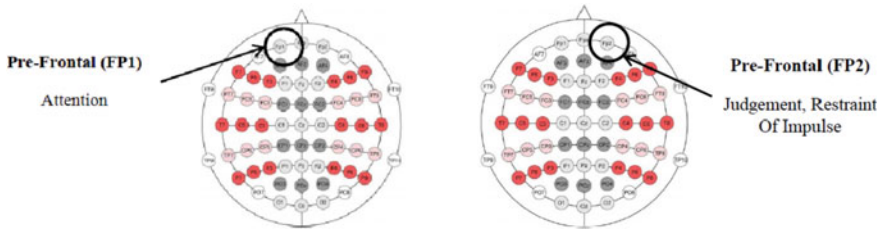


Fig. 6 Fp1 and Fp2 EEG Sensor position

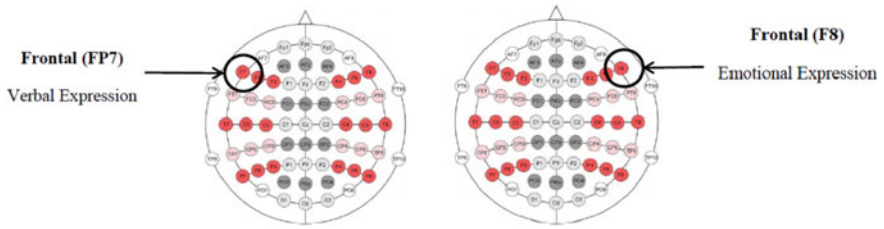


Fig. 7 F7 and F8 EEG sensor position

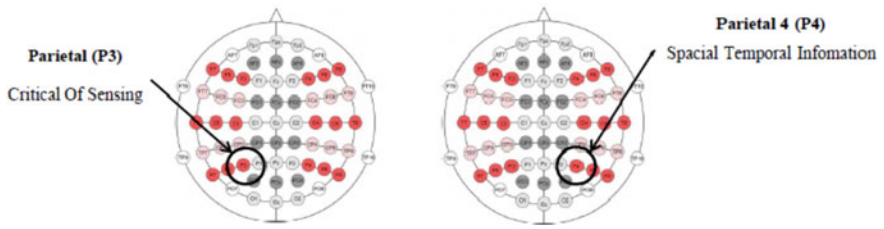


Fig. 8 P3 and P4 EEG sensor position

2.3 EEG Sensors/Electrode Position

In this study, EEG data are measured and analyzed from 6 EEG sensors position as shown in Fig. 6, 7, and 8 which the selected EEG sensors cover both brain hemispheres.

2.4 Feature Extraction

One of the major components of EEG-based BCI program is extraction of EEG features since incorrect selection of EEG features might provide incorrect instruction to device. As a result, device might malfunction. Thus, selecting the best feature is

Table 2 Formula of selected EEG features

Features	Formula
FFT	$f(x) = \sum_{n=1}^{N-1} (X(n)W_N^{kn})$ <p>Where, $W_N = e^{j \frac{2\pi}{N}}$</p>
PSD	$PSD = X(K) ^2$ $= \left \sum_{n=0}^{N-1} X(nTs)e^{-j \frac{2n\pi k}{N}} \right ^2$
Standard deviation	$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2}$ <p>Where, $\mu = \text{mean}$</p>
Mean	$Mean = \frac{1}{N} \sum_{i=1}^N x(i)$

very important to obtain the best output from the device. The formulas for the selected EEG features in frequency domain are Mean, Standard Deviation, FFT and PSD, as illustrated by Table 2.

2.5 EEG Feature Classification Using k-NN

k-NN is one of the powerful and robust linear classifiers but simple in classifying EEG features. In k-NN classification, the output is a class membership. It applies distance metric in finding the class most common among its k-nearest neighbors where k value is typically small which is [11, 12]. The classification accuracy is shown in Eq. 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1}$$

3 Results and Discussion

In this section, the results of the study are illustrated in term of Graphical User Interface (GUI). The GUI is constructed using MATLAB GUI. Here, user can select to analyze EEG data from several EEG measurement areas such as Fp1, Fp2, F7, F8, P3 and P4. Beside measurement area, the GUI consists of the plot of EEG raw data, FFT plot, Power Spectrum plot, Alpha Power, Beta Power, Spectrum Welch plot and

detection indicator. The plot of Alpha and Beta power is shown by bar chart where '1' indicates Alpha power and '2' indicates Beta power band.

The GUI in Fig. 9 indicates poor detection of the device. It is indicated by the red button in the GUI. Here, the plot of Alpha and Beta power spectrum meet the frequency range of Alpha and Beta band. However, since the mean of Alpha band is higher than mean of Beta band and very high standard deviation of Beta band compared to Alpha band, these EEG features that were obtained from Fp1 sensor position might not suitable to be converted to machine code to control device. Meanwhile, the GUI green indicator in Fig. 10 is turn on which indicates that the selected EEG features at F7 sensor position is suitable to be employed to control device. It might be caused by higher mean and standard deviation in Beta band compared to Alpha band.

The selected EG features are also plotted according to the sensor position as illustrated by Fig. 11, 12 and 13. For Fig. 11, it is very obvious that the mean value of EEG Beta power is higher than Alpha power at Fp2, F7, F8, P3 and P4 sensor position. Meanwhile, the standard deviation and spectral centroids value of EEG Beta power is higher than Alpha power at Fp1, F7, F8, P3 and P4 as illustrated by Fig. 12 and 13 respectively. This analysis of EEG features across the sensor position clearly indicate that which sensor position can provide good EEG features to be employed in controlling device.

The selected EEG features in term of mean, standard deviation and spectral centroids are classified using k-NN classifier. The accuracy of the classification is measured and shown in Fig. 14. Here, the classification accuracy for all selected EEG features is achieved at 83% which indicate that the classified features with high accuracy (83%) will be assigned to the class of '1'. Meanwhile, the classified features with low accuracy (17%) will be assigned to the class of '0'. The training and testing

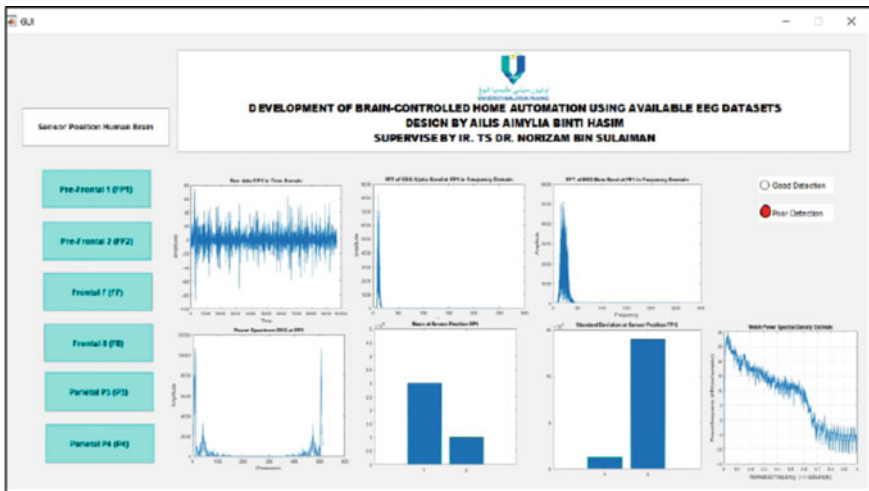


Fig. 9 GUI for Fp1 EEG measurement area

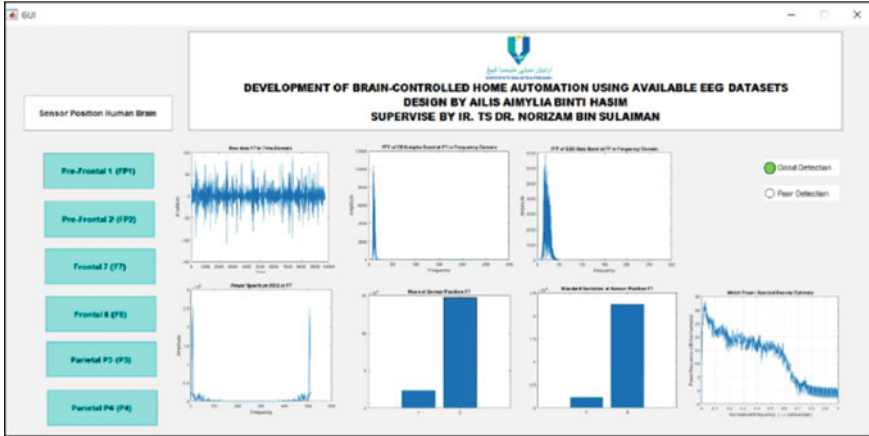


Fig. 10 GUI for Fp7 EEG measurement area

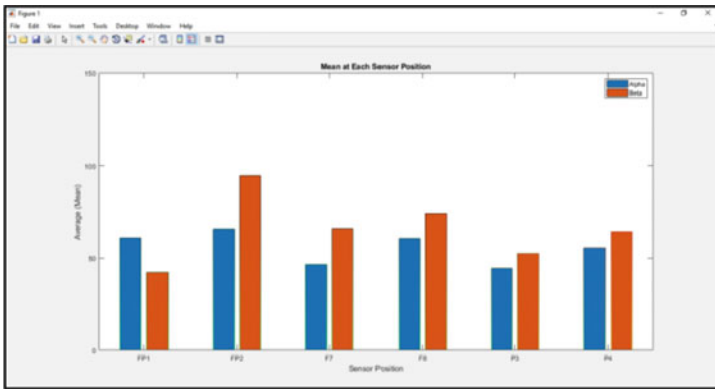


Fig. 11 EEG features (mean) across sensor position

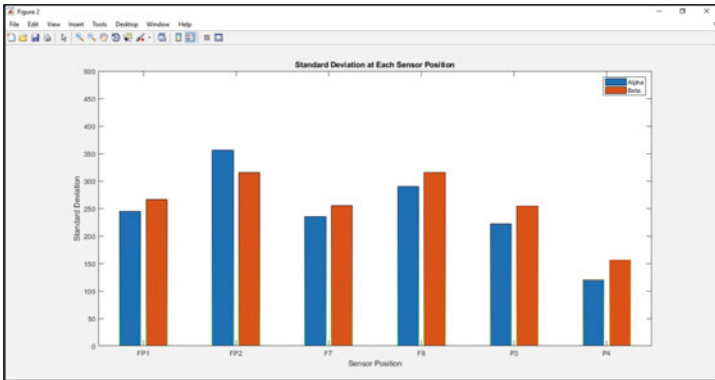


Fig. 12 EEG features (standard deviation) across sensor position

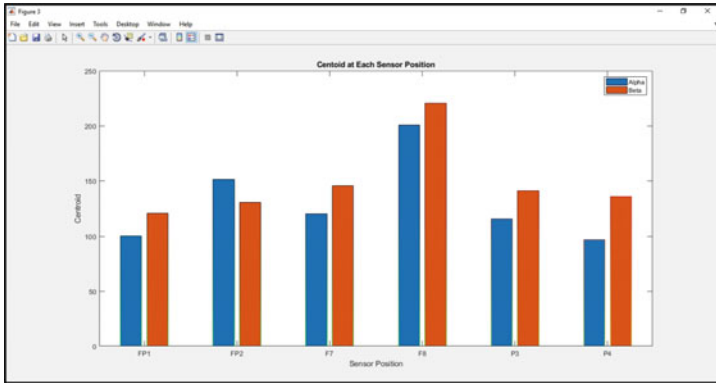
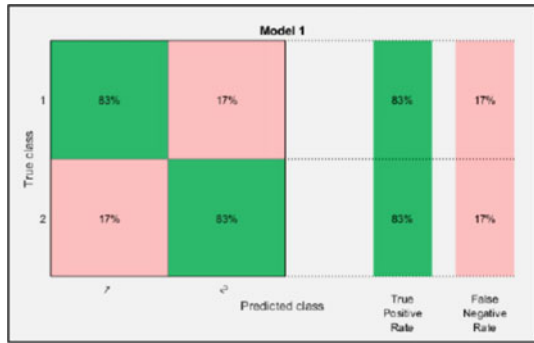


Fig. 13 EEG features (spectral centroids) across sensor position

Fig. 14 Confusion matrix for k-NN classification



ratio are set to 70:30 where 70% of the EEG data with the selected features will be trained and 30% will be tested. In addition, EEG sensor position might affect in controlling device as well. In this study, not all selected sensor position can provide good EEG features to control device as elucidated by Table 3. Meanwhile, good classification accuracy is obtained using all selected EEG features at sensor position F7, F8, P3, P4 which will turn on the device to indicate good device detection where the class is set to ‘1’.

The simulation of the circuit is implemented using Proteus and Arduino software as shown in Fig. 15. Here, Arduino microcontroller is connected to relay, fan and LED light. The classified EEG features are converted to machine code by using translation algorithm where the classified features will be assigned to the class of ‘1’ and ‘0’. Here, class ‘1’ indicates ‘high’ and will turn on the device such as LED light and Fan. Meanwhile, class ‘0’ indicates ‘low’ and will turn off the device. As illustrated by Table 3, the EEG features that generated by the EEG sensors at sensor position F7, F8, P3 and P4 will trigger the device operation to be in ‘ON’ status on certain duration before turn ‘OFF’ when class ‘0’ is triggered.

Table 3 Classification accuracy according to EEG sensor position and features

Sensor position	Device	EEG features	Accuracy (%)	Device operation status
Pre-Frontal 1 (Fp1)	LED light Fan	Mean	17	Poor Detection
		Standard Deviation	83	Good Detection
		Centroid	83	Good Detection
Pre-Frontal 2 (Fp2)	LED light Fan	Mean	83	Good Detection
		Standard Deviation	17	Poor Detection
		Centroid	83	Good Detection
Frontal 7 (F7)	LED light Fan	Mean	83	Good Detection
		Standard Deviation	83	Good Detection
		Centroid	83	Good Detection
Frontal 8 (F8)	LED light Fan	Mean	83	Good Detection
		Standard Deviation	83	Good Detection
		Centroid	83	Good Detection
Parietal 3 (P3)	LED light Fan	Mean	83	Good Detection
		Standard Deviation	83	Good Detection
		Centroid	83	Good Detection
Parietal 4 (P4)	LED light Fan	Mean	83	Good Detection
		Standard Deviation	83	Good Detection
		Centroid	83	Good Detection

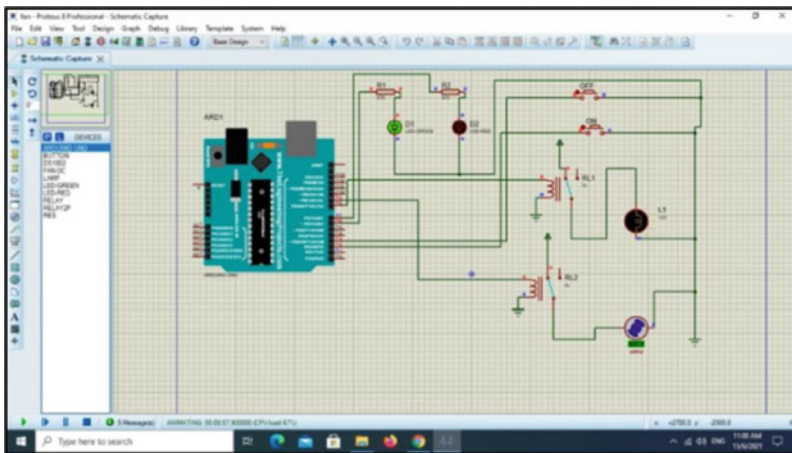


Fig. 15 Circuit simulation using Proteus and Arduino software

After circuit simulation is completed, the circuit is constructed and tested as elucidated by Fig. 16, 17 and 18. Figure 16 shows the hardware setup and components to test the output of the study. As illustrated by Fig. 17, the suitable EEG features from the precise EEG sensor position will turn on the white LED light and USB fan. Meanwhile, the red LED light will be turn on while turning off if the classified

Fig. 16 Device set-up

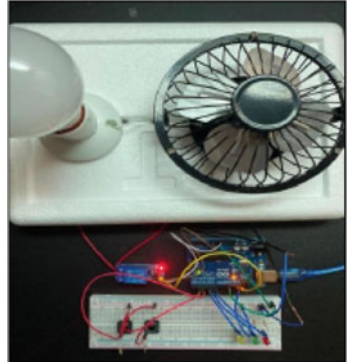


Fig. 17 Yellow LED light triggering

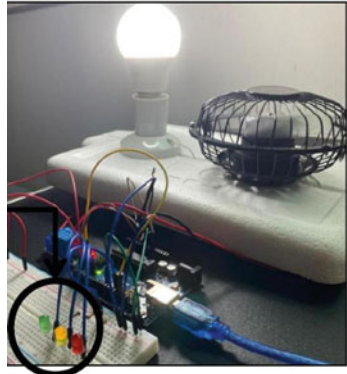
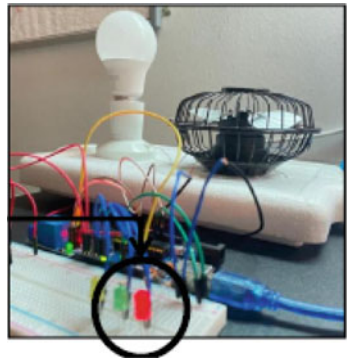


Fig. 18 Red LED light triggering



EEG features from Fp1 and Fp2 sensor position are employed in the Arduino code as shown in Fig. 18.

4 Conclusion

This study is successfully shown the effect of EEG sensor position in providing suitable EEG features in controlling device by analyzing the public EEG datasets that related to Brain-Computer Interface (BCI). From the investigation of the 6 position of EEG sensors, the results of the study show that F7, F8, P3 & P4 sensor position can provide suitable EEG features in term of mean, standard deviation and spectral centroids to control device with the 83% classification accuracy. For future work, more EEG sensors position are required to be analyzed in order to find out which sensor position can provide good EEG features to be employed in controlling device.

Acknowledgements The authors would like to thank Universiti Malaysia Pahang for laboratory facilities as well as financial support under internal research grant, RDU210318.

References

1. Rashid M et al (2020) Analysis of EEG features for brain computer interface application. In: ECCE2019. LNEE, vol 632, pp 529–540. Springer, Singapore. https://doi.org/10.1007/978-981-15-2317-5_45
2. Gurumurthy S, Mahit VS, Ghosh R (2013) Analysis and simulation of brain signal data by EEG signal processing technique using MATLAB. *Int J Technol* 5(3):2771–2776
3. Boostani R, Graitmann B, Moradi MH, Pfurtscheller G (2007) A comparison approach toward finding the best feature and classifier in cue-based BCI. *Med Biol Eng Comput* 45:403–412
4. Alotaiby T, El-Samie FEA, Alshebeili SA, Ahmad I (2015) A review of channel selection algorithms for EEG signal processing. *Eurasip J Adv Signal Process* 66:1–21
5. Birbaumer N, Murguialday AR, Cohen L (2008) Brain computer interface in paralysis. *Curr Opin Neurol* 21(6):634–638
6. Bacy-y-Rita P, Kerckel SW (2003) Sensory substitution and the human-machine interface. *Trends Cognit Sci* 7(12):541–546
7. Liu NH, Chiang CY, Chu HC (2013) Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensor* 13(8):10273–10286
8. Rashid M, Sulaiman N, Mustafa M, Khatun S, Bari BS (2018) The classification of EEG signals using different machine learning techniques for BCI application. In: *International Conference on Robot Intelligence Technology and Applications (RITA)*, pp 207–221
9. Vega-Escobar L, Castro-Ospina AE, Munoz LD (2015) Feature extraction schemes for BCI system. In: *20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*, pp 1–6
10. Lazarou I, Nikolopoulos S, Petranonakis PC, Kompatsiaris I, Tsolaki, M (2018) EEG-based brain computer interface for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century. *Front Hum Neurosci* 12(14):1–18
11. Sulaiman N, Al-Fakih NMMA, Rashid M, Jadin MS, Mustafa M, Samsuri F (2021) Offline EEG-based DC motor control for wheelchair application. In: *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019*. LNEE, vol 666. Springer, Singapore. https://doi.org/10.1007/978-981-15-5281-6_69
12. Leon M, Orellana D, Chuquimarca L, Acaro X (2019) Study of feature extraction method for BCI applications. *Adv Intell Syst Comput (AISC)* 1067:13–23
13. Nikolas M (2020) EEG Emotion Classification, Kaggle EEG Datasets. <https://www.kaggle.com/sannikolas/eeeg-dataset>. Accessed January 2021

Modeling and Analysis of Omnidirectional Wheeled Vehicles Using Velocity-Based Impedance Control



Norsharimie Mat Adam and Addie Irawan

Abstract This paper presents the velocity-based impedance control that would account for the inertia forces acting on the omnidirectional wheeled vehicle during cornering motions. As favorable omni-vehicle, omnidirectional mecanum wheeled vehicle (OMWV) was selected as a platform in this study. Concerning the problem statements in the dynamic analyses, the control design has considered the difference in vehicle forces because the vehicle's interaction forces were indirectly controlled by the vehicle's velocities. The axial velocities control of the OMWV, vertical and horizontal axial motions on cornering periods were highlighted in this research. The simulation results show that with velocity inputs, the different forces on the OMWV axial motion of the vehicle could be reduced. Furthermore, the reduction in vehicle velocity influenced the overall kinetic energy of the system, which reduced the inertia effect.

Keywords Cornering stability · Omnidirectional wheeled vehicle · Motion control · Inertia control · Compliance · And impedance control

1 Introduction

An omnidirectional mecanum wheeled vehicle (OMWV) is equipped with a fully actuated mechanism, which necessitates a different control perspective than a steering wheeled vehicle equipped with an underactuated mechanism. Several studies have been conducted in order to determine a better control for motion planning of the OMWV in order to increase the capability of this type of omnidirectional vehicle.

N. M. Adam (✉) · A. Irawan

Robotics, Intelligent and Control Engineering (RISC) Research Group, Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: sharimie.adam@gmail.com

A. Irawan

e-mail: addieirawan@ump.edu.my

A. Irawan

Center for Automotive Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

According to Chang et al., the speed of each omniwheel can be mathematically defined for control system design based on its orientation angle and direction [1]. Furthermore, according to Adamov et al., the contribution of each motor was calculated by multiplying the velocities by the cosine angle at the desired direction projected on each omni-wheel driven direction [2]. Oo et al., on the other hand, stated that omniwheel selection contributed to the increased flexibility of the wheeled vehicle when the vehicle platforms could move freely in a two-dimensional (2D) space rather than acting like a conventional vehicle [3]. Numerous efforts and research have been made to improve the stability of a wheel vehicle when confronted with the dynamic motion of path planning in a variety of critical situations such as cornering and harsh maneuvers [4, 5]. The inertia of the vehicle is an important element and parameter that must be considered in vehicle dynamics and control system design, but it is typically difficult to control [6].

The inertia may be caused due to the overdriven, slip of the OMWV, over force, sharp turn in cornering, and abrupt stop of OMWV. Realizing the smoothest path motion is the problem that should be resolved in the issue of trajectory planning [7], specifically on cornering tracks. As a holonomic mechanical system subjected to holonomic constraints, OMWV with mecanum wheels tends to change its body orientation, including small spinning during breaking at the cornering track. As reported in [8], most of the efforts emphasized better path planning for the vehicle with different control techniques focusing on system accuracy. The nonlinear friction of the holonomic vehicle is practically unbounded with the different working environments. These controls are inactive, and tracking performance gets insufficient when the nonlinear friction has a change. The change is out of the permitted scope. Thus, the kinodynamic problem is occurred in which needs a solution for both kinematics and dynamics to obtain optimal control input from the current state [9].

2 Velocity Input Shaping Using Impedance Control on OMWV

2.1 Dynamic Model of the Omnidirectional Wheeled Vehicle Platform

The dynamic analysis of the OMWV with mecanum wheels is determined by focusing on the motion of the mecanum wheels. Figure 1 depicts the coordination of this vehicle kinematics chain and the mecanum wheel coordination. In terms of the dynamic model, the resulting forces and torques on the vehicle platform are shown in Fig. 2. The input torques acting on each wheel are shown in Fig. 2(a) and can be calculated using the expression in Eq. 1:

$$\tau_i = \tau_{c_i} - \tau_{f_i} \quad (1)$$

Fig. 1 Top view of OMWV with mecanum wheel overall system coordination

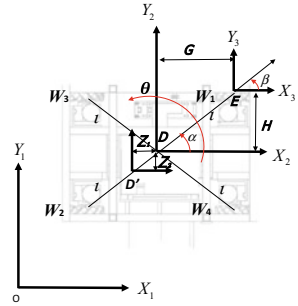
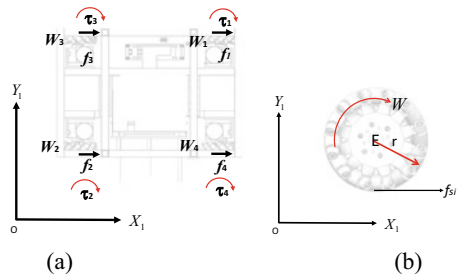


Fig. 2 Dynamic motion frame of OMWV with mecanum wheel: **a** General dynamic motion and **b** Static friction force in the direction of wheel rotation



The torque for each wheel was developed from the difference of the current torque of a wheel (τ_{c_i}) and friction torque (τ_{f_i}). The current torques and friction torques can be calculated using Eqs. 2 and 3.

$$\tau_{c_i} = m_{w_i} a_{oi} r \tag{2}$$

$$\tau_{f_i} = f_i r \tag{3}$$

where m_{w_i} is the mass of the wheel, a_{oi} is the acceleration of the current rotational of the i -wheel, and r is the radius of the wheel shown in Fig. 2(b). The torque of i -wheel of τ_{f_i} is calculated from the forces acting on each mecanum wheel (f_i), which is considered as the static friction force (f_{s_i}), using Eq. 4:

$$f_{s_i} = (\mu_k + \mu_s)(m_b + 4m_w)g \tag{4}$$

In this equation, m_b the platform's mass with g the acceleration due to gravity, μ_k is the coefficient of kinetic friction and μ_s is the coefficient of static friction [10]. In general, the friction is classified as static friction.

The wheel-road contact forces are critical factors since they present the unique interaction between the vehicle and the surface that it is moving on. Therefore, f_i in the direction of the wheel rotation must be considered [11], and the frictional force for i -wheel can be expressed as in Eq. 5:

$$f_i = f_{s_i} \text{sgn}(W_{o_i}) \tag{5}$$

Then, the dynamical model can be expressed using Euler–Lagrange as in Eq. 6:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) - \left(\frac{\partial L}{\partial q} \right) = F_i \tag{6}$$

where q is the generalized coordinated vector of the OMWV and can be defined as in Eq. 7:

$$q = [q_1 \ q_2 \ q_3]^T = [X_1 \ Y_1 \ \theta]^T \tag{7}$$

where F_i is the external generalized force motion and L is the Lagrangian function which is defined by Eq. 8, where K is the total kinetic energy, including those of the platform and the four mecanum wheels, and P is the total of the potential energy of the vehicle given in Eq. 9.

$$L = K - P \tag{8}$$

$$\begin{aligned} K &= K_1 + K_2 + \dots + K_i \\ P &= P_1 + P_2 + \dots + P_i \end{aligned} \tag{9}$$

As shown in the Eq. 9, K_i is the kinetic energy of the link, and P_i is its potential energy. Then, the kinetic energy K of the vehicle can be computed using Eq. 10:

$$K = 1/2 \left[m_b W_{D'}^T W_{D'} + I_b \dot{\theta}^2 + \sum_{i=1}^4 m_w (r W_{O_i})^2 + \sum_{i=1}^4 I_i W_{O_i}^2 \right] \tag{10}$$

In Eq. 10, I_b is the moment of the inertia of the platform, I_i is the moment of inertia of i -wheel ($i = 1 \dots 4$) about its central axis, and r is the radius of each mecanum wheel. Moreover, the potential energy is zero ($P = 0$) [12] since the vehicle is assumed to be moving in a plane, hence Euler–Lagrange equations can be modelled in terms of kinetic energy. The K expression in Eq. 10 is extracted in detail and can be seen in the previous paper [13]. After substitutions and computations, and from the Lagrangian of OMWV with mecanum wheel system, the Cartesian of forces error acting on the vehicle can be expressed in Eq. 11 where $e_f = [e_{fx} \ e_{fy} \ e_{f\theta}]$ refers to the center of mass (CoM) of the vehicle as shown in Fig. 2(a).

$$\begin{aligned} e_{fx} &= [\tau_1 - r \text{sgn}(\dot{W}_{o_1}) f_1] \left[-1/r (\cos \theta - \sin \theta) \right] \\ &+ [\tau_2 - r \text{sgn}(\dot{W}_{o_2}) f_2] \left[-1/r (\cos \theta + \sin \theta) \right] \\ &+ [\tau_3 - r \text{sgn}(\dot{W}_{o_3}) f_3] \left[1/r (\cos \theta - \sin \theta) \right] \end{aligned}$$

$$\begin{aligned}
 & + [\tau_4 - r \operatorname{sgn}(\dot{W}_{o_4}) f_4] \left[\frac{1}{r} (\cos \theta + \sin \theta) \right] \\
 e_{f_Y} = & [\tau_1 - r \operatorname{sgn}(\dot{W}_{o_1}) f_1] \left[-\frac{1}{r} (\sin \theta + \cos \theta) \right] \\
 & + [\tau_2 - r \operatorname{sgn}(\dot{W}_{o_2}) f_2] \left[-\frac{1}{r} (\sin \theta - \cos \theta) \right] \\
 & + [\tau_3 - r \operatorname{sgn}(\dot{W}_{o_3}) f_3] \left[\frac{1}{r} (\sin \theta + \cos \theta) \right] \\
 & + [\tau_4 - r \operatorname{sgn}(\dot{W}_{o_4}) f_4] \left[\frac{1}{r} (\sin \theta - \cos \theta) \right] \\
 e_{f_\theta} = & [\tau_1 + \tau_2 + \tau_3 + \tau_4] \left[-\frac{\sqrt{2}}{r} l \sin(\pi/4 - \alpha) \right] \\
 & + [\operatorname{sgn} W_{o_1}(f_1) + \operatorname{sgn} W_{o_2}(f_2) + \operatorname{sgn} W_{o_3}(f_3) \\
 & + \operatorname{sgn} W_{o_4}(f_4) + 1] \left[\sqrt{2} l \sin(\pi/4 - \alpha) \right]
 \end{aligned} \tag{11}$$

2.2 Velocity Input Shaping Using Impedance Control

Concerning the dynamic model discussed in Sect. 2.1, the velocity state is the lowest for the wheeled type of kinematics system. Therefore, the velocity-based impedance control was derived to cater for the inertia issue in the cornering period of OMWV system. In designing impedance control, the scenario on how reaction forces generated in association with its environment is emphasized. In this study, the inertia force and force error as derived in Eq. 11 were used as reaction forces translation of developed torques on each mecanum wheel, and the velocity of the OMWV is expected to shape with the changes of the virtual suspension system state of control law [14]. The desired impedance model of the OMWV system with inertia as the environment was determined, and the Cartesian impedance control can be derived and expressed in Eq. 12.

$$e_f = M_f \Delta \ddot{q} + D_f \Delta \dot{q} + K_f \Delta q \tag{12}$$

$q \in R^3$ describes the OMWV axes states. The difference between forces before and after the inertia is indicated as interaction forces (e_f) input for impedance equation for the OMWV system, in which, D_f and K_f are damped and stiffed in diagonal form for each X , Y , and θ of the vehicle system as expressed in Eq. 13.

$$\begin{aligned}
 D_f &= \operatorname{diag}(D_X \ D_Y \ D_\theta) \\
 K_f &= \operatorname{diag}(K_X \ K_Y \ K_\theta)
 \end{aligned} \tag{13}$$

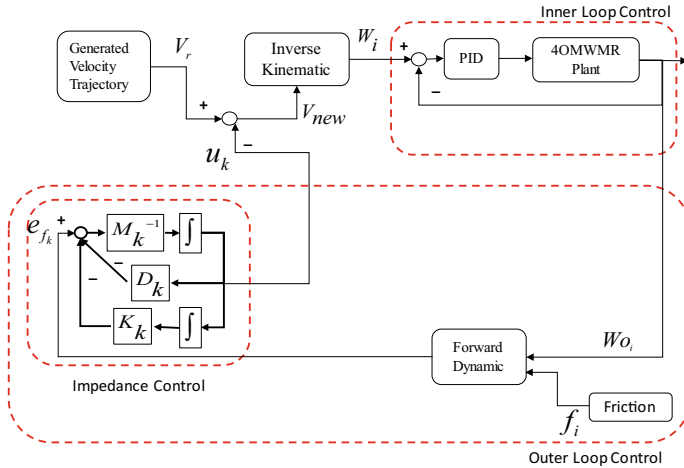


Fig. 3 The overall control architecture of OMWV with proposed velocity-based impedance control

On the other hand, M_f are diagonal of mass of the vehicle for each axial with a diagonal of 200 kg (without load) represented as M_k with $k = \{X, Y, \theta\}$ as shown in Fig. 3. To control the inertia of the OMWV with mecanum wheel via inertia forces, the relationship between D_k , K_k , and M_k need to be determined with $e_f \rightarrow 0$ and Newton’s Second Law are applied as in Eq. 14 [15] as follows;

$$\Delta\ddot{q} + \left(\frac{D_k}{M_k}\right)\Delta\dot{q} + \left(\frac{K_k}{M_k}\right)\Delta q = 0 \tag{14}$$

Thus, the natural frequency (ω_0) and damping ratio (ζ_f) of the impedance model can be expressed as in Eq. 15 as follows.

$$\omega_0 = \sqrt{\frac{K_k}{M_k}}, \zeta_f = \frac{D_k}{2\sqrt{M_k K_k}} \tag{15}$$

Therefore, the relationship of D_k with K_k , and M_k are as Eq. 16 by assuming the impedance is in underdamped mode.

$$D_k = \zeta_k 2\sqrt{K_k M_k} \tag{16}$$

As a result, the control law for velocity shaping for each vehicle’s axes can be formulated as in Eq. 17. By integrating Eqs. 15 to 12 and using the first order of the Δq as a control input as follows.

$$u_k(q) = \Delta\dot{q} = (K_k M_k)^{\frac{1}{2}}(e_{f_k} - M_k \Delta\ddot{q} - K_k \Delta q) \tag{17}$$

As previously mentioned, the cornering period is the case of this study. Thus, velocity shaping is implemented. On the other hand, $V_R \in^n$ are the reference inputs of velocity from the trajectory generation. The overall model of the control structure of OMWV with a mecanum wheel with the proposed velocity-based impedance control on Cartesian velocity input is shown in Fig. 3, where wheel speed control is considered the inner-loop control section and impedance control as the outer-loop control section.

3 Results and Analysis

3.1 Kinematic Analysis of Omnidirectional Mecanum Wheeled Vehicle

The simulation and analyses were done using the designed discrete trajectory waypoint for the X - Y position of the OMWV as input to the simulation model. The trajectory has been taken from the local frame coordinate data (X - Y position) to analyze the performance of the OMWV during the cornering period. Noted that, the crucial terrain is the asphalt (concrete) environment. Therefore, the simulation was done in three different tuning styles for parameters K_x and K_y in three different simulation sessions T_1 , T_2 , and T_3 ; shown in Table 1 for the asphalt concrete environment. In addition to the first simulation, the value of each K were assumed and the performance of ζ for each axis showed differently in starting from Fig. 5 to Fig. 6.

As shown in Fig. 4, the motion all turning motion WithImp shows following the reference of trajectory input as compare to the NoImp. The measured vehicle trajectories output shows very minimum tracking error in cornering period for with WithImp makes OMWV having a minimum overdriven situation. For the case of NoImp performances, the average trajectory motion output increased by approximately 1.2 m from the expected input trajectory (Full trajectory) as compare with WithImp that can reduced the trajectory tracking error by approximately 0.7 m. The results are significantly related to velocity and kinetic energy with reference to the performance in Fig. 5 and 6 respectively.

According to the results in Figs. 5(a) and 5(b), the result of ζ affected the shape of the velocity input of the OMWV. The result of the velocity input on the OMWV

Table 1 Type of styles stiffness and damping ratio tuning for each session of simulation

Simulation session	Tuning stiffness (Nm^{-1}) and damping ratio (ζ)			
	K_x	ζ_x	K_y	ζ_y
T_1	1	1	1	1
T_2	2	9	6	4
T_3	2	15	9	10

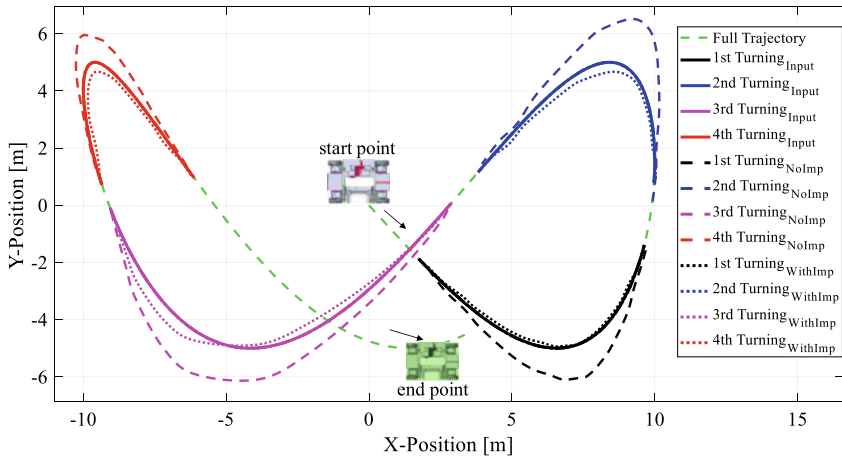
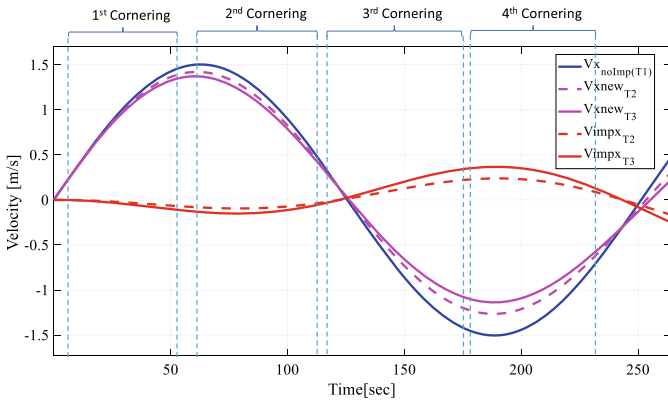
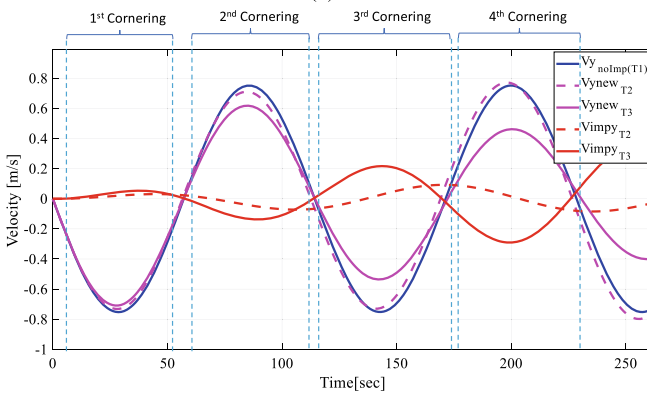


Fig. 4 Sample of the path planning input versus path planning output for OMWV X–Y motion without impedance (NoImp) and impedance control (WithImp)

is shown in Fig. 5(a). The horizontal of the robot system (X -axis) velocity decreased approximately by 0.02 m/s when K_X increased for T_2 and 0.15 m/s for T_3 . Similarly, the situation happened to the Y -axis motion, in which the velocity was reduced by 0.08 m/s in T_2 and 0.17 m/s in T_3 as shown in Fig. 5(b). As depicted in Figs. 5, the OMWV took the axle of weight concentration as a rotation centre and began rotating around it. It is clear that the OMWV displayed speed and rotational behaviour as a result of weight concentration. If the mass is shifted near the frontal axle, the OMWV advances, and vice versa. The speed of motion and rotation centre will change if the centre of mass is moved. There is an interaction between the location of the centre of mass and the direction of the driving force. This interaction will determine the translation’s direction and the rate at which the curvature centre changes. As previously stated, if no driving moment is applied to the wheel, the wheel will not rotate, but the roller in contact with the surface will slip. The diagonal motion (forward-right direction) was chosen to demonstrate the effect of shifting the centre of mass on the speed of OMWV. As depicted in Fig. 6, the energy reduced to approximately 18% for T_2 and 26% for T_3 . These results show that the higher the speed, the higher the inertia force, the bigger the energy used for a vehicle to interact between the system and environment. This situation can be compared with the sample of Y -axis velocity performance of the OMWV in the Fig. 5(b) respectively.



(a)



(b)

Fig. 5 Sample of velocity input performance **a** velocity input on X-axis OMWV system at cornering period **b** velocity input on Y-axis OMWV system at cornering period

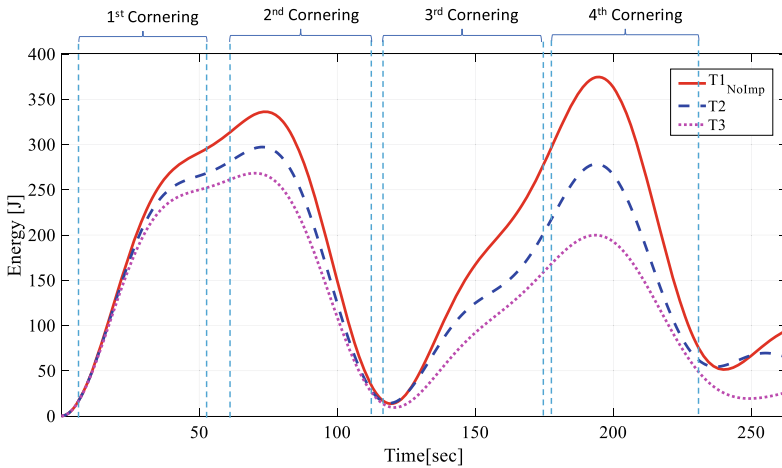


Fig. 6 Sample of the kinetic energy of OMWV system at cornering period

4 Conclusion

Inertia control through velocity-based impedance shaping is presented, showing the velocity input shaping each axis motion with force error reduction. In terms of overall performance, the inertia caused by friction forces during the cornering period was reduced to approximately 45% for the axes of Cartesian motion by tuning the stiffness and damping ratio values. The proposed control reduced vehicle speed to make the vehicle slower during the cornering session, which correlated to kinetic energy reducing. The proposed impedance control parameter has the potential to be optimised using a learning system, which will be one of the future works in this research, rather than proceeding with actual system implementation.

Acknowledgements The authors would like to thank the Universiti Malaysia Pahang for providing financial support under the Postgraduate Research Grant Scheme (Grant No. PGRS200348) and laboratory facilities.

References

1. Chang PIT et al (2020) Signal-based and model-based wheel fault detection of omni-directional vehicle with mecanum wheel. In: 2020 International Automatic Control Conference (CACS)
2. Adamov BI, Saypulaev GR (2020) A study of the dynamics of an omnidirectional platform, taking into account the design of mecanum wheels and multicomponent contact friction. In: 2020 International Conference Nonlinearity, Information and Robotics (NIR)
3. Oo AS et al (2019) Experiment on real-time image processing in the controlling of mecanum wheel robotic car. In: 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)

4. Seo B (2019) Path following control for military automobiles using dynamic inversion method. In: 2019 International Conference on Information and Communication Technology Convergence (ICTC)
5. Cheng S et al (2020) Robust LMI-based H-infinite controller integrating AFS and DYC of autonomous vehicles with parametric uncertainties. *IEEE Trans Syst Man Cybern Syst* 2020: 1–10
6. Yu Z, Huang X, Wang J (2015) A least-squares regression based method for vehicle yaw moment of inertia estimation. In: 2015 American Control Conference (ACC), 2015
7. Katrakazas C et al (2015) Real-time motion planning methods for autonomous on-road driving: state-of-the-art and future research directions. *Transp Res Part C Emerg Technol* 60:416–442
8. Faudzi FK, Irawan A (2020) Steering angle control of rack steering vehicle using antiwindup-PI-control. *Int J Electr Eng Appl Sci (IJEEAS)* 3(1):61–66
9. Moon CB, Chung W (2015) Kinodynamic planner dual-tree RRT (DT-RRT) for two-wheeled mobile robots using the rapidly exploring random tree. *IEEE Trans Ind Electron* 62(2):1080–1090
10. Ferriere L, Raucant, B (1998) ROLLMOBS, a new universal wheel concept. In: Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No. 98CH36146)
11. Vlantis P et al (2016) Fault tolerant control for omni-directional mobile platforms with 4 mecanum wheels. In: 2016 IEEE International Conference on Robotics and Automation (ICRA)
12. Cox JJ, Lozano-Perez T, Wilfong GT (2012) *Autonomous Robot Vehicles*. Springer, New York. <https://doi.org/10.1007/978-1-4613-8997-2>
13. Adam NM et al (2018) Dynamic modeling and analysis of omnidirectional wheeled robot: turning motion analysis. *J. Telecommun. Electron Comput Eng (JTEC)* 10(1–3):103–108
14. Hogan N (1984) Impedance control: an approach to manipulation. In: American Control Conference, 1984
15. Alam MM, Irawan A, Yin TY (2015) Buoyancy effect control in multi legged robot locomotion on seabed using integrated impedance-fuzzy logic approach. *Indian J Geo-marine Sci* 44(12):1937–1945

An Observation of Different Clustering Algorithms and Clustering Evaluation Criteria for a Feature Selection Based on Linear Discriminant Analysis



K. H. Tie, A. Senawi, and Z. L. Chuan

Abstract Linear discriminant analysis (LDA) is a very popular method for dimensionality reduction in machine learning. Yet, the LDA cannot be implemented directly on unsupervised data as it requires the presence of class labels to train the algorithm. Thus, a clustering algorithm is needed to predict the class labels before the LDA can be utilized. However, different clustering algorithms have different parameters that need to be specified. The objective of this paper is to investigate how the parameters behave with a measurement criterion for feature selection, that is, the total error reduction ratio (TERR). The k -means and the Gaussian mixture distribution were adopted as the clustering algorithms and each algorithm was tested on four datasets with four distinct clustering evaluation criteria: Calinski-Harabasz, Davies-Bouldin, Gap and Silhouette. Overall, the k -means outperforms the Gaussian mixture distribution in selecting smaller feature subsets. It was found that if a certain threshold value of the TERR is set and the k -means algorithm is applied, the Calinski-Harabasz, Davies-Bouldin, and Silhouette criteria yield the same number of selected features, less than the feature subset size given by the Gap criterion. When the Gaussian mixture distribution algorithm is adopted, none of the criteria can consistently select features with the least number. The higher the TERR threshold value is set, the more the feature subset size will be, regardless of the type of clustering algorithm and the clustering evaluation criterion are used. These results are essential for future work direction in designing a robust unsupervised feature selection based on LDA.

Keywords Unsupervised feature selection · Linear discriminant analysis · Clustering algorithm · Clustering evaluation criterion

K. H. Tie · A. Senawi (✉) · Z. L. Chuan

Centre for Mathematical Sciences, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia
e-mail: azlyna@ump.edu.my

Z. L. Chuan

e-mail: chuanzl@ump.edu.my

1 Introduction

Data with large number of features or variables are known as high dimensional data. The number of features of high dimensional data can be two, tens, hundreds, thousands or even up to millions. The larger the number of features means the higher the dimension of the data. Data of high dimensionality are often associated with the problem of high complexities in modelling. In many cases, some of the features are highly correlated and redundant [11]. Therefore, a pre-processing step that transforms high dimensionality data into lower dimensionality data is necessary and thereby allow significant features to be identified.

Lots of work were devoted on dimensionality reduction using mutual information stated in [8] and [22], correlation-based criterion stated in [5] and [21], principal component analysis (PCA) stated in [1] and [23] and linear discriminant analysis (LDA) stated in [2] and [14]. The LDA-based method is one of the popular approaches for feature selection. The focus of LDA is to maximize the separation of multiple classes. It has high performance in classifying unknown dataset and works by finding the discriminating function that gives clear separation of the data samples.

LDA has shortcomings when the data is unsupervised since it was specifically designed for supervised data. In particular, the number of class labels of the data is required for the LDA to be utilized. Nevertheless, it still can be adopted for unsupervised data by performing some clustering onto the data so that the number of class labels can be predicted in advance [24].

It is worth to note that a combination of LDA and k -means clustering algorithm is believed to offer higher classification accuracy compared to a combination of PCA and k -means or just k -means alone [6]. In addition, a combination of LDA with Gaussian mixture clustering is expected to perform well in nonparametric regression [9].

This paper aims to present an analysis of how the number of selected features behave using LDA-based feature selection under different clustering algorithms and clustering evaluation criteria. Note that the feature selection method being used is an unsupervised approach. The fewer the number of features selected, the better the clustering algorithm and the clustering evaluation criterion.

The next section of this paper discusses the clustering algorithms and clustering evaluation criteria that were used to group the samples of the data sets before the LDA-based feature selection can be carried out. The discussions are given in Sect. 2.1 and Sect. 2.2, respectively. The feature selection method applied for the observation is explained in brief in Sect. 2.3, while Sect. 3 describes how the experiment was performed. Section 4 is reserved for presenting the results and discussing the experimental findings. Finally, Sect. 5 delivers some concluding remarks of the study.

2 Related Works

2.1 Clustering Algorithms

Clustering is a process of creating a number of clusters by dividing the data based on their characteristics [19]. Clustering algorithms are used to solve classification problems involving data that do not have target or dependent variables. They will put together the data samples which are similar to each other in the same cluster while separating them as different as possible from other clusters. In this paper, two clustering algorithms are considered: k -means and Gaussian mixture distribution. These two clustering algorithms were chosen for the analysis as they have distinct clustering approaches. The k -means is a hard-clustering method [12, 17]. On the other hand, Gaussian mixture distribution is a soft clustering method [10, 25]. Hard clustering methods assign the data points only to a cluster only while soft clustering methods may put data into different clusters.

k -means. The k -means is a clustering algorithm that makes centroid a base cluster and minimizes the sum of distances between the data samples and their respective cluster centroid. There are many ways of measuring the distance, but Euclidean distance is the most commonly used [20]. The k -means has been proven easy to use, consume less memory and has high computation efficiency when compared to other clustering algorithms [18].

Gaussian Mixture Distribution. Gaussian mixture distribution is a clustering algorithm based on the superposition of multiple Gaussian distributions which is also known as Gaussian mixture model (GMM). The advantage of using the Gaussian mixture distribution is that it can create a model-based framework where the number of clusters and contribution of each feature in the clustering process can be shown [16].

2.2 Clustering Evaluation Criteria

Clustering evaluation criterion is a core component in finding different groups in data. Basically, an evaluation criterion is a distance value that quantifies which group a sample belongs. There are many criteria that can be used to determine the best division with an optimal number of clusters for a dataset. Four of them are discussed below: Calinski-Harabasz index, Davies-Bouldin index, Silhouette index and Gap statistic.

Calinski-Harabasz. The Calinski-Harabasz focuses on the ratio of the sum of cluster scattering within a cluster and cluster scattering between the clusters. Higher ratio indicates that the clusters are stable and have a good division between different clusters. This criterion is a good choice when quick results are desired as it is simple and thus can be computed within a short time.

Davies-Bouldin. The Davies-Bouldin evaluates the cluster by comparing the distance between clusters and the size of each cluster. The lower the index means, the better the separation between the clusters. It only needs a few controlled parameters to determine the number of clusters [26]. However, this criterion is limited to Euclidean distance only.

Silhouette. The Silhouette calculates the difference between the distance of points of different clusters and the adjacency of points within the cluster. The difference is then compared to the maximum distance of points of different clusters or the maximum adjacency of points within the cluster. The higher the index value means the better the separation of the clusters. This criterion has its benefit as it can determine whether the data lying in the correct cluster, incorrect cluster or overlapping cluster. It also has been proven to give higher performance efficiency in predicting the number of clusters than the Calinski-Harabasz and Davies-Bouldin [3]. However, it has greater complexity when compared to the Calinski-Harabasz and Davies-Bouldin.

Gap Statistic. The gap statistic compares the changes of dispersion within a cluster with the expected error for the same number of clusters is under a null reference distribution. However, the gap statistic is likely to fail if underestimation and overestimation of cluster occurred. Its performance was found to be relatively higher when combined with k -means compared to basic k -means [7].

2.3 MSOLS Feature Selection Guided by LDA

Multiple sequential orthogonal least squares (MSOLS) algorithm is a feature selection and ranking method to that rank significant features by using principal component analysis (PCA) to guide the selection [4]. In the method, the principal components are treated as dependent variables while the original features are taken as independent variables in the multiple regression models. The measurement criterion used to rank features importance in the MSOLS method is called as total error reduction ratio (TERR). In this paper, a similar approach proposed by [4] is adopted. However, instead of PCA, this paper utilizes the LDA as the dependent variables so as the characteristics of the LDA is taken into account to guide the feature selection and ranking. Detail steps to perform the LDA can be found in [15].

3 Methodology

Four public datasets were used to analyze how the number of selected features using the LDA-based feature selection described earlier behave under different clustering algorithms and clustering evaluation criteria. The clustering algorithms and clustering evaluation criteria discussed in Sect. 2.1 and 2.2, respectively, were applied in the

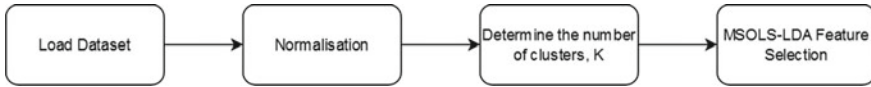


Fig. 1 Overall flowchart

experiment. In addition, three different threshold values of the TERR criterion were used to perform the observation: 80, 90 and 95%. One unsupervised dataset (Alate Adelges) and three supervised datasets (Iris, Wine, and Breast Cancer) with neglected class label were employed. Hence, 24 different tests were conducted for each dataset based on the two clustering algorithms, four evaluation criteria and three threshold values. Note that all features of the datasets are numeric values. The number of selected features was recorded for every test.

The Alate Adelges dataset consists of 19 features which were collected based on 40 winged aphids. The full dataset can be found from [13]. The Iris Data consists of 4 features, collected from 3 classes of Iris plants where 50 observations were recorded for each class. The Wine datasets consists of 13 features collected based on 178 observations from 3 types of wine. Meanwhile, the Breast Cancer dataset consists of 10 features which were collected based on 699 observations. The Iris, Wine, and Breast Cancer datasets can be obtained from the UCI Machine Learning Repository. The overall flowchart is shown in Fig. 1.

4 Results and Discussions

The results of the experiments on the four datasets are given in Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, Table 7 and Table 8. It can be observed that the *k*-means algorithm shows either lower or same average number of selected features than the Gaussian mixture distribution algorithm for all threshold values being considered except for the case 95% threshold value when tested with the Alate Adelges dataset. This proves that the feature selection method selects a smaller number of features when coupled with the *k*-means algorithm. The Calinski-Harabasz, Silhouette and

Table 1 Number of selected features of Iris dataset at different threshold values of TERR and clustering evaluation criteria under the *k*-means algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	2	2	2	2	2
90	2	2	3	2	2
95	2	2	3	2	2
Average	2	2	2	2	

Table 2 Number of selected features of Iris dataset at different threshold values of TERR and clustering evaluation criteria under the Gaussian mixture distribution algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	2	2	2	2	2
90	2	2	3	2	2.25
95	2	2	3	2	2.25
Average	2	2	2.67	2	

Table 3 Number of selected features of Alate Adelges dataset at different threshold values of TERR and clustering evaluation criteria under the *k*-means algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	3	3	3	3	3
90	6	6	7	6	6.25
95	13	13	14	13	13.25
Average	7.33	7.33	8	7.33	

Table 4 Number of selected features of Alate Adelges dataset at different threshold values of TERR and clustering evaluation criteria under the Gaussian mixture distribution algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	4	4	3	3	3.5
90	7	7	6	7	6.75
95	12	12	13	10	11.75
Average	7.67	7.67	7.33	6.67	

Table 5 Number of selected features of Wine dataset at different threshold values of TERR and clustering evaluation criteria under the *k*-means algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	7	7	8	7	7.25
90	10	10	10	10	10
95	11	11	11	11	11
Average	7.33	7.33	8	7.33	

Table 6 Number of selected features of Wine dataset at different threshold values of TERR and clustering evaluation criteria under the Gaussian mixture distribution algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	7	7	7	8	7.25
90	10	10	10	11	10.25
95	11	11	11	12	11.25
Average	7.67	7.67	7.33	6.67	

Table 7 Number of selected features of Breast Cancer dataset at different threshold values of TERR and clustering evaluation criteria under the *k*-means algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	5	5	4	5	4.75
90	8	8	7	8	7.75
95	9	9	8	9	8.75
Average	7.33	7.33	6.33	7.33	

Table 8 Number of selected features of Breast Cancer dataset at different threshold values of TERR and clustering evaluation criteria under the Gaussian mixture distribution algorithm

TERR threshold (%)	Clustering evaluation criterion				Average
	Calinski-Harabasz	Silhouette	Gap	Davies-Bouldin	
	Number of selected features				
80	5	5	5	5	5
90	8	8	8	8	8
95	9	9	9	9	9
Average	7.33	7.33	7.33	7.33	

Davies-Bouldin clustering evaluation criteria shows the same average number of selected features for each dataset under the *k*-means algorithm. This can be seen from Table 1, Table 3, Table 5, and Table 7. In the meantime, the Gap statistic fail to give a consistent pattern if compared to the other three criteria under the same clustering algorithm.

If the Gaussian mixture distribution is considered, only the Calinski-Harabasz and Silhouette evaluation criteria seem to show the same number of selected features when the same threshold values are considered through all four datasets. However, none of the criteria able to consistently select features with the least number when the Gaussian mixture distribution is applied.

It can also be observed that the number of selected features can increase significantly when the threshold value increases for all evaluation criteria, under both clustering algorithms as depicted in Table 3 and Table 4. Thus, the higher the TERR threshold value is set, the more the number of features may be selected, regardless which type clustering algorithm as well as the clustering evaluation criterion.

5 Conclusion

This paper analyses how different combinations of clustering algorithms and clustering evaluation criteria behave with an LDA-based feature selection. Through the experiment, it can be inferred that the feature selection method selects a smaller number of features when coupled with the k -means algorithm than the Gaussian mixture distribution algorithm. The Calinski-Harabasz, Silhouette, and Davies-Bouldin select the same number of features with the k -means algorithm, which is better than the Gap clustering criterion. As for the TERR threshold value, the higher the value is set, the more the number of features will be selected. Thus, one can anticipate that a higher TERR value will lead to a better classification result.

Acknowledgements The authors would like to thank the International Islamic University Malaysia (IIUM), Universiti Malaysia Pahang (UMP) and the Universiti Teknologi MARA (UiTM) for providing financial support under the IIUM-UMP-UiTM Sustainable Research Collaboration Grant 2020 (Vote Number: RDU200722).

References

1. Adegbola OA, Adeyemo IA, Semire FA, Popoola SI, Atayero AA (2020) A principal component analysis-based feature dimensionality reduction scheme for content-based image retrieval system. *Telkomnika* 18(4):1892–1896
2. Alharbi AS, Li Y, Xu Y (2017) Integrating LDA with clustering technique for relevance feature selection. In: Peng W, Alahakoon D, Li X (eds) *Advances in Artificial Intelligence: 30th Australasian Joint Conference*. Springer, Melbourne, pp 274–286
3. Baarsch J, Celebi ME (2012) Investigation of internal validity measures for k -means clustering. In: *Proceedings of the International MultiConference of Engineers and Computer Scientists*, pp 471–476. Newswood Limited, Hong Kong
4. Billings SA, Wei HL (2005) A multiple sequential orthogonal least squares algorithm for feature ranking and subset selection. ACSE Research Report (908). University of Sheffield
5. Chormunge S, Jena S (2018) Correlation based feature selection with clustering for high dimensional data. *J Electr Syst Inf Technol* 5(3): 542–549
6. Ding C, Li T (2007) Adaptive dimension reduction using discriminant analysis and K -means clustering. In: Ghahramani Z (ed) *ACM International Conference Proceeding Series*, vol 227. Association for Computing Machinery, New York, pp 521–528
7. El-Mandouh AM, Mahmoud HA, Abd-Elmegid LA, Haggag MH (2019) Optimized K -means clustering model based on gap statistic. *Int J Adv Comput Sci Appl (IJACSA)* 10(1):183–188
8. Gao W, Hu L, Zhang P (2020) Feature redundancy term variation for mutual information-based feature selection. *Appl Intell* 50(4):1272–1288

9. Hastie T, Tibshirani R (1996) Discriminant analysis by Gaussian mixtures. *J Roy Stat Soc Ser B (Methodol)* 58(1):155–176
10. He C, Fu H, Guo C, Luk W, Yang G (2017) A fully-pipelined hardware design for Gaussian mixture models. *IEEE Trans Comput* 66(11):1837–1850
11. Houari R, Bounceur A, Kechadi MT, Tari AK, Euler R (2016) Dimensionality reduction in data mining. *Expert Syst Appl Int J* 64(C): 247–260
12. Kamper H, Livescu K, Goldwater S (2017) An embedded segmental K -means model for unsupervised segmentation and clustering of speech. In: *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp 719–726
13. Krzanowski WJ (2018) Attribute selection in correspondence analysis of incidence matrices. *J Roy Stat Soc: Ser C (Appl Stat)* 42(3):529–541
14. Kumar, BS, Ravi V (2017) LDA based feature selection for document clustering. In: *Proceedings of the 10th Annual ACM India Compute Conference*, pp. 125–130. Association for Computing Machinery, New York
15. Lu J, Plataniotis KN, Venetsanopoulos AN (2003) Face recognition using LDA-based algorithms. *IEEE Trans Neural Netw* 14(1):195–200
16. Maugis C, Celeux G, Martin-Magniette ML (2009) Variable selection for clustering with Gaussian mixture models. *Biometrics* 65(3):701–709
17. Mohd MRS, Herman SH, Sharif Z (2017) Application of K -Means clustering in hot spot detection for thermal infrared images. In: *IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp 107–110
18. Morissette L, Chartier S (2013) The k -means clustering technique: general considerations and implementation in Mathematica. *Tutor Quant Methods Psychol* 9(1)
19. Nazari Z, Kang D, Asharif MR, Sung Y, Ogawa S (2016) A new hierarchical clustering algorithm. In: *ICIIBMS 2015–International Conference on Intelligent Informatics and Biomedical Sciences*, pp 148–152
20. Duda O, Peter E, Hart DGS (eds) (2000) *Pattern Classification*. 2nd edn. Wiley, United States
21. Senawi A, Wei HL, Billings SA (2017) A new maximum relevance–minimum multicollinearity (MRmMC) method for feature selection and ranking. *Pattern Recognit.* 67: 47–61
22. Sharmin S, Shoyaib M, Ali AA, Khan MAH, Chae O (2019) Simultaneous feature selection and discretization based on mutual information. *Pattern Recogn* 91:162–174
23. Uddin, MP, Mamun, MA, Hossain, MA (2020) PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Techn Rev* 1–21
24. Ünü R, Xanthopoulos P (2019) Estimating the number of clusters in a dataset via consensus clustering. *Expert Syst Appl* 125:33–39
25. Vashishth V, Chhabra A (2019) GMMR: a Gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic IoT networks. *Comput Commun* 134:138–148
26. Xiao J, Lu J, Li X (2017) Davies bouldin index based hierarchical initialization K -means. *Intell Data Anal* 21(6):1327–1338

Recent Advances and Open Challenges in RFID Antenna Applications



Nazmus Sakib Khair, Nurhafizah Abu Talip Yusof,
Mohd Hisyam Mohd Ariff, Yasmin Abdul Wahab, and Bifta Sama Bari

Abstract Ultra-high-frequency (UHF) radio frequency identification (RFID) technology has been gaining significance with the progression of wireless communication systems and equipment having a wide variety of applications. This technology has been promoted as a low-cost, low-energy method of operation. A huge number of papers demonstrate the possibility of foreign materials or specific objects being tracked and monitored in food packaging, healthcare, agriculture, and environmental conditions. Maximum work focuses on area coverage, and significant information has been gathered to exhibit possibilities, but some open challenges and limitations have also been manifested in the previous research. Thus, further information is needed for a profound understanding of the RFID antenna method to make them dependable and applicable. From a system point of view, the challenges and state-of-the-art techniques of the UHF RFID antenna are comprehensively summarized and clearly highlighted in terms of sensing and communication. Oncoming aptitudes with recommended challenges to mitigate current limitations of RFID antenna design are also discussed.

Keywords UHF antenna · RFID · Wireless communication systems · Antenna design

1 Introduction

Radio frequency identification (RFID) is a microwave-based technology that is used for recognition using radio frequency electromagnetic waves to transfer data [1, 2]. RFID is a thriving technology throughout the world [3]. This technology is used in the identification and tracking of individual objects, including people, animals, or other

N. S. Khair · N. A. T. Yusof (✉) · M. H. M. Ariff · Y. A. Wahab · B. S. Bari
Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, 26600
Pekan, Pahang, Malaysia
e-mail: hafizahs@ump.edu.my

N. A. T. Yusof
Centre for Research in Advanced Fluid and Processes (Fluid Centre), Universiti Malaysia Pahang,
Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia

items [4–6]. Normally, three types of frequencies are used in RFID systems. Low-frequency (LF) RFID ranges from 125 to 134 kHz, the high-frequency (HF) RFID range is 13.56 MHz, and the ultra-high-frequency (UHF) RFID ranges from 860 to 960 MHz. Different countries use different variations of the UHF system. Using UHF systems, systems can cover a large area with a flexible reading range rather than LF and HF [7, 8]. UHF is normally used in radio services, Wi-Fi, Bluetooth, satellite communications, cell phones, television broadcasting, and other various applications [9]. Low frequencies are typically used in the military, radio broadcasting, radio navigation signals and other different sectors. High frequencies are usually used in military communications, ship-to-ship communications, in the aviation sector, radar, etc. LF tags usually have slower transfer rates than HF and UHF. LF can be used at a few centimeters' distance. The HF operation range varies from millimeters to meter, and its data rates are acceptable for different uses. UHF tags can be read from a long distance and have fast identification characteristics [10–12]. Due to its long range, fast transfer rates, and low cost, UHF has become popular for a variety of applications, including industrial applications such as tool tracking, IT asset tracking, and laundry management. The UHF antennas' receiving and sending signal range varies from 300 MHz to 3 GHz. Common UHF antenna designs include microstrip antennas, fan dipole antennas, Yagi antennas, UHF reflectors, planar antennas, and log periodic dipole antennas. UHF RFID antennas are beneficial because of their passive, wireless, low-profile, solid size, and particular application in infrastructures. However, there are some limitations to the existing antenna system in antenna performance, including antenna bandwidth, gain, directivity, and size. In this paper, we have discussed a concise idea about different types of UHF antennas. Moreover, numerous applications of UHF RFID antennas in different sectors have been presented here, which is the main focus of this study. In addition, possible solutions to the current issues with RFID antenna systems and their future directions have also been discussed.

The rest of this paper is organized as follows: Section 2 represents a concise description of different UHF RFID antenna designs. Section 3 discusses the application of different existing UHF RFID antennas in different sectors. Finally, possible solutions have been shown in Section 4, followed by a conclusion in Section 5. This gives a sense of where the study will go next.

2 Types of UHF Antenna

2.1 Microstrip Antenna

The microstrip antenna is a comparatively recent creation. It is allowed to implement a handy integration of an antenna with other circuits of a communication system on a common printed-circuit board (PCB) or a semiconductor chip. Li et al. [13] proposed a microstrip array antenna with a gain of 9.5 dBi in which the integrated circuit technique is used to fabricate the antenna to achieve good accuracy. Pratigya

M. et al. [14] proposed a series-fed microstrip array antenna whose gain is 15.5 dBi and reflection loss is about -10 dB. Both antennas used the array technique, but their goals were not the same. Microstrip antennas are popular because of their light weight, lower cost, lower profile, superior portability, ease of fabrication, and suitability for arrays and planar configuration. But it has some disadvantages too, such as lower gain and narrow bandwidth. The operational range of microstrip antennas is from 100 MHz to 100 GHz. However, they normally operate in the GHz range.

2.2 Reflector Antenna

The reflector antenna is popular in spacecraft systems because of its simple construction, light weight, and advanced manufacturing design. Moreover, Park et al. [15] proposed a single-beam parabolic reflector antenna for radar applications. The function of a reflector antenna is to reflect electromagnetic waves. Derry et al. [16] proposed an antenna with narrow-wall slots and a unique design of a V-shaped metal reflector with a bandwidth of 780 MHz and a VSWR of 1.5. The main objectives are to increase the gain and enhance the side lobe level. The antenna is made for X band applications. These two antennas can be differentiated by design, application, and gain. The major disadvantage is that the reflector requires symmetrizing to eliminate obstacles at the feed point. A reflector antenna can be made up of different types of reflectors, whose shapes can be hyperbolic, ellipsoid, parabolic, etc. The most common reflector antenna that is used in several applications is the parabolic type.

2.3 Yagi Antenna

The Yagi antenna, also known as the Yagi-Uda antenna or aerial, is a remarkably popular antenna where gain and directivity are vital. Singh et al. [17] proposed a compact Yagi antenna for near field UHF RFID systems. The Yagi antenna has a dipole as the main radiating component, and power is directly engaged from a feeder. P. Ngamjanyaporn et al. [18] present in their work a switch-beam circular array using a Yagi antenna for space communications, which has 7 main beam directions and the gain of the antenna is 7 dBi. A reflector is also added to the radiating component, which is about 5% longer than the radiating component, resulting in better gain and directivity than other normal dipole antennas. This kind of antenna is easy to fabricate and lower in cost. But it has some disadvantages too. For a higher gain antenna, the size becomes too long.

Table 1 Characteristics of different UHF RFID antennas.

Authors	Antenna type	Design technique	Antenna size (mm)	Gain (dBi)	Bandwidth
Li et al. [13]	A low-profile dual-polarized microstrip antenna array	n/a	120 × 112 × 5.3	9.5	200–300 MHz
Park et al. [15]	A single beam parabolic reflector antenna	n/a	n/a	n/a	24–77 GHz
Singh et al. [17]	A compact Yagi-Uda antenna	n/a	n/a	n/a	868 MHz
Hua et al. [19]	A high-gain planar dual reflector antenna	CST	122 × 78 × 3	16	32.7–37.6 GHz
Zengin et al. [21]	A planar log periodic antenna	CST	n/a	7.8	2–6 GHz

2.4 Planar Antenna

The planar antenna is a two-dimensional antenna because the active and parasitic components are set in one plane. The planar antenna can also be classified as a microstrip antenna. Hua et al. [19] proposed a high-gain planar dual reflector antenna using computer simulation technology (CST) with 16 dBi gain. A planar antenna is an ultra-wideband antenna, and it has lower cost advantages. Tian Ling Zhang et al. [20] proposed an oversized rectangular antenna for the Ka band which has a low profile structure and good impedance bandwidth and high efficiency. These two planar antennas have simple and low-profile structures. Owing to its low profile and size, it is perfect for wireless communications. However, some planar antennas have low gain, low radiation efficiency, and a narrow bandwidth.

2.5 Log Periodic Dipole Antenna

The log periodic antenna, or aerial, also known as LPDA, is an antenna that can provide directivity and gain at the same time over a wide band of frequencies. Zengin et al. [21] proposed a trapezoidal dipole array antenna with 7.8 dBi gain. The log periodic dipole antenna consists of several dipole components. It can be used in different frequency applications where the pattern and radiation of the antenna do not depend on frequency. Xin-Rong Li et al. [22] demonstrated a new high-gain printed log-periodic dipole antenna with a gain of 9.1–12.6 dBi based on substrate integrated waveguide methods. The wide bandwidth is the main advantage of the development of LPDA antennas. However, it has some disadvantages too. For example, the

antenna has lower gain compared to other antennas, and it cannot be used in omnidirectional applications. The characteristics of different UHF RFID antennas have been shown in Table 1.

3 Review of UHF Antenna Applications

3.1 Telecommunication Technology

In the telecommunication sector, the use of UHF RFID technology is increasing day by day for its cost-effective performance in tracking data-center assets and costly instruments. Mobile phone producers realize that RFID-enabled mobiles will be embraced by consumers to make payments at stores, buy tickets on mass transit, and manage other financial services. Hence, since RFID systems have a limited communication range, it is necessary to use several RFID readers and related devices in a network area to cover all tags. Due to the communication range limitation, configuring an RFID system in a large area is very significant and difficult. Zhao et al. [23] came up with a solution to deal with these problems by figuring out how to deal with interference and coverage issues while taking into account the cost. Reliability and performance issues have to be taken into account in telecommunication services as well [24]. This is due to poor monitoring of construction, professionals' activities and construction areas, which have been liable for different occupational insecurities. An experiment was done by Osunsanmi et al. [25] to make sure construction workers were safe, which used radio frequency identification (RFID) and mobile technology.

3.2 Biomedical Technology

Different types of antennas are used in biomedical technology to detect health diseases [26, 27]. Every year, millions of people are affected by different kinds of chronic diseases, and millions of them die. Various kinds of innovations have been invented for the purpose of the identification and treatment of different types of chronic diseases. Nowadays, different types of UHF miniature patch antennas are applied for various biomedical applications such as cancer detection, tumor detection, or stroke imaging. Besides, there have been approaches to developing respirational rate monitors recently in order to improve the system's health. Those approaches sprout with the awareness that patient safety is a serious matter in global public health concerns. Mistreatments in the healthcare industry are on the rise as a result of medical errors caused by misidentifications. RFID is implemented to develop safety, proper identification, efficiency, and reduce cost [28]. Najera et al. [29] tried to give two kinds of solutions in their study: one solution is for medical equipment tracking for real-time location and theft prevention; the other is for taking care of and monitoring the patient in the hospital. Recently, the use of RFID antennas in different ways is growing in a variety of biomedical fields.

3.3 Agricultural Technology

The advancement of UHF RFID and related innovations such as antennas, transponders, and other electronic devices provides enormous opportunities for agricultural research, development, and innovative thinking. Owing to its track-down capacity, UHF RFID is used for monitoring the food chain. UHF RFID is also used for soil water monitoring, irrigation, environmental monitoring, and other purposes. But it has some challenges in collecting data in a rough environment with dirt and high temperatures [30]. In cold climate countries, UHF RFID technology is very good at modernizing greenhouse management because it is low-cost and power-efficient. [31]. Wang et al. [32] presented the effect of a passive RFID sensor on the growth and different characteristics of a specific harvest. Deng et al. [33] proposed a technique in their work soil environment control system using RFID sensors and short-range (LoRa) wireless communication to know long-term and low-cost monitoring.


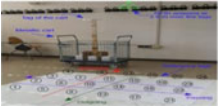

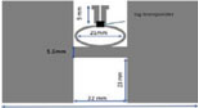
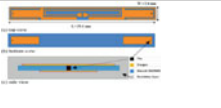
3.4 Livestock Farming Technology

RFID technology is becoming a very common method for the livestock farming industry for animal identification and tracking. The awareness that animals carry diseases and other environmental issues has motivated the research and development of individual identification and tracking systems for livestock. Moreover, the RFID system is cost-effective and reduce the incidence of livestock theft. Thus, RFID technology can be said to be very effective for monitoring and taking care of livestock. Basic RFID systems for animals consist of a reader and an identification tag. Transponders of different sizes, factors, and weights have been developed, together with ear tags, small capsules, and rumen boluses [34]. In animal husbandry, LF transponders are not very sensitive to radio wave trespassing and water. In addition, LF transponders will not be influenced by metal in the environment, thus making them a very good option for animal facilities [35]. For the purpose of animal behavior studies, UHF RFID antennas are currently being used by different experimentation groups. By implementing passive HF RFID with an operating frequency of 13.56 MHz, the feeding of piglets was successfully identified in a semi-commercial state [36]. Meanwhile, Maselyne et al. [37] proposed a HF RFID system to validate its suitability for pigs' feeding models at a round trough in the housing area. Thus, based on the research that has been reported, the use of UHF RFID technology can be said to be increasing very fast in present days.

3.5 UHF RFID Antenna Application in Other Technologies

Besides the above applications mentioned in sub-topics 3.1 to 3.4, UHF RFID technology is also used in some other areas, including flood detection systems, humidity identification, green house monitoring, Bluetooth connection, television broadcasting, radio services, etc. It has also been reported that the UHF RFID perfect

Table 2 Comparison of different UHF RFID antenna applications.

Author	Type of antenna project	Applications	Size (mm)	Performance	Prototype
Li et al. [1]	A flooding warning system based on an RFID tag array for energy facilities	To measure the performance of buried tags for flood detection.	n/a	n/a	
Narci- andi et al. [39]	A UHF-RFID gate control system based on a convolutional neural network	To classify transpallet actions at gates or check points.	n/a	n/a	
Sarkar et al. [40]	A dual frequency circularly polarized UHF RFID/WLAN circular patch antenna for RFID readers	To gain dual band operability and circular polarization at both respective frequencies.	84 × 84 × 1.6	Band-width: 915 MHz & 2.45 GHz. Gain: 0.58 dBi and 6.84 dBi	
Turki et al. [41]	Chemical vapor detects passive RFID tags	To detect the threshold of certain chemical vapors.	80 × 60	Band-width: EU band 865–868 MHz and US band (902–928) MHz	
Ngyen et al. [42]	Wireless interrogation of small animal phantoms with a miniature implanted UHF RFID tag	To identify small animals in different positions.	30.5 × 30.5	868 MHz	

for use in disaster-hit areas and sewerage systems. Rennane et al. [31] presented a self-dependent multi sensing passive UHF RFID tag for greenhouse monitoring. While Borgese et al. [38] proposed a chipless RFID sensor based on an artificial impedance surface, which result in the tag’s having a lower error chance when it is used as a threshold sensor. Table 2 shows a comparison of different characteristics of UHF RFID antenna applications.

4 Issues with Possible Solutions

UHF RFID has great advantages comparable to LF systems, such as a longer read range, faster data transfer rate, and programmability. However, there are some limitations associated with the antenna system. Firstly, the antenna performance can be

affected by the variation of the antenna substrate material. Besides, antennas with low gains and bandwidth will disrupt the whole antenna performance. This means that an antenna with a good value of gain and suitable bandwidth is highly desirable. Furthermore, the antenna radiation pattern can have an impact on how well the antenna works as a whole. This makes the antenna less directed.

To mitigate the above limitations, several techniques should be implemented in the design of an RFID antenna. Based on the trends and previous analysis of antenna design techniques, some observations are proposed, particularly for the efficient applications of the RFID antenna methods. Different flexible substrates can be used for antenna design according to the applications. The chipless RFID tag can be made on flexible substrates by printing techniques using conductive inks because it does not contain weighty chips but only a metal model as an antenna, lowering the fabrication cost significantly [43]. The printability of the tags on extensionable substrates is also desired to qualify the RFID tag as suitable for any surface [44]. Conductive ink plays a significant role in printable antennas. The manufacture of conductive inks and obtaining conductive models by using several sintering methods could be recommended [45]. The backscattered signal is subject to environmental multipath routes to the reader that cause both frequency and time-selective effects [46]. It is recommended to apply different techniques to make the system sensitive and perform well. Several techniques, including reflector layer and partial ground techniques, could be used to enhance antenna gain, bandwidth, and directivity [47, 48]. In addition, impedance matching could be recommended in antenna design to achieve the desired performance. To sum up the discussion, low fabrication costs, simple processing steps, less waste, the appropriate substrate material, and simple patterning techniques are advantages for cost-effective antenna manufacturing [49].

5 Conclusion

This paper presents an overview of the applications of UHF antenna systems based on RFID technology. The related issues have been summarized, and the challenges, reasons, and state-of-the-art have been presented. This study offers a comprehensive understanding of problems and instructions in the RFID area. This review paper also includes a brief discussion on antenna design techniques with the intention of giving an overall overview to the new researchers working on a wide range of RFID antenna systems and applications.

Acknowledgements The authors would like to thank the UMP-IIUM-UITM research collaboration grant RDU200742 and UMP internal grant PDU203218 for funding this project.

References

1. Li CH, Lao KW, Tam KW (2018) A flooding warning system based on RFID tag array for energy facility. In: RFID-TA 2018–2018 IEEE International Conference on RFID Technology & Application, pp 1–4. <https://doi.org/10.1109/RFID-TA.2018.8552767>.
2. El Matbouly H, Tedjini S, Zannas K, Duroc Y (2019) Chipless sensing system compliant with the standard radio frequency regulations. *IEEE J Radio Freq Identif* 3(2):83–90. <https://doi.org/10.1109/JRFID.2019.2909092>
3. Marindra AMJ, Tian GY (2019) Multiresonance chipless RFID sensor tag for metal defect characterization using principal component analysis. *IEEE Sens J* 19(18):8037–8046. <https://doi.org/10.1109/JSEN.2019.2917840>
4. Tariq N et al (2019) Orientation independent chipless RFID tag using novel trefoil resonators. *IEEE Access* 7:122398–122407. <https://doi.org/10.1109/ACCESS.2019.2937131>
5. Ben Abdallah E, Egels M, Bergeret E, Benzaim O, Pannier PA (2018) A comparative performances analysis of UHF RFID two crossed dipoles antenna based on FR4 substrate. *Microw Opt Technol Lett* 60(9):2217–2224. <https://doi.org/10.1002/mop.31322>
6. Huang H-F, Su Li: A Compact Dual-Polarized Chipless RFID Tag by Using Nested-Yahoo Search Results. https://search.yahoo.com/search/?hspart=__default&hsimp=__default&q=Huang. Accessed 04 Aug 2021
7. Hessel, EF, Saeys, W, Van Nuffel, A, Gallmann, E (2019) A review of passive radio frequency identification systems for animal monitoring in livestock facilities, vol 35, no 4, pp 579–591
8. Anam H, Habib, A, Jafri SI, Amin Y, Tenhunen H (2017) Directly printable frequency signa-tured chipless RFID tag for IoT applications. *Radioengineering* 26(1). <https://doi.org/10.13164/re.2017.0139>.
9. Chen ZN, Qing X, Chung HL (2009) A universal UHF RFID reader antenna. *IEEE Trans Microw Theory Tech* 57(1):1275–1282. <https://doi.org/10.1109/TMTT.2009.2017290>
10. Cho C, Choo H, Park I (2005) Broadband RFID tag antenna with quasi-isotropic radiation pattern. *Electron Lett* 41(20):1091–1092. <https://doi.org/10.1049/EL:20052337>
11. Abdulghafor R et al (2021) Recent advances in passive UHF-RFID tag antenna design for improved read range in product packaging applications: a comprehensive review. *IEEE Access* 9:63611–63635. <https://doi.org/10.1109/ACCESS.2021.3074339>
12. Ng WH, Lim EH, Bong FL, Chung BK (2019) E-shaped folded-patch antenna with multiple tuning parameters for on-metal UHF RFID tag. *IEEE Trans Antennas Propag* 67(1):56–64. <https://doi.org/10.1109/TAP.2018.2874795>
13. Li H, Kang L, Wei F, Cai YM, Yin YZ (2017) A low-profile dual-polarized microstrip antenna array for dual-mode OAM applications. *IEEE Antennas Wirel Propag Lett* 16:3022–3025. <https://doi.org/10.1109/LAWP.2017.2758520>
14. Mathur P, Arrawatia M (2020) High gain series fed planar microstrip antenna array using printed L-Probe feed. In: 2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting IEEECONF 2020–Proceedings, pp. 589–590. <https://doi.org/10.1109/IEEECONF35879.2020.9329574>
15. Park DH (2019) Design and analysis of single beam parabolic reflector antenna in LTCC for millimeter wave automotive radar. In: 2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems COMCAS 2019, pp 20–22. <https://doi.org/10.1109/COMCAS44984.2019.8958132>
16. Yusuf DP, Zulkifli FY, Rahardjo ET (2019) Design of narrow-wall slotted waveguide antenna with V-shaped metal reflector for XBand radar application. In: ISAP 2018–2018 International Symposium on Antennas and Propagation (ISAP), pp 18–19
17. Singh RK, Michel A, Nepa P, Salvatore A (2019) Design of a compact Yagi-Uda antenna for near field UHF RFID smart gloves. In: 2019 IEEE International Conference on RFID Technology & Application RFID-TA, pp 453–457. <https://doi.org/10.1109/RFID-TA.2019.8892028>

18. Ngamjanyaporn P, Kittiyapunya C, Krairiksh M (2017) A switch-beam circular array antenna using pattern reconfigurable Yagi-Uda antenna for space communications. In: 2017 International Symposium on Antennas Propagation, ISAP 2017, vol 2017-January, pp 1–2. <https://doi.org/10.1109/ISANP.2017.8228872>
19. Zong H, Zhang H, Lin S, Li H, Liu B, Wu Q (2017) A high-gain planar dual reflector antenna. In: ISAP 2016-International Symposium on Antennas Propagation, pp 746–747
20. Zhang T, Chen L, Guo C, Zhang Z (2017) Low profile oversized rectangular waveguide planar antenna fed with quasi-TEM wave in ka band. In: 2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL), vol 2017-November, pp 1690–1692. <https://doi.org/10.1109/PIERS-FALL.2017.8293406>
21. Zengin F (2019) The effects of the trapezoidal dipole array elements on planar log periodic antenna. In: In 2019 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC 2019), pp 333–336. <https://doi.org/10.1109/APWC.2019.8870439>
22. Li XR, Ye M, Chu QX (2016) Novel high gain printed log-periodic dipole antenna. In: In 2016 IEEE International Symposium on Antennas and Propagation (APSURSI 2016)–Proceedings, pp 1647–1648. <https://doi.org/10.1109/APS.2016.7696530>
23. Zhao C et al (2017) Decomposition-based multi-objective firefly algorithm for RFID network planning with uncertainty. *Appl Soft Comput J* 55:549–564. <https://doi.org/10.1016/j.asoc.2017.02.009>
24. De Barros Filho IE, Silva I, Costa DG, Viegas CMD, Ferrari P (2021) A reliability and performance GSPN-Based model for anti-collision RFID algorithms under noisy channels in industrial internet of things. *Comput Ind* 125:103381. <https://doi.org/10.1016/j.compind.2020.103381>
25. Osunsanmi TO, Oke AE, Aigbavboa CO (2019) Survey dataset on fusing RFID with mobile technology for efficient safety of construction professionals. *Data Br* 25:104290. <https://doi.org/10.1016/j.dib.2019.104290>
26. Bari BS et al (2020) Ultra wide band (UWB) based early breast cancer detection using artificial intelligence. In: Nasir ANK et al (eds) *InECCE2019*. LNEE, vol 632, pp 550–515. Springer, Singapore. https://doi.org/10.1007/978-981-15-2317-5_43
27. Markkandan S, Malarvizhi C, Raja L, Kalloor J, Karthi J, Atla R (2021) Highly compact sized circular microstrip patch antenna with partial ground for biomedical applications. *Mater Today Proc* (xxxx):12–14. <https://doi.org/10.1016/j.matpr.2021.04.480>
28. Haddara M, Staaby A (2018) RFID applications and adoptions in healthcare: a review on patient safety. *Procedia Comput Sci* 138:80–88. <https://doi.org/10.1016/j.procs.2018.10.012>
29. Najera P, Lopez J, Roman R (2011) Real-time location and inpatient care systems based on passive RFID. *J Netw Comput Appl* 34(3):980–989. <https://doi.org/10.1016/j.jnca.2010.04.011>
30. Ruiz-Garcia L, Lunadei L (2011) The role of RFID in agriculture: applications, limitations and challenges. *Comput Electron Agric* 79(1):42–50. <https://doi.org/10.1016/j.compag.2011.08.010>
31. Rennane A, Benmahmoud F, Tayeb Cherif A, Touhami R, Tedjini S (2021) Design of autonomous multi-sensing passive UHF RFID tag for greenhouse monitoring. *Sens Actuators A Phys* 331: 112922. <https://doi.org/10.1016/j.sna.2021.112922>
32. Wang W, Zhang F, Feng T, Zhou X, Li (2021) Facilities based on passive RFID sensors on growth and physiological characteristics of *Lycium barbarum* seedlings. *Meas J Int Meas Confed* 168(January 2020):108385. <https://doi.org/10.1016/j.measurement.2020.108385>
33. Deng F, Zuo P, Wen K, Wu X (2020) Novel soil environment monitoring system based on RFID sensor and LoRa. *Comput Electron Agric* 169(September 2019):105169. <https://doi.org/10.1016/j.compag.2019.105169>
34. Rossing W (1999) Animal identification: introduction and history. *Comput Electron Agric* 24(1–2):1–4. [https://doi.org/10.1016/S0168-1699\(99\)00033-2](https://doi.org/10.1016/S0168-1699(99)00033-2)
35. Top 7: Considerations to Choose the Right RFID Tag
36. Reiners K, Hegger A, Hessel EF, Böck S, Wendl G, Van den Weghe HFA (2009) Application of RFID technology using passive HF transponders for the individual identification of weaned piglets at the feed trough. *Comput Electron Agric* 68(2):178–184. <https://doi.org/10.1016/J.COMPAG.2009.05.010>

37. Maselyne J et al (2014) Validation of a high frequency radio frequency identification (HF RFID) system for registering feeding patterns of growing-finishing pigs. *Comput Electron Agric* 102:10–18. <https://doi.org/10.1016/J.COMPAG.2013.12.015>
38. Borgese M, Dicandia FA, Costa F, Genovesi S, Manara G (2017) An inkjet printed chipless RFID sensor for wireless humidity monitoring. *IEEE Sens J* 17(15):4699–4707. <https://doi.org/10.1109/JSEN.2017.2712190>
39. Alvarez-Narciandi G, Motroni A, Pino MR, Buffi A, Nepa P (2019) A UHF-RFID gate control system based on a convolutional neural network. In: 2019 IEEE International Conference on RFID Technology and Applications (RFID-TA 2019), pp 353–356. <https://doi.org/10.1109/RFID-TA.2019.8892080>
40. Sarkar S, Gupta B (2019) A dual frequency circularly polarized UHF-RFID/WLAN circular patch antenna for RFID readers. In: 2019 IEEE International Conference on RFID Technology and Applications (RFID-TA 2019), pp 448–452. <https://doi.org/10.1109/RFID-TA.2019.8892107>
41. Turki B et al (2019) Chemical vapor detecting passive RFID tag. In: 2019 IEEE International Conference on RFID Technology and Applications (RFID-TA 2019), pp 113–115. <https://doi.org/10.1109/RFID-TA.2019.8892161>
42. Nguyen VH, Diallo A, Le Thuc P, Staraj R, Lanteri S, Carle GF (2018) Wireless interrogation of small animal phantoms with a miniature implanted UHF RFID tag. In: 2017 IEEE Conference on Antenna Measurements & Applications (CAMA 2017), vol 2018-Janua, pp 306–309. <https://doi.org/10.1109/CAMA.2017.8273434>
43. Javed N, Habib A, Amin Y, Loo J, Akram A, Tenhunen H (2016) Directly printable moisture sensor tag for intelligent packaging. *IEEE Sens J* 16(16):6147–6148. <https://doi.org/10.1109/JSEN.2016.2582847>
44. Kim J, Wang Z, Kim WS (2014) Stretchable RFID for wireless strain sensing with silver nano ink. *IEEE Sens J* 14(12):4395–4401. <https://doi.org/10.1109/JSEN.2014.2335743>
45. Kamyshny A, Magdassi S (2014) Conductive nanomaterials for printed electronics. *Small* 10(17):3515–3535. <https://doi.org/10.1002/SMLL.201303000>
46. Boyer C, Roy S (2013) Space time coding for backscatter RFID. *IEEE Trans Wirel Commun* 12(5):2272–2280. <https://doi.org/10.1109/TWC.2013.031313.120917>
47. Bari BS et al (2020) Bandwidth and gain enhancement of a modified ultra-wideband (UWB) micro-strip patch antenna using a reflecting layer. In: Nasir ANK et al (eds) *InECCE2019*. LNEE, vol 632, pp 463–473. Springer, Singapore. https://doi.org/10.1007/978-981-15-2317-5_39
48. Bari BS et al (2020) Performance comparison of early breast cancer detection precision using AI and Ultra-Wideband (UWB) bio-antennas. In: Bhuiyan T, Rahman MM, Ali MA (eds) *Cyber Security and Computer Science*. ICONCS 2020. LNICS, Social Informatics and Telecommunications Engineering, vol 325, pp 354–365. Springer, Cham. https://doi.org/10.1007/978-3-030-52856-0_28
49. Khan S, Tinku S, Lorenzelli L, Dahiya RS (2015) Flexible tactile sensors using screen-printed P(VDF-TrFE) and MWCNT/PDMS composites. *IEEE Sens J* 15(6):3146–3155. <https://doi.org/10.1109/JSEN.2014.2368989>

Investigating Vehicle Characteristics Behaviour for Roundabout Cornering



Ashaa Supramaniam, Muhammad Aizzat Zakaria,
Mohamad Heerwan Peeie, and Gerogious Papaioannou

Abstract The allowable range of speed that a vehicle can tolerate in a constant radius turn is crucial for the development of smart assistance systems. Although the development of advanced system observers has been grown since early days of its introduction, extensive study is required in monitoring the vehicle's behaviour in the conditions such as variation of vehicle dynamic parameters and terrain type. Autonomous vehicles will fail to judge the parameter of the road cornering due to the safety constraints of the vehicle. Thus, the primary concern of this paper is to study the vehicle's behaviour for different curvature profiles. A real-time simulation for a typical Sedan is presented to test a constant roundabout turning with a radius of 50 m for this measure. In prior to that, a detailed analysis on the vehicle stability and handling responses are discussed. The vehicle is found to be traveling in a stable region at a speed from 10 to 74 km/h. The vehicle enters a critical area when speed is more than 74 km/h. Therefore, that the allowable range of speed that the vehicle can travel in a 50 m radius turn lies between 10 to 74 km/h. The stability is evaluated by two criterions which are the yaw rate and sideslip angle.

Keywords Vehicle dynamics · Roundabout cornering · Stability analysis

1 Introduction

Vehicles has been a very important invention throughout the study of technology and its' growth of it in the automotive industry has shown tremendous improvement

A. Supramaniam · M. A. Zakaria (✉) · M. H. Peeie
Autonomous Vehicle Laboratory, Automotive Engineering Centre, Universiti Malaysia Pahang,
Pekan, Pahang, Malaysia
e-mail: maizzat@ump.edu.my

A. Supramaniam · M. A. Zakaria
Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and
Mechatronics Engineering Technology, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia

G. Papaioannou
KTH Vehicles Dynamics, Department of Engineering Mechanics, KTH Royal Institute of
Technology, SE-100 44, Stockholm, Sweden

over the past few decades. Due to this, number of vehicles, especially cars have been increasingly seen on the roads. Increasing number of vehicles on road directly influence the frequency of road crashes and therefore an identification of the cause of this is needed to help reduce the issue. In prior to that matter, vehicle stability has been the biggest concern in reducing the risk of road accidents. Active driver assistance systems embedded in vehicles has been proven sustainable in most of the automotive industries. Among the popular vehicle stability controls are the Electronic Stability Control (ESC), Adaptive Cruise Control (ACC), Forward Collision Warning System, Active Rollover Protection and Anti-lock Braking System (ABS) [1]. All these active control systems were introduced into the market in an effort to reduce the number and severity loss-of-control automobile accidents. These systems are not only certified to human driven vehicles, but also in autonomous vehicles (AV). Generally speaking, the stability active systems are made up a monitor and a controller [2]. The monitor detects when the vehicle leaves a stable behavior using acceptable vehicle dynamics variables. In this paper, the primary concern is to monitor the vehicle's stability and behavior for a roundabout cornering.

The stability of the vehicle is observable by determining several criteria such as the lateral acceleration, the vehicle slip, side slip and the yaw rate of the vehicle. In order to conjugate with curvy road conditions, the vehicle's behavior needs to meet the requirements accordingly. However, the problem arises when the vehicle navigates at different cornering speeds, especially tight roundabout corners. Autonomous vehicles will fail to judge the parameter of the road cornering due to the safety constraints of the vehicle. Considering the kinematic of cornering, the yaw rate and longitudinal speed of the vehicle alone does not aid in computing those stability criterions. Factors such as large vehicle slip angles and lateral acceleration will act as a disturbance in cornering [3]. This results in AV to experience instability during driving. Thus, the main objective of this paper is to classify the stability region of the vehicle for a 50 m radius road radius using vehicle simulator. The paper is organized as follows: Sect. 3 reviews on the previous work done by other researches. Section 4 presents the process design of the vehicle simulator and its application to the vehicle system. Section 5 presents the stability and handling analysis and the simulation results are illustrated. Section 5 concludes the paper and shows recommendations for future works.

2 Previous Work

The lateral stability is the crucial feature in every type of vehicle, whether in a distributed drive electric vehicle (DDEV), ordinary electric vehicle (EV), autonomous vehicle (AV) or fossil fuel-based vehicles. A high-speed vehicle in a sharp turn may lose the lateral stability when it encounters fast varied road adhesion coefficients. Many researches have considered the study on lateral motion as a major part in vehicle stability, performance analysis and control [4, 5]. The real performance of the vehicle's lateral dynamics control mechanism is closely connected to the lateral stability of the vehicle. As far as the study of the vehicle stability area

is concerned, few cases have been documented as opposed to the aforementioned vehicle lateral dynamics control [6].

These methodologies show different indication accordingly. Some work requires constant and the negligence of parameters to suit with the type of study the researches are working on. The scope of the work in stability analysis for transient vehicle varies according to the need of the running conditions. Despite of the methodology used, the lateral stability analyses in [7, 8] indicated that the reliability of the balance point disappears as the speed of the vehicle and the steering angle increase above a certain limit. This shows that the vehicle exhibits instability behavior and is affected by the inputs of the longitudinal speed and steering angle while turning. The authors have also assumed the constant velocity and tire slip ratio. Therefore, the effect of the tire slip ration on lateral stability was ignored.

From the point of view of road safety, driving patterns or the style of driving raise or decrease the risk of an accident, and the driver's reaction to a sudden and damaged condition will decide the severity of the accident. Hence, the importance and under-standing the safety limits at all times. Several scholars have defined the safety limits of automobiles according to the stability boundaries. Maneuver ability refers to the capacities of a vehicle moving to the correct heading after the control of the steering mechanism and the steering wheels. Stability refers to the vehicle capability to stand up to the external impedances and maintain the original heading of travel.

Most of the methods usually take into account the yaw rate to ensure cornering stability [9]. In fact, the sideslip angle [10] is also a crucial factor that determines the cornering stability and the dynamic behavior of the vehicle determined by both the yaw rate and the sideslip angle. It is widely believed that the sideslip angle and yaw rate of the vehicle are closely related to stability, especially when the case is related with lateral motion of the vehicle [11]. In order to minimize the risk of skidding or turnover, a relatively small angle of sideslip angle and a fair rate of yaw should be maintained in the vehicle while taking a turn. State variables, such as yaw rate and sideslip angle must be maintained within a safe range [12]. Therefore, the analysis of sideslip angle of the vehicle is needed [13–17].

3 Methodology

3.1 Vehicle Configuration

The vehicle simulator used in this work is the CarSim 2017. The vehicle configuration displays the basic parameters of the dynamics and kinematics of the vehicle. This includes the vehicle body, animator data and systems involved inside the vehicle. The vehicle body section consists of two main parts which are the rigid sprung mass and aerodynamics. In this study, a generic C-Class vehicle (Sedan) is used for the experiment runs. The vehicle parameter configurations are shown in Table 1.

Table 1 Vehicle parameters

Parameter	Value	Description
l_f	1.015 m	Distance of center of gravity to the front axles
l_r	1.895 m	Distance of center of gravity to the rear axles
m	1270 kg	Vehicle mass
I_{xx}	536.6 kg/m ²	Roll inertia
I_{yy}	1536.7 kg/m ²	Pitch inertia
I_{zz}	1536.7 kg/m ²	Yaw inertia

According to SAE convention, the reference point is located at the midpoint of the wheelbase, the mid-point of the track, at ground level, and the reference length is equal to the wheelbase.

The systems include the powertrain, brake system and steering system. The internal engine model is 125 kW. Generic 125 kW powertrain with an automatic transmission. The torque converter data has been tuned to work with the 125 kW to give reasonable performance. It also uses a viscous type differential of a gear ratio 4: 1 and an internal transmission model (18 gears). The brake system applied is without ABS. The front and the rear suspensions have spin inertia of 0.9 kg/m² for each side. The un-sprung mass for both sides is 71 kg. The effective rolling radius is 325 mm and un-loaded (free) radius is 334 mm. These radius constants are used for the calculation of the vertical force, F_z . The maximum allowed force is set to 100,000 N.

3.2 Procedure Configuration

This section gives us information on the controls and environmental settings used for the runs. The set up follows the following condition:

- The speed control is set to a constant target speed. Throughout the run, this speed acts as a variable to determine the allowable range of speed.
- No open-loop braking is applied.
- Shifting is performed with a closed-loop controller that will use the shift schedules for the vehicle, using all gears available in the vehicle.
- Steering is set with a closed-loop path follower that will follow the target path (road centerline) using 1.0 s preview time.
- The direction on the road is set such that the vehicle goes forward. The road was defined such that going forward implies a turn to the left.
- Condition are set to stop the run when the vehicle has covers exactly one lap around a looped driver path.

Table 2 Input and output of the simulation model

Input	Output
<ul style="list-style-type: none"> • Longitudinal velocity (m/s) • Road radius (m) 	<ul style="list-style-type: none"> • Yaw rate (rad/deg) • Lateral acceleration (m/s²) • Sideslip angle (°) • Steering angle (δ)

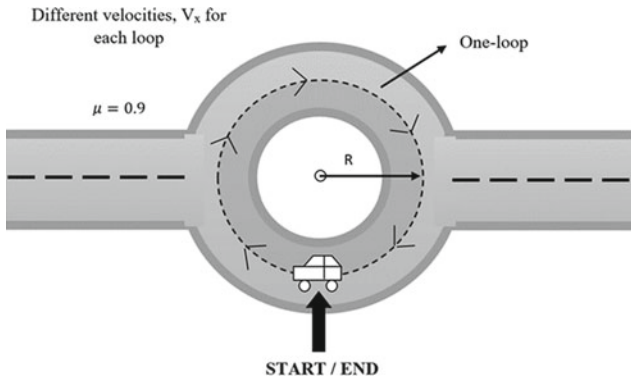


Fig. 1 Roundabout configuration

3.3 Road Configuration

The segment builder in the vehicle simulator dataset makes the provision that the road should be looped, and specifies a single turn to the left with an angle of 360° and a radius of 50 m. The length of the roundabout is about 628.32 m. The dataset for the entire road environment includes the surface friction dataset and some animation data. For this study, the friction coefficient is fixed to 0.9. The test used to generate the plots is ISO 4138. ISO 4138 recommends that the radius should not be less than 40 m [18]. Table 2 shows the description of the case studies that will be investigated (the illustration as shown in Fig. 1).

3.4 Stability Analysis

Prior to that, studies on determining the turning radius based on stability and safety factors are carried out. From the data obtained, the author can easily find out the range of allowable speed of the vehicle turning around a roundabout. The value of the longitudinal speed, V_x gradually increases at a gain of 10 until the vehicle is spotted to be in a critical area. From the test, it can be observed that there is a threshold where the vehicle starts to lose the trajectory but remains with a stable margin. Only few criteria are violated and the author propose to use this information

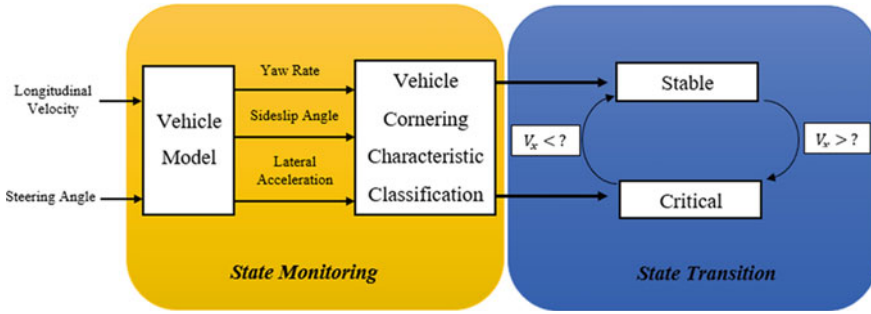


Fig. 2 State transition

to introduce a stable or critical area. The conditions will be elaborated in the next section.

The variable here is the road radius, R , whereby a range of constants are picked to be investigated. This work only includes the study on road radius of 50 m. For every test run, the road radius remain constant as the speed varies. The responding variables are the yaw rate, lateral acceleration, sideslip angle and steering angle. Table 2 represents the input and output of the simulation model. The input parameter given to the vehicle is the longitudinal velocity. From the outputs, the vehicle is classified into two main categories based on two main criteria which are the yaw rate and sideslip angle. The state transition is determined by the allowable range of speed that the vehicle can tolerate as shown in Fig. 2.

4 Results and Discussions

4.1 Yaw Rate

Vehicle yaw rate stability is bounded by (1). For each velocity profile, there is a maximum limit for allowable yaw rate dependent on the tire-road friction coefficient (TRFC) and longitudinal velocity of the vehicle. Thus, the limit varies for different vehicle velocity and TRFC. Based on calculations, the yaw rate limit increases with decreasing vehicle velocity. As observed, during the lateral motion, the longitudinal velocity of the vehicle decreases and the decrement is higher at increased velocity. On the contrary, the limit value decreases as the velocity increases. Referring to Table 3, from 10 to 70 km/h speed, the actual yaw rate of the vehicle has not reached its maximum limit. Therefore, the vehicle is said to be in a very stable manner and the handling capability is still safe. The table shows a constant value of the lateral acceleration for the corresponding velocity of the vehicle.

$$|\dot{\phi}| \leq (\mu g)/V_x \tag{1}$$

Table 3 Yaw rate at different velocities

Velocity (km/h)	Yaw rate (rad/deg)	Maximum yaw rate (rad/deg)	Percentage difference (%)
10	3.191	181.97	98.2
20	6.388	90.98	92.9
30	9.587	60.73	84.22
40	12.8	45.53	71.89
50	16.0	36.42	56.07
60	19.1	30.34	37.05
70	22.18	26.02	15.14

where ϕ is the maximum yaw rate of the vehicle, μ is the TRFC, g is the gravitational acceleration and V_x is the longitudinal velocity.

At 10 km/h, the yaw rate limit is extremely high due to high TRFC and very low vehicle velocity. Based on Eq. (1), the yaw rate limit is measured based on TRFC and vehicle velocity. Thus, the vehicle velocity 10 km/h causes the limit to be very high. Moreover, a slow moving vehicle does not easily get affected by instability and that is why the percentage difference is evidently proven to be very high. However, at 70 km/h, the yaw limit becomes lower due to very high vehicle velocity. At higher velocity, the yaw rate of the vehicle is increased. Besides, a decreasing trend can be seen in terms on the percentage difference between the yaw rate of the vehicle and the maximum limit of yaw rate as the vehicle travels faster in a lateral motion.

At 80 km/h, the graph turns to nonlinear and the vehicle is found less stable. The yaw rate of the vehicle is no more constant and starts to vary at different points of the roundabout. The first 3 and a half seconds of the run, the vehicle seems to experience a higher yaw rate. The increase in yaw rate is due to sudden change in turning radius of the vehicle which causes the vehicle to experience oversteer. However, the yaw rate of the vehicle remains lower than the stability limit. The yaw rate appears to decrease until the 8th second and remain a slightly constant rate after that and finally increases a little at the end of the lap. The vehicle is seen to be moving away from the original path at this point as the turning radius continue to increase as the vehicle is making a turn whereby the longitudinal velocity remains at 80 km/h. The turning of the wheels sharply during a fast driving to make a turning can cause back end of the vehicle to slide out (as shown in Fig. 3).

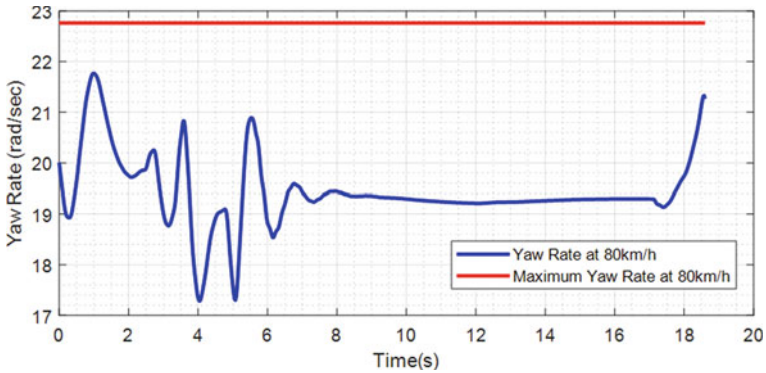


Fig. 3 Yaw rate of the vehicle at 80 km/h

4.2 Sideslip Angle

Another prior element that identifies the vehicle’s stability is the side slip angle. Limiting vehicle sideslip is important to ensure the vehicle does not slip from its intended path and crash. In fact, the sideslip angle of the vehicle indicates if the vehicle is moving in the right orientation with respect to its trajectory. This element also is bounded by (2) to compute the acceptable value of vehicle sideslip angle for a specific turn radius.

The maximum sideslip angle allowed for a 50 m radius is 10°. Based on Fig. 4, each sideslip angle for corresponding speed of the vehicle shows a value less than 10°. This evidence proves that the vehicle does not experience any skidding or turnovers while making a turn. Plus, the values are significantly small, thus it exhibits a good stability and handling capabilities. However, at 70 km/h, the sideslip appears to be a negative point. A negative sideslip means that the vehicle is experiencing skidding.

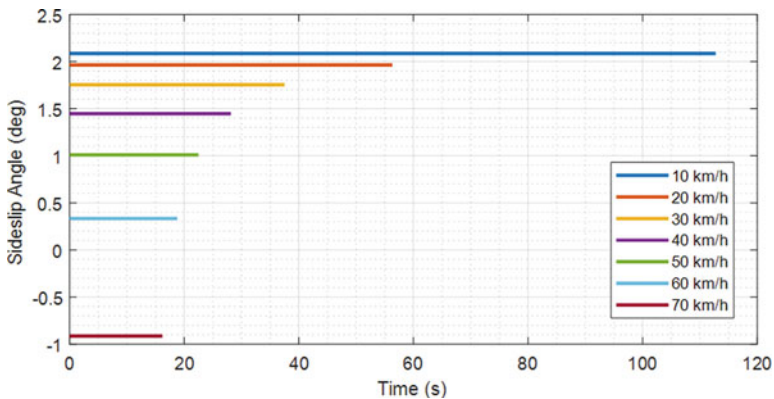


Fig. 4 Sideslip angle at 80 km/h

A skid happens when the tire loses their grip on the road and as for this scenario is due to high velocity of the vehicle which bring to the a oversteer or understeer condition. Although it is a very small change and it does not give a noticeable change, the vehicle can be said to be in a critical region. The small change is due to the road condition which is dry and not slippery.

$$\beta = \tan^{-1} 0.02 \mu g \quad (2)$$

where β is the allowable sideslip angle.

5 Conclusion

From the simulation results, the allowable speed range is determined. The stability analysis of the vehicle is examined based on two important criterion (a) yaw rate and (b) sideslip angle. The yaw rate of the vehicle seems to be maintained in a linear region from 10 to 74 km/h and starts to oscillate after 74 km/h. Based on the other criterion measured by the sideslip angle, the typical Sedan is said to have a linear and positive value from 10 to 60 km/h. The vehicle experiences skidding which happens to approach a negative value starting from 70 km/h onwards. According to the study, the allowable speed for this specific constant radius turn is within 10 to 74 km/h. This shows that as long as the vehicle stays within this range of speed, the vehicle is said to be in a stable position. The underlying equations is important in order for a vehicle to make a perfect cornering. For future research, different road radius can be investigated, focusing on sharp cornering (bigger curvature) for a better and wider insight on the estimation of the vehicle's behaviour during lateral motion.

Acknowledgements The authors would like to thank Ministry of Higher Education (KPT) and Universiti Malaysia Pahang (www.ump.edu.my) for financial supports given under FRGS/1/2018/TK08/UMP/02/1 (University Reference: RDU190104), RDU1803139. The authors also thank the research team from Autonomous Vehicle Laboratory AEC, Innovative Manufacturing, Mechatronics and Sports Laboratory (iMAMS); who provided insight and expertise that greatly assisted in the present research work.

References

1. Kritayakirana K, Gerdes JC (2012) Autonomous vehicle control at the limits of handling. *Int J Veh Auton Syst* 10(4):271–296. <https://doi.org/10.1504/IJVAS.2012.051270>
2. Gasmi A, Boudali MT, Orjuela R, Basset M (2019) Multi-criteria stability combi-nation for yaw stability control of autonomous vehicles. *IFAC-PapersOnLine* 52(5):465–470. <https://doi.org/10.1016/j.ifacol.2019.09.074>
3. Borroni F, Tanelli M (2018) A weighting approach to the shared-control of lateral vehicle dynamics. *IFAC-PapersOnLine* 51(9):305–310. <https://doi.org/10.1016/j.ifacol.2018.07.050>

4. Yavasoglu HA, Tetik YE, Gokce K (2019) Implementation of machine learning based real time range estimation method without destination knowledge for BEVs. *Energy* 172:1179–1186. <https://doi.org/10.1016/j.energy.2019.02.032>
5. Jin X, Yin G, Chen J, Chen N (2019) Analysis of lateral stability region for lightweight electric vehicle using phase plane approach. In: *Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019*, pp 5461–5466. <https://doi.org/10.1109/CCDC.2019.8832747>
6. Chen K, Pei X, Ma G, Guo X (2016) Longitudinal/lateral stability analysis of vehicle motion in the nonlinear region. *Math Probl Eng.* <https://doi.org/10.1155/2016/3419108>
7. Jin XJ, Yin G, Chen N (2015) Gain-scheduled robust control for lateral stability of four-wheel-independent-drive electric vehicles via linear parameter-varying technique. *Mechatronics.* <https://doi.org/10.1016/j.mechatronics.2014.12.008>
8. Wang W, Fan J, Xiong R, Sun F (2016) Lateral stability control of four wheels independently drive articulated electric vehicle. In: *2016 IEEE Transportation Electrification Conference and Expo, ITEC 2016*, pp 1–5. <https://doi.org/10.1109/ITEC.2016.7520218>
9. Daily R, Bevly DM (2004) The use of GPS for vehicle stability control systems. *IEEE Trans Ind Electron.* <https://doi.org/10.1109/TIE.2004.824851>
10. Farrelly J, Wellstead P (1996) Estimation of vehicle lateral velocity. In: *IEEE Conference on Control Applications—Proceedings.* <https://doi.org/10.1109/cca.1996.558920>
11. Fukada Y (1999) Slip-angle estimation for vehicle stability control. *Veh Syst Dyn.* <https://doi.org/10.1076/vesd.32.4.375.2079>
12. Kimbrough S (1991) Coordinated braking and steering control for emergency stops and accelerations. American Society of Mechanical Engineers, Design Engineering Division (Publication) DE
13. Rajamani R, Phanomchoeng G, Piyabongkarn D, Lew JY (2012) Algorithms for real-time estimation of individual wheel tire-road friction coefficients. *IEEE/ASME Trans Mechatron.* <https://doi.org/10.1109/TMECH.2011.2159240>
14. Babak SJ, Hussain SA, Karakas B, Cetin S (2017) Control of autonomous ground vehicles: a brief technical review. *IOP Conf Ser Mater Sci and Eng* 224(1). <https://doi.org/10.1088/1757-899X/224/1/012029>
15. Yang S, Zhou Y, Liu Y, Liu D (2014) Simulation analysis of an off-road SUV for steady-state performance. In: *IEEE Transportation Electrification Conference and Expo, ITEC Asia-Pacific 2014—Conference Proceedings*, vol 7, pp 5–7. <https://doi.org/10.1109/ITEC-AP.2014.6941254>
16. To CN, Hormoz Quoc CDV, Simic M, Khayyam H, Marzbani Jazar RN (2018) Autodriver autonomous vehicles control strategy. *Procedia Comput Sci* 126: 870–877. <https://doi.org/10.1016/j.procs.2018.08.021>
17. Rajamani R (2012) Lateral Vehicle Dynamics, pp 15–46. https://doi.org/10.1007/978-1-4614-1433-9_2
18. Karlsson A (2014) Test Procedures and Evaluation Tools for Passenger Vehicle Dynamics, p 95. <http://publications.lib.chalmers.se/records/fulltext/211557/211557.pdf>

Rice Disease Identification Through Leaf Image and IoT Based Smart Rice Field Monitoring System



Md Nahidul Islam, Fahim Ahmed, Md Tanvir Ahammed, Mamunur Rashid, and Bifta Sama Bari

Abstract Rice disease identification in early-stage, proper medication in case of disease affection and managing irrigation at the appropriate time is the most consequential phenomena to increase the production level of the rice. In this paper, a novel technique to diagnose the rice diseases and smart medication prescription system have been proposed. Furthermore, the internet of things (IoT) based smart rice field monitoring system has also been proposed. To identify rice diseases, the leaf image dataset (consists of healthy and three different diseases) has been analyzed through the convolutional neural network (CNN). The obtained rice disease diagnosis accuracy of the proposed system was 98.7%. In a real-time system, the leaf image data has been collected remotely using Raspberry Pi and the data has been sent to a server to be tested by a trained CNN model. Some sensors including soil moisture sensor, pressure sensor, humidity sensor, and temperature sensor have been implanted in the targeted field which aims to record the current scenario of the rice field and send the sensors data to the server. On a web page, proper medications have been displayed if any rice disease identified. Moreover, the user may monitor his field remotely which facilitates irrigation in opportune time.

Keywords IoT · Image processing · Deep learning · Convolutional Neural Networks (CNN) · Leaf disease detection

M. N. Islam (✉) · M. Rashid · B. S. Bari
Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia
e-mail: nahidul76.edu@gmail.com

F. Ahmed
Department of Electrical and Electronic Engineering, Khulna University of Engineering and
Technology (KUET), Khulna 9203, Bangladesh

M. T. Ahammed
Department of Electrical and Electronic Engineering, Jashore University of Science and
Technology, Jashore 7408, Bangladesh

1 Introduction

According to United Nation (UN), world population will reach 9.7 billion by 2050, this may cause a lot of challenges, mainly on food production. To face these challenges, it is mandatory to increase food production by 70% globally to feed the increased population in 2050 [1]. Climate change, unnecessary use of pesticide, irregular irrigation are the main reasons behind the reduction of food production. The problem regarding to economical disease protection is closely associated with the issues of sustainable agriculture as well as climate change [2]. Analyzing and monitoring of rice plant diseases are manually done by the farmers that needs more time and effort for them to assess certain damage to the plant. Inexperienced and over usage of pesticide can lead long-term resistance development of the pathogens. In practice, the disease detection is done by eye observation of the symptoms by a trained professional, so the person should have good observational skills. But it becomes difficult for most of the farmers and hobbyists. An automated plant disease identification system can play a vital rule for most of the farmers and hobbyists by classifying symptoms. It can reduce unnecessary waste of pesticide and can minimize financial cost of the resources.

Various techniques are recently used for plant diseases detection using machine learning technique. Proper agricultural field monitoring and precious disease detection can improve production level of crops. Now a day's different computer-based techniques are used to identify plant disease. A method of disease identification by color feature extraction is presented in. Due to rapid improvement in technology, many researches are focused on implementation of Artificial Intelligence (AI) in agriculture sector. Support Vector Machine based sugar beet diseases detection is presented in [3]. Zhang et al. [4] proposed a three-channel CNN model based on RGB colors to detect vegetable leaf diseases. Konstantinos et al. [5] performed several pre-trained CNN models on a large open leaves' dataset. Their studies show that CNN is highly suitable for automatic plant disease identification. Lee et al. [6] introduce a hybrid model to extract contextual information of leaf features using CNN and Deconvolutional Networks (DN). Deep Neural Networks based recognition of different plant diseases by leaf image classification is presented in [7]. In smart irrigation system proper identification of plant disease as well as proper monitoring of field condition is very essential. Forecasting about rain and irrigation at perfect time can reduce irrigation cost as well as waste of water. Smart irrigation over Internet of Things (IoT) is presented by many authors [8], they showed the effectiveness of smart irrigation in agricultural field. From the literature review it is observed that many researchers work with plant disease detection, but nobody has done combined agricultural field monitoring system and plant disease identification remotely where people can see their field condition from anywhere.

The major contributions of our work are as follows:

1. In this paper, the main target is to design a low-cost wireless system for agricultural field monitoring and plant disease identification with possible medication.

2. IoT based digital method is applied for monitoring agricultural field monitoring system where the user can get information from anyplace.
3. The proposed system can provide remote host and a user can access to his ground from any place.

In this paper, Raspberry Pi, Arduino, Pi camera, moisture sensor, humidity sensor, temperature sensor and pressure sensor are used to build the model. All the results and observation shown that the model is most effective for an agricultural field monitoring system.

This paper is organized as follows: Sect. 2 describes the proposed system whereas the operational workflow is shown in Sect. 3. Section 4 exposes image processing techniques and Sect. 5 shows results and discussion; followed by a conclusion in Sect. 6.

2 Brief Explanation of Proposed System

IoT provides remote collection and exchange of different sensor data over internet. A method has been proposed which is able to send sensor data from field to a webpage and after monitoring the field condition it is possible to on/off pump through internet. Apache web server has been used in Raspberry Pi for this purpose, which enables it to serve web pages, the server and the client can communicate via HTTP protocol. Then, Ngrok communication protocol has been used which is a cross platform tunneling software that can be used to establish secure tunnels from the internet to the local network. It is also possible to monitor the current image of the crops and view the analyzing result on the webpage by commanding the server.

2.1 Data Collection

The rice leaf dataset has been collected from Online (Kaggle) [9]. Total four class leaf data have been collected (Brown spot, Hispa, Leaf Blast and Healthy). From the dataset 500 images has been collected as per class. Total 2000 images have been collected. In this study we build an automatic system which can collect leaf image from lab, able to detect leaf condition and send the result to webpage. These entire incidents happen by putting a single command on the webpage. In our model, real-time image can be captured in Raspberry Pi camera [10] by putting a command on the webpage and send the result to webpage. In order to check the validation of our model, online dataset has been used in this study.

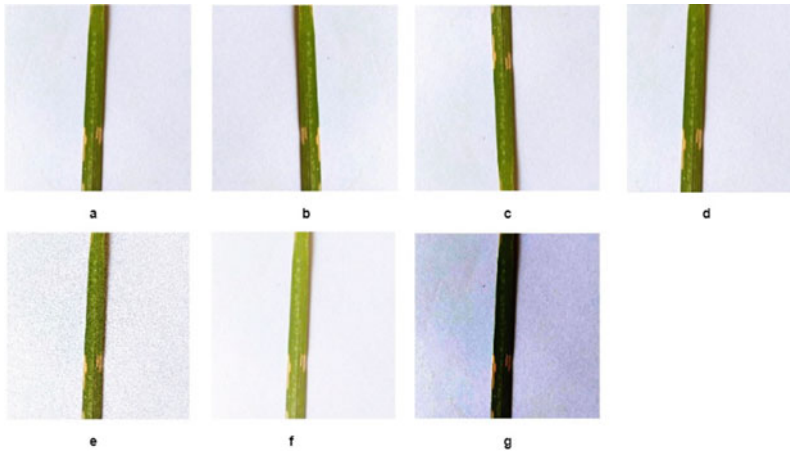


Fig. 1 Data augmentation of rice leaf disease images: **a** Original image **b** Horizontal flip **c** Vertical flip **d** Cropping one side of the image by 30% **e** Gaussian noise **f** High brightness **g** Low contrast

2.2 Data Augmentation

Data augmentation is the process to overcome the overfitting during the training stage of CNNs. The overfitting problem occurs when random noise or errors, rather than under-lying relationship, are described [11]. With the help of data augmentation, more image was generated from each image to learn the model as many irrelevant patterns occurs during the training process. For data augmentation operations, several techniques are used as rotation transformations, horizontal and vertical flips, and intensity disturbance which includes disturbances of brightness. Cropping removes columns or of pixels at the sides of images. In the example below, we are cropping one side of the image by 30%. A Gaussian noise processing operation is also applied Natural sources like thermal are responsible for the Gaussian noise. Atom vibration and distinct radiation character of warm artifacts. In digital images, Gaussian noise the gray values. To train the model with training data set Gaussian noise images were used for better results. Via the above operations, 7 new diseased images are generated from each image, as shown in Fig. 1. Finally, the dataset containing 14,000 images have been created using data augmentation.

3 Operational Workflow

3.1 Sending Sensor's Data to User

In this research three types of sensors i.e. Barometric Pressure Sensor—MPL115A1 Breakout, Grove—Moisture Sensor, DHT22 DHT-22 AM2302 Digital Temperature

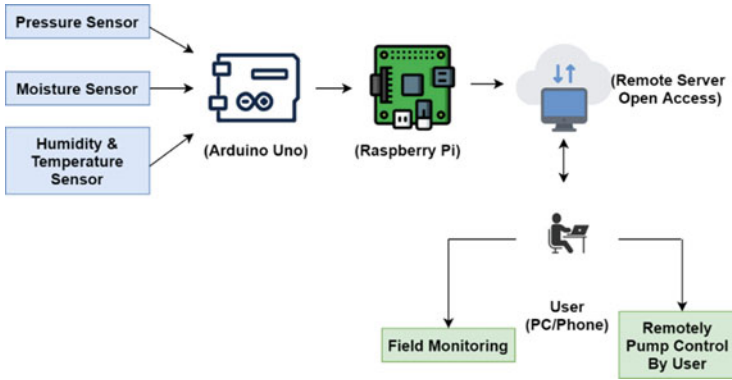


Fig. 2 Sensor’s data shown in remote server and user can control pump via server

and Humidity Temperature Sensor have been used. In order to measure the soil moisture of plants, The Grove—Moisture Sensor has been used. Using analog-to-digital converter (ADC) it can be used very effectively by simply inserting the sensor into the soil and reading data. To measure the accurate pressure measurements between 50 and 115 kPa, The Barometric Pressure Sensor has been used. DHT22 output calibrated digital signal. With the help of this sensor, the temperature and humidity can be measured. Then all the sensors data are read by Arduino Uno and the values are sent to Raspberry Pi Apache Server [12] via internet. With the help of these sensors data, a user can reckon the environment around his field. Then, the user can decide easily whether to irrigate or not and if required, the user can on or off his pump through server remotely. The process is shown in Fig. 2.

3.2 Sending Leaf Image and Condition to User

Real time image of the leaf can be seen easily through server by commanding the server to capture image. If the leaf contains a disease, the server will tell the disease name, symptoms and its medication to the user by analyzing the image and in case of healthy leaf, the server will show “Healthy Leaf”. In this study, python language has been used for developing the CNN model and analyzing the images. The process is shown in Fig. 3.

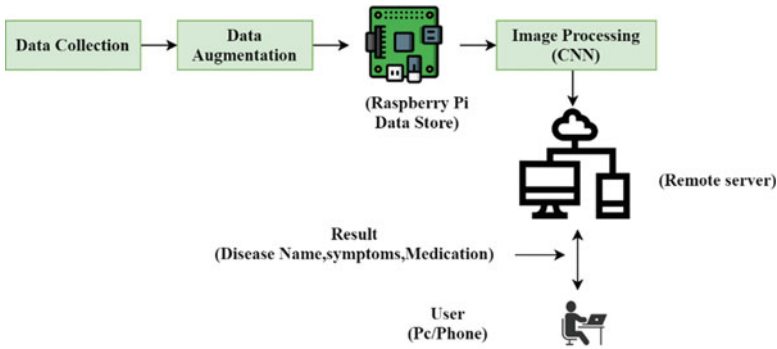


Fig. 3 Analyzing result (name, symptoms and medication) shown in server

4 Image Processing Using CNN

Convolutional neural networks (CNN) [13] proposed by Lecun have become one of the most successful methods in the field of pattern recognition. Locally trained filters are used in these structures for extracting visual features over the input image. In the pooling operation the feature maps size has been reduced and use as input image of the next convolution. This process continues until deep features are extracted. Following these steps, a classifier usually takes a decision on these features [14]. A deep convolutional neural network generally refers to a structure involving convolutional layers, pooling layers, and a fully connected network [15]. Convolution operations are used for extraction of features in this structure, while full connected network is a classifier on this feature. For classification purposes the fully connected component can end up with a SoftMax output layer.

4.1 Convolution Layer

Convolution layer is a structure with several fixed-size filters that allows complex functions to be applied to the input image [14]. This process is carried out by sliding the locally trained filters over the image. During this system each filter has the same weight and bias values throughout the image. This is called the weight sharing mechanism and this mechanism enables the whole image to be represented with the same feature. Local receptive field of a neuron represents the area which the neuron is connected in previous layer. The receptive field size is determined by the filter size. Let $m \times n$ and $c \times c$ be the size of the input image and the size of the kernel, i represent the image, w and b are weight and bias values of the filter, respectively. Output $O_{0,0}$ can be calculated as in Eq. (1) where f is activation function. Either ReLu or sigmoid can be used as activation function on this process[15]. The behavior of the ReLu activation function can also be seen in Eq. (2),

$$O_{0,0} = f \left(b + \sum_{t=0}^c \sum_{r=0}^c w_{t,r} i_{0+t,0+r} \right) \tag{1}$$

$$f(x) = \begin{cases} x & x > 0 \\ 0 & else \end{cases} . \tag{2}$$

4.2 Pooling Layer

In the feature maps the pooling process is applied and that have passed through convolution and activation function. This makes it possible to generate smaller feature maps, which are the summaries of the input feature maps. Pooling is performed by sliding a window on the image to apply the selected operation. The well-known pooling operations are maximum, average and L2 pooling. While average pooling takes the average of the input values, maximum pooling passes the maximum value throughout and L2 pooling calculates the L2 norm of the input. The main advantages of the pooling operation are the reduction of the image size and the extraction of the visual features independently on the image [16].

4.3 Fully Connected Layer

The data is converted into a one-dimensional vector after the convolution and the pooling layers. This vector will be the input of the fully connected network. The fully connected structure may have one or more hidden layers. Each neuron multiplies the connection weights by the data from previous layer and adds a bias value. Before transmission to the next layer, the calculated value passes through the activation function. The calculations made by a neuron in this layer can be seen in Eq. (3).

$$fc1 = f \left(b + \sum_{q=1}^M w_{1,q} * O_q \right) \tag{3}$$

where, f is the activation function, w is the weight vector, O is input vector of the q th neuron and b is the bias value.

4.4 Softmax

The softmax activation function is a multi-class adapted version of the logistic regression and normally used for classification purposes in the output layer. It is defined

as in Eq. (4) [17],

$$class_j = \frac{\exp(sf_j)}{\sum_q \exp(sf_q)} \tag{4}$$

5 Result and Discussions

This paper has been proposed a method combining IoT and image processing. In IoT part, three sensors (moisture, pressure and humidity) has been used. After getting all the sensors value, a user can concern about the current environment of his agricultural field. The dataset has been trained using CNN. Here Raspberry Pi 3 Model B [18] is used as computer. In this purpose Python has been used for image processing in CNN and free hosting of Raspberry Pi is used for web application. In this paper, we have worked with three different disease and healthy leaf of rice. We have collected 2000 rice leaf images and via image augmentation, we achieved 14,000 images for classification. Finally, 80% of leaf data has been stored in the training directory in order to train the dataset and the remaining 20% has been used for testing. The Fig. 4 demonstrate that the accuracy and loss curve and the performance of CNN. Finally, 98.7% accuracy has been obtained at 50 epochs and the learning rate was 0.0001.

The real-time leaf image can be taken from a remote place that can be used for viewing in the webpage. we also made a module that can be tested the captured image and able to show the leaf condition automatically to the webpage. In this paper, three common diseases that frequently occur in rice leaf is tested. The whole outcomes are discussed below:

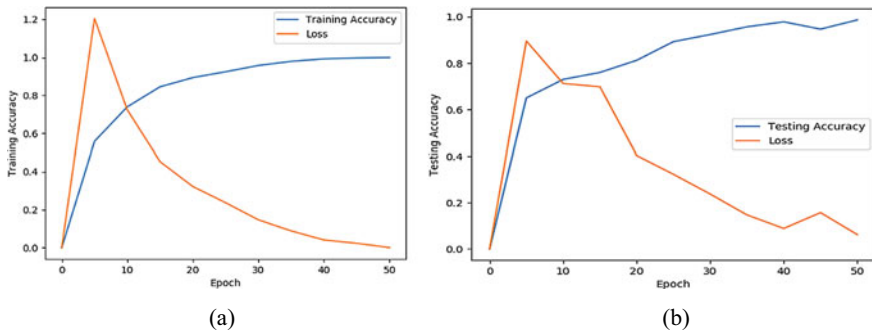
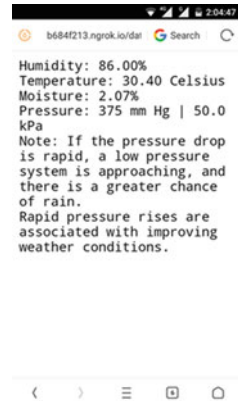


Fig. 4 a and b represent the Performance of CNN

Fig. 5 Sensors data shown in webpage



5.1 Sensors Data in Webpage

If there is a rapid drop of air pressure, a low-pressure environment is approaching that indicates a greater chance of rain. The User can also check humidity, temperature and moisture value through the webpage. After checking all these sensors data, he can concern about the field condition and control the pump manually by staying at home through the remote server. The result is shown in Fig. 5.

5.2 Leaf Image, Disease Name and Medication Showing in Webpage

Three types of rice diseases (Brown spot, Rice blast, hispa) and healthy has been predicted which occur frequently. The image can be achieved by commanding the server and it can be seen through a web-browser. The result is shown in Fig. 6. The image commanding by the user gets compared with the trained CNN model and shows the result to the webpage as shown in Fig. 7. Other two class result can be mentioned in the same way.

From all the figures, it is clear that the proposed model is able to detect the rice leaf disease and then prescribes its possible medication.

Fig. 6 Leaf image of rice shown in webpage

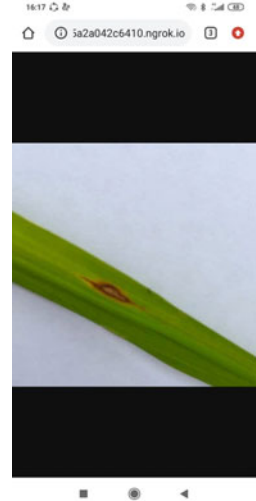
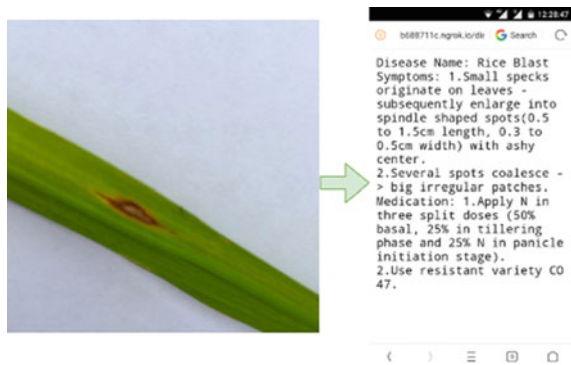


Fig. 7 Result showing in webpage (for rice blast disease)



6 Conclusion

Every year millions of crops are spoiling due to lack of proper irrigation, congruous monitoring to the plants and by the attack of various diseases. Identification and classification of rice disease is very difficult for farmers by naked eye. It needs a large quantity of time and human effort besides professional knowledge. To overcome the mentioned problems, a wireless system based smart agricultural system is developed here. The developed system can monitor agricultural fields, forecast weather condition and deliver plant leaf conditions to user. By using the sensor's data, a better forecasting system is possible for users. The user can reckon proper irrigation time using this system. The developed wireless system can identify three

different frequently occurred disease of rice crop. The system shows testing accuracy of 98.7% for the dataset. Therefore, this proposed method has been made fully automated for the classification of rice diseases.

References

1. World Population Growth - Our World in Data. <https://ourworldindata.org/world-population-growth>. Accessed 29 May 2020
2. Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE (2006) Climate change effects on plant disease: genomes to ecosystems. *Ann Rev Phytopathol* 44:489–509. <https://doi.org/10.1146/annurev.phyto.44.070505.143420>
3. Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plümer L (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput Electron Agric* 74:91–99. <https://doi.org/10.1016/j.compag.2010.06.009>
4. Zhang S, Huang W, Zhang C (2019) Three-channel convolutional neural networks for vegetable leaf disease recognition. *Cogn Syst Res* 53:31–41. <https://doi.org/10.1016/j.cogsys.2018.04.006>
5. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. *Comput Electron Agric* 145:311–318. <https://doi.org/10.1016/j.compag.2018.01.009>
6. Lee SH, Chan CS, Mayo SJ, Remagnino P (2017) How deep learning extracts and learns leaf features for plant classification. *Pattern Recogn* 71:1–13. <https://doi.org/10.1016/j.patcog.2017.05.015>
7. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep neural networks based recognition of plant diseases by leaf image classification. *Comput Intell Neurosci* 2016:1–11. <https://doi.org/10.1155/2016/3289801>
8. Roopaei M, Rad P, Choo KKR (2017) Cloud of things in smart agriculture: intelligent irrigation monitoring by thermal imaging. *IEEE Cloud Comput* 4:10–15. <https://doi.org/10.1109/MCC.2017.5>
9. Rice Diseases Image Dataset | Kaggle, <https://www.kaggle.com/minhhuy2810/rice-diseases-image-dataset?> Accessed 29 May 2020
10. Buy a Camera Module V2 – Raspberry Pi. <https://www.raspberrypi.org/products/camera-module-v2/>. Accessed 29 May 2020
11. Heisel S, Kovačević T, Briesen H, Schembecker G, Wohlgemuth K (2017) Variable selection and training set design for particle classification using a linear and a non-linear classifier. *Chem Eng Sci* 173:131–144. <https://doi.org/10.1016/j.ces.2017.07.030>
12. Setting up an Apache Web Server on a Raspberry Pi - Raspberry Pi Documentation. <https://www.raspberrypi.org/documentation/remote-access/web-server/apache.md>. Accessed 29 May 2020
13. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2323. <https://doi.org/10.1109/5.726791>
14. Ravi D et al (2017) Deep Learning for Health Informatics. *IEEE J Biomed Heal Inform* 21:4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
15. Nielsen M *Neural networks and deep learning* (2015). Determination Press
16. Earliest Diabetic Retinopathy Classification Using Deep Convolution Neural Networks.pdf | Request PDF. https://www.researchgate.net/publication/309260065_Earliest_Diabetic_Retinopathy_Classification_Using_Deep_Convolution_Neural_Networkspdf. Accessed 30 May 2020
17. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y (2013) OverFeat: integrated recognition, localization and detection using convolutional networks
18. Raspberry Pi Foundation (2016) Raspberry Pi 3 Model B. Raspberry Pi Website, vol 2837

Chili Plant Classification Using Transfer Learning Models Through Object Detection



Muhammad Nur Aiman Shapiee, Amirul Asyraf Abdul Manan,
Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin,
and Anwar P. P. Abdul Majeed

Abstract This study presents the use of a Convolutional Neural Network (CNN) based detector to detect chili and its leaves in the chili plant image. Detecting chili on its plant is essential for the development of robotic vision and monitoring. Thus, helps us supervise the plant growth, furthermore, analyses their productivity and quality. This paper aims to develop a system that can monitor and identify bird's eye chili plants by implementing machine learning. First, the development of methodology for efficient detection of bird's eye chili and its leaf was made. The image labeling will provide the source of images between chili and leaf. The dataset would be split into training, verification and test set for 70:20:10%, correspondingly. The images YOLO V4 Darknet was implemented to train the dataset. After a series of experiments were conducted, the model is compared with other transfer learning models like YOLO V4 Tiny, Faster R-CNN, and EfficientDet. The classification performance of these transfer learning models has been calculated and compared with each other. The experimental result would discuss on part of hyper parameter optimization and transfer learning application. Firstly, the optimization of hyper parameter shows that the YOLO V4 Darknet model achieves mAP of 76.54%, followed by EfficientDet at 73.66% for 512×512 input layers. Next, the application of transfer learning. The result shows that YOLO V4 Darknet achieves highest mAP value, 75.69% follow by EfficientDet, with mAP of 71.85%.

Keywords Transfer learning · Machine learning · Object detection · Precision agriculture · Chili plant

M. N. A. Shapiee · A. A. Abdul Manan · M. A. Mohd Razman · I. Mohd Khairuddin ·
A. P. P. Abdul Majeed (✉)

Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and
Mechatronic Engineering Technology, Universiti Malaysia
Pahang, 26600 Pekan, Pahang Darul Makmur, Malaysia
e-mail: amajeed@ump.edu.my

A. P. P. Abdul Majeed
School of Robotics, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong - Liverpool
University, Suzhou 215123, P. R. China

1 Introduction

Agriculture is a potential commodity that has high economic value and a high potential for development. From the supply or production side, Malaysia land area with its Agro-climate diversity enables the development of various crops. Based on the Department of Statistics Malaysia, the agriculture sector contributed 7.1% to the Gross Domestic Product (GDP) in 2019, equal to RM 101.5 billion. Pepper production is one of several selected crops that show the increment of production from 2018 to 2019 [1]. In addition, the production of chillies has slightly escalated from 32.3 thousand tonnes in 2018 to 33.9 thousand tonnes in 2019. In chili production, farmers are the ones who are manually sorting the chili into their categories. Furthermore, this task also involves humans as decision makers to identify the quality of the chili, resulting in manually identifying a lot of time-consuming and labour-intensive tasks. Besides, due to the human visual limitations, product variety is also obtained, and it may also causing inconsistency in the selection process.

The application of object detection in agriculture is to monitor the condition of the crops, especially on a larger production scale. This research seeks to address how object detection can help reduce the time consumption of chili plant identification for crop production and help increase the chili production rate. Therefore, this study proposes to monitor the plant growth rate through image processing. Hence, will ensure that the condition of a plant, which is healthy and non-healthy, can be determined from time to time.

Unmanned Aerial Vehicle (UAV) is equipped with many sensors that can make a notable impact in field-based crop monitoring, providing the information and data from above such as phenotyping, estimation of plant density, or plant quantity proper and systematic order [2]. Many machine learning approaches have been introduced to solve classification problems through image processing and object detection, which employed classifiers such as k -Nearest Neighbor (k -NN) and Support Vector Machine (SVM) and not just for crop monitoring [3, 4]. Additionally, machine learning models have improved their accuracy and shown an excellent performance in classifying object-based in a wide range of applications [5]. Besides, there is much more potential for their use in agriculture-focused analysis. Many researchers used machine learning to solve agricultural-related problems such as leaf counting, plant counting, leaf segmentation, and field yield prediction.

On the features extraction aspects, it was studied to use five basic geometric approaches [6]. The detection features they used are the leaf diameter, width, length, perimeter, and area. Another study has proposed using Probabilistic Neural Network (PNN) to classify leaves that used the vein, shape, texture, and colour features in which the author manage to obtain a 93.75% average accuracy [7]. Likewise, research was done to study leaf biometric features and use hidden naive Bayes (HNB) as a classifier [8]. The experiment data shows an accuracy of 97% by using HNB performed with the Flavia dataset.

Nowadays, Convolutional Neural Network (CNN) outperform all other traditional methods in computer vision tasks [9], with some of them superior to human ability

on ImageNet image classification challenge [10]. Image classification is a task that trying to analyse an entire image as a whole [11]. The image will classify the image by assigning it a specific label obtained by labelling or classifying images into pre-determined categories [12]. AlexNet [13] is one of the convolutional networks that perform image classification. Object detection, however, involves the combination of classification and localisation of various objects in an image. Many object detector convolutional neural networks use a sliding window approach to inspect different regions of an image sequentially [13]. As a result, reducing their speed but have higher accuracy output. Some object detectors such as YOLO take another approach where the model examines the whole image and its context simultaneously, significantly speeding up the process [14].

Deep learning, such as object detection algorithms, is a great way to detect plant health and diseases. A study was conducted to detect diseases from plant leaves that proposed method using four different types of models for the detection, which are DarkNet-19, MobileNet, Inception v2, and ResNet-101. Finally, Darknet-53 achieved a 99.10% of success rate in the classification [15].

A proposed method of using CNN for the law-protected tree species detection in which they used three different object detection methods, Faster Region-based Convolutional Neural Network (Faster R-CNN), You Only Look Once (YOLO), and RetinaNet [16]. The models were then evaluated, and the acquired around 92% of average precision. The present study aims at evaluating different Transfer Learning pipelines for the detection of bird's eye chili.

2 Material and Methods

2.1 Experimental Setup

The purpose of this study is to answer the research objective to monitor and identify the plants and one or more object detection models trained to detect bird's eye chili. Nonetheless, bird's eye chili classification consists of four steps, namely dataset mining, pre-processing, model training, and evaluation. The processes done is shown in Fig. 1 where dataset mining will begin before the pre-processing stage starts. The image labelling will provide the source of images between leaf and chili. The model training will consist of data splitting of the images in turns provide the training of the models. Moreover, the hyper parameter optimization would be done to investigate

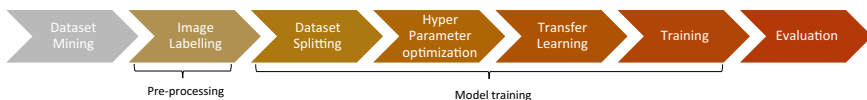


Fig. 1 Flow of machine learning process



Fig. 2 Bird' eye chili images sample

the relationship between the sizes of the input layer of the models and mAP correspondingly. Then, the transfer learning stage is applied as to improve the accuracy of the models. Lastly, the evaluation part will compared the performance of each models and determine the optimum classifier that could detect the chili plant.

2.2 Dataset Mining

A total of 298 bird' eye chilli images as shown in Fig. 2 and 324 images of healthy and unhealthy capsicum leaf as in Fig. 3 have been collected from capturing images of bird' eye chilli from various angles and open-source for model training [17].

2.3 Pre-processing

The images are then labelled by drawing a bounding box around the objects found in the image. This process was used to allow the object detection algorithm to learn to identify the trained objects. The annotation file contained the detail about the objects in the image such as the image name, label name, class name of the objects and the coordinates of the bounding boxes surrounding the object that was present in the image as in Fig. 4.

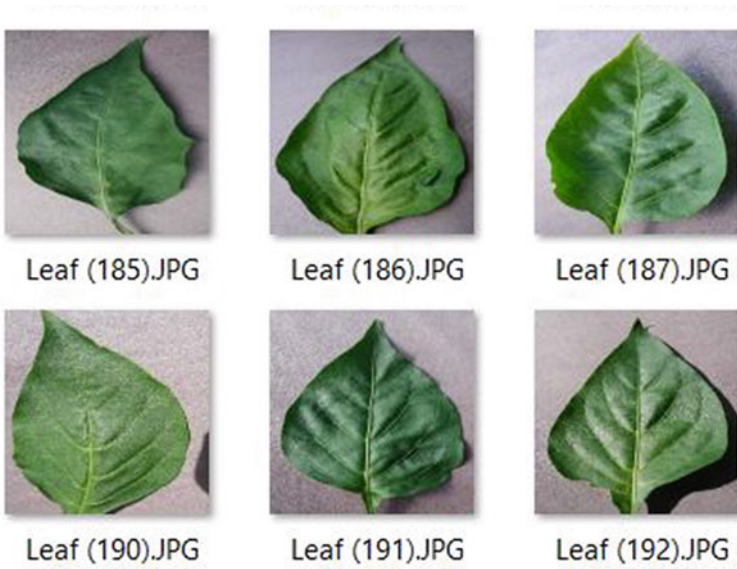


Fig. 3 Sample from plant village dataset

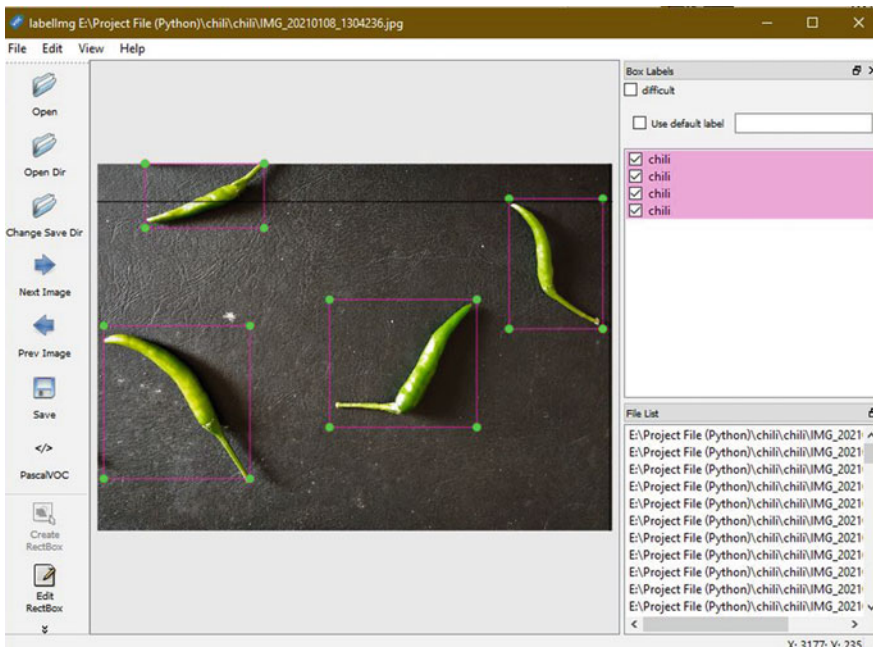


Fig. 4 Dataset labeling

Table 1 Parameter information

Name	Description
Decay	0.005
Total batch	6000
Batch size	32 images
Momentum	0.9
Learning rate	0.01

2.4 Model Training

In this section, a python code called training data splitter was used to split the dataset [18]. The dataset was randomly split into three sets: training set, verification and a test set with a distribution of 70, 20 and 10% of the images, respectively. A validation set was used for the model's evaluation in the selected iteration of the neural networks.

You Only Look Once, more commonly known as YOLO, is an algorithm used to detect objects of various sizes. YOLO is an object detection architecture composed of a single neural network [14]. The approach YOLO utilised is a single convolutional network that can simultaneously anticipate numerous bounding boxes and makes the model look once at a picture to foresee what objects are shown and where they are. Hence, it makes YOLO fast in detecting objects. Fast YOLO [14] or Tiny YOLO is an alternate implementation of YOLO V4 where it is less accurate but has faster training rate. The Faster Region-based Convolutional Network (Faster R-CNN) is an image processing algorithm that was derived from the Fast R-CNN model [19]. The Fast R-CNN is developed over previous concepts of R-CNN but with more efficiency added to identify objects. The model list from EfficientDet-D0 until EfficientDet-D7. It was claimed that their most basic model structure of EfficientDet-D0 can match YOLOv3 accuracy while being faster [20]. Furthermore, a study shows that EfficientDet-D2 achieves an mAP of 43.0% on the COCO test set, which would be 1% higher than Faster R-CNN [21].

The model is trained on the training set in which employed five constant parameters used for the model. Table 1 describes all the parameters. Furthermore, several hyper parameters are tested where the input layer of the model is modified into 512×512 , 416×416 , and 224×224 resolution.

2.5 Performance Evaluation

The model is evaluated on the test set and the evaluation metrics called mean Mean Average Precision (mAP), the method of evaluating the accuracy of the object detection model [14, 22, 23]. Hence, the metrics will be used in this study to evaluate the model's detection performance. Equation 1 shows the formula for Intersection over union (IoU).

$$\text{Intersection over union} = \frac{\text{area of overlap}}{\text{area of union}} \tag{1}$$

If the detection or IoU is more than 0.5, the object detection is classified as True Positive (TP), but if the IoU is less than 0.5, it is classified as False Positive (FP) as it is a false detection. Then, False Negative (FN) is when the model failed to detect the ground truth although the ground truth is in the image.

3 Results and Discussions

The outcome of the study regarding the performance of YOLO V4 Darknet, YOLO V4 Tiny, Faster R-CNN, and EfficientDet will be described. An analysis was done to evaluate the models' performance according to the image bounding boxes area in the ground truth to confirm the models' ability and performance in detecting the classes. Figure 5 illustrates the prediction made by YOLOv4 Darknet.

3.1 Accuracy of the Models on Hyper Parameter Optimization

Hyper-parameter optimization experiments suggest that in the neural network, there exists a direct relationship between the size of the input layer of the models and mAP

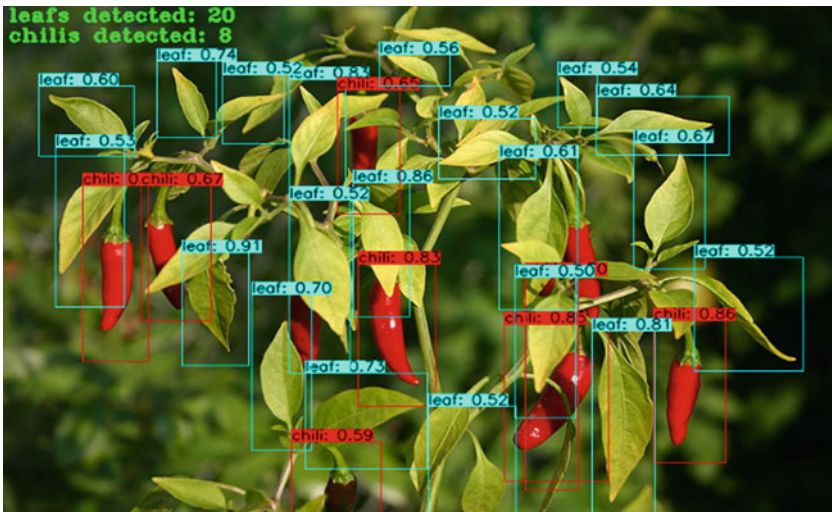


Fig. 5 Predictions made by YOLOv4 Darknet

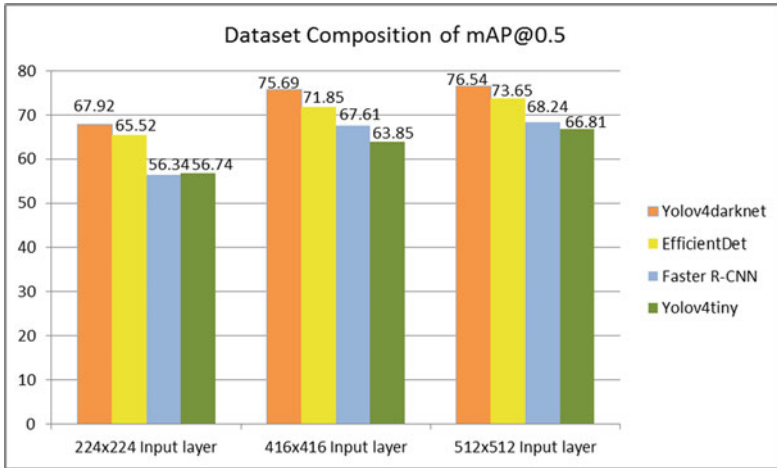


Fig. 6 Relationship between mAP and model input layers

respectively, as in the Fig. 6. From instance, YOLO V4 Tiny increases from 56.74, 63.85, and to 66.81%. Even though Faster R-CNN has the lowest accuracy at 56.34% for 224×224 input layer, it surpassed YOLO V4 Tiny when using 416×416 input layer with 67.61% and 512×512 input layer with 68.24%, which may be due to the architecture simplicity of Faster R-CNN [14]. The YOLO V4 Darknet model has the highest accuracy that rose from 67.92, 75.69 then 76.54% from 224×224 , 416×416 to 512×512 input layer, respectively. Meantime, the EfficientDet model came short with its best value of 73.65% for 512×512 input layer. Overall, the result suggests that the increasing the dimension of the input layer would produced a more accurate model.

3.2 Accuracy of the Models on Transfer Learning

The Fig. 7 below displays the average precision comparison between the models. From the Fig. 7, the average precision of bird’s eyes chili leaf is lower compared to bird’s eye chili across four models. Where YOLO Darknet achieved the highest mAP of 68.04%, followed by EfficientDet, Faster R-CNN and YOLO v4 Tiny at 63.18, 59.99 and 52.85%, respectively. Meanwhile, for bird’s eye chili, YOLO V4 Darknet attained 83.33% as the highest mAP, followed by EfficientDet at 80.52%, Faster R-CNN at 75.23% and YOLO v4 Tiny at 74.84%, correspondingly. When considering the average accuracy of the maximum mAP @ 0.5 over all the classes, as displayed in Fig. 8, the YOLO V4 Darknet gets a higher accuracy with 75.69%, followed by EfficientDet, Faster R-CNN and YOLO V4 Tiny with 80.52, 75.23 and 74.84%, correspondingly.

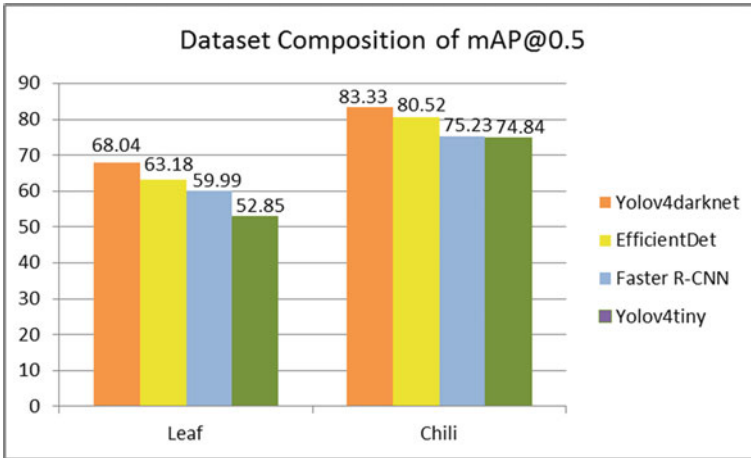


Fig. 7 Models' average precision comparison

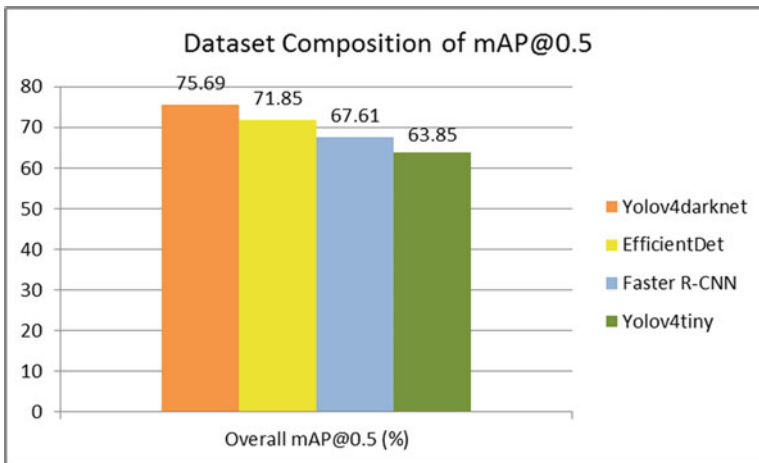


Fig. 8 Models' overall mAP comparison

4 Conclusion

The present study evaluated different object detection models in the classification of bird's eyes chili. It was shown the model can identify the bird's eyes chili and its leaves with a satisfying result. The model is also able to inform the quantity of the objects that are successfully detected. From four algorithms used for the comparisons, YOLO V4 Darknet, YOLO V4 Tiny, Faster R-CNN, and EfficientDet, the experiment result reveals that YOLO V4 Darknet performs well against the other models at the highest

mAP of 76.54 and 75.69% when using dataset composition with hyper parameter optimization and transfer learning, respectively.

Acknowledgements The authors would like to acknowledge Universiti Malaysia Pahang for funding this study under the Research Grant (RDU200332).

References

1. Department of Statistics Malaysia Official Portal (2019) Value of gross output of agriculture sector registered an annual growth rate of 11.1% to RM91.2 billion, April 2019
2. Koh JCO, Hayden M, Daetwyler H, Kant S (2019) Estimation of crop plant density at early mixed growth stages using UAV imagery. *Plant Methods* 15(1):1–9
3. Shapiee MNA, Ibrahim MAR, Razman MAM, Abdullah MA, Musa RM, Abdul Majeed APP (2020) The classification of skateboarding tricks by means of the integration of transfer learning and machine learning models. In: Mohd Razman M, Mat Jizat J, Mat Yahya N, Myung H, Zainal Abidin A, Abdul Karim M (eds) *Embracing Industry 4.0*. LNEE, vol 678, pp 219–226. Springer, Singapore. https://doi.org/10.1007/978-981-15-6025-5_20
4. Ng H, Nguyen VD, Vonikakis V, Winkler S (2015) Deep learning for emotion recognition on small datasets using transfer learning, pp 443–449
5. Puteh S, Rodzali NFM, Razman MAM, Ibrahim ZZ, Shapiee MNA (2020) Features extraction of capsicum frutescens (C.F) NDVI values using image processing. *MEKATRONIKA* 2(1):38–46
6. Karami A, Crawford M, Delp EJ (2020) Automatic plant counting and location based on a few-shot learning technique. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:5872–5886
7. Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL (2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In: *ISSPIT 2007 - 2007 IEEE international symposium. signal processing information technology*, pp 11–16
8. Kadir A, Nugroho LE, Susanto A, Santosa PI (2013) Leaf classification using shape, color, and texture features, pp 225–230
9. Eid HF, Hassanien AE, Kim TH (2016) Leaf plant identification system based on hidden Naïve Bays classifier. In: *Proceedings - 2015 4th International conference advanced information technology sensor applications. AITS 2015*, pp 76–79, August 2016
10. Alom MZ et al (2018) The history began from AlexNet: A comprehensive survey on deep learning approaches. arXiv
11. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification - IEEE conference publication. In: *Proceedings IEEE international conference computing vision*
12. Wang S, Su Z (2019) Metamorphic testing for object detection systems. arXiv
13. Kamavisdar P, Saluja S, Agrawal S (2013) A survey on image classification approaches and techniques. *Int J Adv Res Comput Commun Eng* 2(1):1005–1009
14. Krizhevsky A (2012) ImageNet classification with deep convolutional neural networks, pp 1–1432
15. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings IEEE computing society conference computing vision pattern recognition*, vol 2016, pp 779–788
16. Anjanadevi B, Charmila I, Ns A, Anusha R (2020) An improved deep learning model for plant disease detection. *Int J Recent Technol Eng* 8(6):5389–5392
17. dos Santos AA et al (2019) Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)* 19(16):1–11

18. PlantVillage Dataset | Kaggle. <https://www.kaggle.com/emmarex/plantdisease>. Accessed 28 July 2021
19. GitHub - bigrobinsn/Training-Data-Splitter: A simple python script to split data into train, test, and validation sets
20. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39(6):1137–1149
21. Tan M, Pang R, Le QV (2019) EfficientDet: Scalable and efficient object detection. *arXiv*, pp 10781–10790
22. Detection Evaluation, Common Objects in Context. <https://cocodataset.org/#detection-eval>
23. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: *Proceedings - 30th IEEE conference computing vision pattern recognition, CVPR 2017*, vol 2017, pp 6517–6525
24. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. *Comput Vis Pattern Recogn* 34(5):241–247

The Classification of Impact Signal of 6 DOF Cobot by Means of Machine Learning Model



Gavin Lim Jiann Kai, Ahmad Fakhri Ab. Nasir, Anwar P. P. Abdul Majeed, Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin, and Lim Thai Li

Abstract Collaborative robot (Cobot) has seen a rise in adoption rate in the industry as the Industry 4.0 era marches in. Cobot were introduced to replace human operators in harsh environments or repetitive work processes. The health condition monitoring of these cobot have not been standardized due to lack of widely available standardized fault dataset and the high complexity of diagnostic. This study aims to use machine learning algorithms as a mean to identify the cobot pick and place process offset error using vibrational signals. The vibrational sensor was attached to the end effector of the cobot where the vibration signal of 3 axis were collected. The features were then extracted, standardized, and 544 features were selected from 2337 features based on a hypothesis testing method. The dataset was then spilt into training and testing by a ratio of 80:20. Three machine learning models namely, the k -Nearest Neighbors (k -NN), Neural Network (NN), and Support Vector Machine (SVM) classifier were tested, and the classification accuracy of the models was analyzed. A grid search approach was used to identify the best hyperparameter for each model. The model with the highest classification accuracy of 95.2% was the MLP model compared to SVM (92.4%) and kNN (79%). Therefore, it could be established from the study that a comparable classification efficacy is attainable through the identification of significant features. The findings are non-trivial, particularly with respect to the implementation of the developed classifier in real-time.

Keywords Machine learning · Condition-based monitoring · Feature selection

G. L. J. Kai · A. P. P. A. Majeed · M. A. M. Razman (✉) · I. M. Khairuddin
Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia
e-mail: mohdazraai@ump.edu.my

A. F. Ab. Nasir
Faculty of Computing, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

A. F. Ab. Nasir · A. P. P. A. Majeed
Centre for Software Development and Integrated Computing, Universiti Malaysia Pahang,
26600 Pekan, Malaysia

L. T. Li
TT Vision Holdings Berhad, 11900 Plot 106, Sungai Hilir Keluang 5, Bayan Lepas FIZ.4, Pulau
Pinang, Bayan Lepas, Malaysia

1 Introduction

The industrial era has seen an increase in the adoption of industrial 4.0 where digital information of manufacturing systems are gathered and analyzed to provide vital insight into the production floor [1]. Entire factories are fitted with sensors and camera to provide real-time data for management to keep track of everything occurring. All of this is possible with the advent of Internet of Things (IoT), cloud services, and big data, creating the cyber-physical system of Industry 4.0 [2]. These emerging technologies have caused an uprising in machine monitoring using machine learning. Using prediction models by leveraging statistical techniques, machine health or condition could be extracted from sensor data [3]. The adoption of robotic manipulator or robotic arm in industrial setting have been increasing to replace human operators in repetitive tasks such as a pick and place, handover, and screwing operation [4]. In a production line, downtime is costly where a broken robot will cause an unexpected disruption. Thus, the ability to anticipate robot failure is an important tool to schedule for maintenance before a breakdown occurs [5].

Previous study on collaborative robots (cobots) have mainly focused on the research perspectives of safety while interacting with human workers [6, 7]. Sharp stated that even with increased efforts on improving fault diagnosis, there is still no standardized methodology for applying Prognostic and Health Management (PHM) that have been developed and adopted [8]. One part of PHM is conditional-based monitoring (CBM) where the health of machine parts within a predetermined scope can be assessed with a high degree of confidence thus reducing unnecessary maintenance but still be able to identify potential faults in a machine. Even though CBM has been in use for a long time, the integration with machine learning is novel and still provides challenges due to the complexity of industrial machines and the lack of comprehensive fault datasets. Time series data have shown to be favored where the peak of impact can be determined more easily as compared to data in the frequency domain or wavelet domain [9]. This research investigated the classification of impact vibration signals from a six degree of freedom collaborative robot by means of machine learning models.

2 Methodology

2.1 Data Acquisition Device and Configuration

The Balluff condition monitoring sensor was used to collect vibrational data in terms of velocity in mm per second and acceleration in mm per second per second. The sensor uses the MEMS technology and can detect velocity up to 220 mm per second at 79.4 Hz and acceleration of 16 times the gravitational acceleration. The sensor was attached to the end effector of the cobot to acquire vibration data in mm per second

Fig. 1 Balluff vibration sensor



Fig. 2 OMRON TM5-700 collaborative robot



per second. The time domain data was collected and used in this research (Figs. 1 and 2).

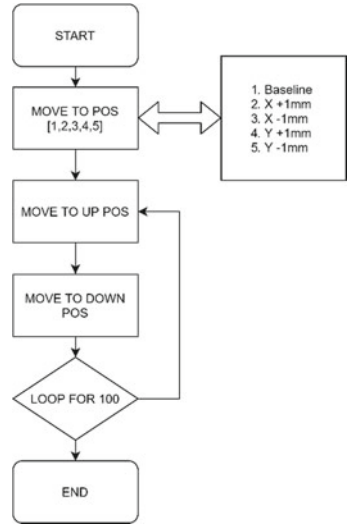
2.2 OMRON TM5 Collaborative Robot

The OMRON TM5 Collaborative robot is a six degree of freedom robotic manipulator which was designed to work alongside human operators and machines. The TM5 cobot weights 22.1 kg and can carry payload around four to six kg depending on the specific product. The reach of the cobot is 700 mm with a repeatability of 0.05 mm. The typical speed of the cobot is at 1.1 m/s.

2.3 Experimental Protocol

The experiment was carried out at TT Vision Sdn Bhd as a collaborative project with Universiti Malaysia Pahang. The cobot was placed in a clean workplace to facilitate the process of experimental investigation for data acquisition. After a few discussions and consultation with the research and development manager at TTV, an experimental outline was decided. The scope of the experiment will be based on a pick and place operation for the cobot, and the objective was to investigate the

Fig. 3 Cobot programming flow



performance of machine learning algorithms. A metal pin and cylindrical part was used to simulate a part insertion using the cobot. A 3D part was designed and printed using ABS filament to attach the metal pin to the end effector and the cylindrical part to the workbench.

The flowchart of the cobot program was shown in the figure below. The baseline offset for the pin was identified and fixed as position 1. The coordinates of the position 1 was set to align the pin concentric to the cylindrical part in the X and Y axis with a 20 mm offset in the Z axis (Up Position). The end effector than move downwards in the negative Z axis where the pin was inserted (Down Position). The insertion process was repeated for 100 cycles followed by the next class. The classes were predetermined to be Baseline (Class 1), Xp1 (Class 2), Xn1 (Class 3), Yp1 (Class 4), and Yn1 (Class 5). The signals were collected from the Balluff vibration sensor through a JavaScript software named Node-Red to save the data into Comma-Separated Values (.csv) format (Fig. 3).

2.4 Signal Preprocessing, Feature Extraction and Selection

The collected data were preprocessed by filtering and trimming the blank data and separating the data into five classes with their respective X, Y and Z axis vibration signals. A snapshot of the vibration data was shown in Fig. 5 where the spike indicates the vibration caused by the insertion operation. The data was the reshaped and classes were added. The features of the data were extracted using a python library called Tsfresh [10, 11] where fresh stands for FeatuRe Extraction based on Hypothesis tests. The total features extracted were 2337 (779 features each for channel X, Y

and Z respectively). Some examples of the features extracted apart from the standard statistical feature e.g., mean, max, min, mode, and median, were absolute energy (1), aggregate correlation (2), and autocorrelation (3). The features were standardized using the StandardScaler to remove outliers and centering the mean of the data to zero with a standard deviation of one. The features were then evaluated individually and independently with respect to its significance. The vector of p-values were evaluated based on the Benjamini-Yekutieli [12] procedure and features were selected if the p-value is higher than 0.05. A total of 544 features were selected from 2337 using this method (Fig. 4).

$$E = \sum_{i=1, \dots, n} x_i^2 \tag{1}$$

$$R(l) = \frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} (X_t - \mu)(X_{t+1} - \mu) \tag{2}$$

$$\frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} (X_t - \mu)(X_{t+1} - \mu) \tag{3}$$

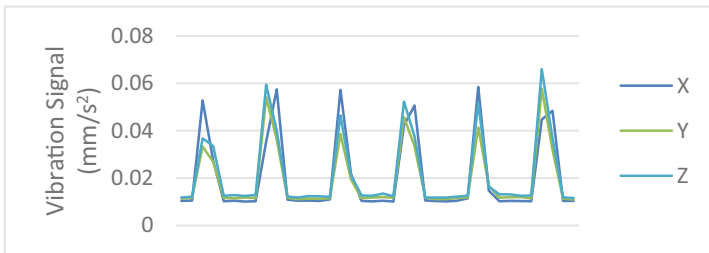


Fig. 4 Vibartion signals from accelerometer

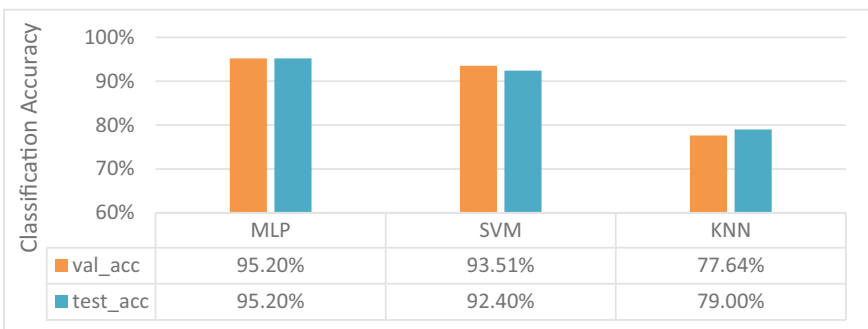


Fig. 5 Classification accuracy for the models

Table 1 Tested hyperparameters for each model

Model	All tested parameters
MLP	'hidden_layer_sizes': [(50,50,50),(50,100,50),(100,300,100)],
	'activation': ['tanh', 'relu'],
	'solver': ['sgd', 'adam'],
	'alpha': [0.0001, 0.05],
	'learning_rate': ['constant', 'adaptive']
SVM	'C': [1,10,20,30], 'kernel': ['rbf', 'linear']
KNN	'n_neighbors': [4,5,6]

2.5 Classification

The k-Nearest Neighbor (kNN) supervised learning was used to appraise the evaluated features in the classification of the five-class vibration signals [13, 14]. kNN machine learning was chosen as it has been regarded as one of the basic models due to the small number of hyperparameter while multilayer perceptron (MLP) also known as NN and SVM are used as they are popular classifier based on previous studies [13]. The parameters of the machine learning models were evaluated based on an exhaustive Grid Search Cross Validation (GridSearchCV from scikit-learn) approach where each model parameters were specified in a dictionary form. The models would be fitted and the best model with the highest accuracy will be returned. A total of three models namely kNN, SVM and MLP were used in this study. The parameters included for the models were included in Table 1. The models were fitted with a five-fold cross validation technique. The total of 500 instances (100 instances for five class) were split randomly into 80% for training and 20% for testing.

3 Result and Discussion

The result of the research was shown in Fig. 5. The model with the best classification accuracy was the MLP model with a validation accuracy and test accuracy of 95.2%. The MLP model shown higher classification accuracy by 2.8% compared to SVM and 16.2% compared to kNN. Both the MLP and SVM models shown no signs of overfitting while the kNN model was overfitted. This was most likely due to the nature of the kNN model where the decision boundary was more rough and less defined as the number of hyperparameters was low. The hyperparameter for MLP that achieved the highest classification accuracy were Rectified Linear Unit (ReLU) as its activation function, alpha value of 0.05, constant learning rate, 'adam' solver, and (50,100,50) for the three hidden layers. The hyperparameter for SVM were C value of 1 and radial basis function kernel while the k value for kNN was 5. From Table 2, MLP takes the longest time to train and test the models compared to SVM

Table 2 Computational time and best parameters of models

Model	Training time (s)	Testing time (s)	Best parameters
MLP	684.88187	2.6818292	'activation': 'relu', 'alpha': 0.05, 'hidden_layer_sizes': (50, 100, 50), 'learning_rate': 'constant', 'solver': 'adam'
SVM	9.327976	0.232331	'C': 1, 'kernel': 'rbf'
KNN	4.125072	0.0508382	'n_neighbors': 5

and kNN models. The total training time for MLP was around 685 s, 9 s for SVM and 4 s for kNN. The MLP took a significantly longer time to train the model compared to the other models, but the testing time was relatively short where it will be more suitable in an industrial setting.

4 Conclusion

In this study, the statistical features of the vibrational signals collection from the vibration sensor were extracted. In addition to that, the extracted features were selected using hypothesis testing which reduced the number of features from 2337 to 544. Three machine learning models were investigated and tested. The MLP model achieved the best classification accuracy compared to the other tested model. With a small amount of classification accuracy increase, the amount of time that MLP need was far longer due to the large amount of calculation. It is worth noting that the study is still in its preliminary stage as the study is a novel technique and the limited amount of previous study in the same scope. Hence, future studies shall investigate the effect of other feature selection techniques and classifiers as well as its hyperparameters.

Acknowledgements The authors would like to thank TT Vision Holdings Berhad for providing the Balluff sensor and OMRON TM Cobot to make this evaluation possible as well as for funding the study in collaboration with Universiti Malaysia Pahang via UIC200817 and RDU202406.


References

- Chiarello F, Trivelli L, Bonaccorsi A, Fantoni G (2018) Extracting and mapping industry 4.0 technologies using wikipedia. *Comput Ind* 100:244–257. <https://doi.org/10.1016/j.compind.2018.04.006>
- Wang S, Wan J, Zhang D, Li D, Zhang, C (2016) Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Comput Networks* 101:158–168. <https://doi.org/10.1016/j.comnet.2015.12.017>
- Bullon J, González Arrieta A, Hernández Encinas A, Queiruga Dios A (2017) Manufacturing processes in the textile industry. *Expert systems for fabrics production. ADCAIJ Adv Distrib Comput Artif Intell J* 6:15–23. <https://doi.org/10.14201/ADCAIJ2017641523>

4. El Zaatari S, Marei M, Li W, Usman Z (2019) Cobot programming for collaborative industrial tasks: an overview. *Rob Auton Syst* 116:162–180. <https://doi.org/10.1016/j.robot.2019.03.003>
5. Zekveld M, Hancke GP (2018) Vibration condition monitoring using machine learning. In: *Proceedings IECON 2018 - 44th annual conference IEEE industrial electronics society*, vol 1, pp 4742–4747. <https://doi.org/10.1109/IECON.2018.8591167>
6. Heo YJ, Kim D, Lee W, Kim H, Park J, Chung WK (2019) Collision detection for industrial collaborative robots: a deep learning approach. *IEEE Robot Autom Lett.* 4:740–746. <https://doi.org/10.1109/LRA.2019.2893400>
7. Min F, Wang G, Liu N (2019) Collision detection and identification on robot manipulators based on vibration analysis. *Sensors (Switzerland)*. 19:1–17. <https://doi.org/10.3390/s19051080>
8. Sharp M (2019) Observations on developing reliability information utilization in a manufacturing environment with case study: robotic arm manipulators. *Int J Adv Manuf Technol* 102:3243–3264. <https://doi.org/10.1007/s00170-018-03263-z>
9. Adam A, Zainal Abidin AF, Md Yusof Z, Mokhtar N, Shapiai MI (2019) A review on time-domain peak detection and classification algorithms for electroencephalogram signals. *Mekatronika* 1:115–121. <https://doi.org/10.15282/mekatronika.v1i2.4995>
10. Christ M, Braun N, Neuffer J, Kempa-Liehr AW (2018) Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a Python package). *Neurocomputing* 307:72–77. <https://doi.org/10.1016/j.neucom.2018.03.067>
11. Christ M, Kempa-Liehr AW, Feindt M (2016) Distributed and parallel time series feature extraction for industrial big data applications
12. February R, Academy I (2001) The control of the false discovery rate in multiple testing under dependency by Yoav Benjamini 1 and Daniel Yekutieli 2. *Ann Stat* 29:1165–1188
13. Krishna DH, Pasha IA, Savithri TS (2016) Classification of EEG motor imagery multi class signals based on cross correlation. *Procedia Comput Sci* 85:490–495. <https://doi.org/10.1016/j.procs.2016.05.198>
14. Wang LL, Zheng WL, Ma HW, Lu BL (2016) Measuring sleep quality from EEG with machine learning approaches. In: *Proceedings international joint conference neural networks*, pp 905–912, October 2016. <https://doi.org/10.1109/IJCNN.2016.7727295>

A Cluster Analysis of Identifying Team and Individual Sports Athlete Based on Anthropometric, Health and Skill Related Components



Noor Aishah Kamarudin, Mohamad Razali Abdullah, Vijayamurugan Eswaramoorthi, Ahmad Bisyril Husin Musawi Maliki, Aina Munirah Ab Rasid, Anwar P. P. Abdul Majeed, Mohd Azraai Mohd Razman, and Rabiul Muazu Musa 

Abstract The purpose of this study is to identify the essential fitness attributes in an individual sport and team sport. A total number of 218 male competitive youth athletes aged 13 to 21 years old were examined anthropometric, health and skill-related performance parameters (PP). Anthropometric parameters namely, weight, height, sitting height, and arm span were tested while the health-related parameters consisting of sit and reach, 1 min sit up, push up, handgrip, predicted VO_{2max} , and medicine ball throw were also collected. The 20 m speed, vertical jump, standing wide jump, stork stand test, and t-test represented the skill-related component tests. Based on the three components, hierarchical agglomerative cluster analysis (HACA) was utilized to group the athlete concerning their similarities in the PP examined. The athletes' performance within the clusters differs in three areas: age and arm span are significant in individual sports while height, weight, and sitting height are vital for team sports. Muscle endurance is shown to be essential in individual sports. Meanwhile, vertical jump and 20 m speed are more attributed to individual sports while standing broad jump, stork stand test, and t-test are more inclined to

N. A. Kamarudin · V. Eswaramoorthi
Faculty of Health Sciences, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia

M. R. Abdullah
East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia

A. B. H. M. Maliki
Academy of Defence Fitness, National Defence University of Malaysia, Kem Sungai Besi, 57000 Kuala Lumpur, Malaysia

A. M. Ab Rasid · R. M. Musa (✉)
Centre for Fundamental and Continuing Education, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
e-mail: rabiul.muazu@umt.edu.my

A. P. P. A. Majeed · M. A. M. Razman
Innovative Manufacturing, Mechatronics and Sports Laboratory, Universiti Malaysia Pahang, 26600 Pekan, Pahang Darul Makmur, Malaysia

team sports. Some fundamental traits could determine the type of sport suitable for athletes. These traits should be prioritized for identifying team or individual sports athletes. The findings could aid coaches in making decisions about the types of sports that could best suit an athlete for better performance delivery.

Keywords Hierarchical Agglomerative Clustering · Physical fitness components · Performance parameter · Team sport · Individual sport · Youth athletes

1 Introduction

A team sport is considered as any type of sport in which athletes compete in-group against other athletes. During a competition, team members work together to achieve a shared objective of winning [1]. The team members create goals, make choices, communicate, handle conflict, and solve problems in a helpful and trusting environment. Examples of such teams' sports are basketball, volleyball, and rugby. On the other hand, individual sports are those in which an athlete competes against another athlete either directly or indirectly. To win a competition, an athlete must rely exclusively on his abilities, strength, and stamina. For example, in archery, shooting, or lawn bowls, the athlete makes his own decision, talks, and calms himself to achieve the desired result [2].

Physical fitness related components are the prerequisite for any sporting activity. The ability of an athlete to perform sports activities without feeling overwhelmed is referred to as physical fitness performance in sports [3]. Physical fitness is proven to play a significant part in contributing to athlete performance in sports [4]. Physical fitness is separated into two components: physical fitness linked to health as well as physical fitness associated with motor behaviour [5]. The health-related components include cardiovascular endurance (aerobic fitness), body composition, muscle strength, muscular endurance, and flexibility. The skill-related component consists of power, agility, speed, coordination, response time, and balance [6]. The possession of aforesaid components is found to play a significant role in determining athlete achievement, as evidenced by the findings of the previous study [7]. The study further revealed that a combination of several skills of each component of physical fitness could accelerate athletes' achievement in their choosing sports.

The nature of a particular sport could warrant the possession of certain components that is non-trivial in shaping and producing athletes who can play in a team or individual sports. Hence, the current investigation is aimed at identifying the relevant essential parameter that could discriminate against an athlete in a team or individual sports through the application of multivariate analysis.

2 Methodology

2.1 Participants

The study's subjects are Terengganu junior athletes aged 13 to 21 who participated in the SUKMA 2020 games (Nationally organized competitive games). A total number of eight different sports with five consisting of individual sports and three different team sports were examined in this study. The athlete's overall physical fitness test was which included anthropometric, health and skill-related components were carried out. Before the commencement of the data collection, the purpose of the study was communicated to the council, coaches and all the relevant stakeholders. All of the athletes signed consent papers, which were collected in writing. The Terengganu Sports Institute (ISNT Research)'s Ethics Board examined and approved all of the study's protocols, and equipment, with the approval number MSN.TR.15 (22).

2.2 Anthropometric Evaluation

Weight, height, sitting height, and arm span were measured following the standard anthropometric assessment. A wooden stadiometer set on the wall was used to measure standing height to the nearest 0.5 cm. A standardized electronic digital scale was used to determine body weight to the closest 0.01 kg. The height of the seated buttocks was measured from the vertex of the head to the seated buttocks, and the values were recorded to the nearest 0.5 cm. The individual's sternal notch (centre of the breastbone) to their middle finger as it is stretched out to one side was measured with tape to determine the arm span of the athletes.

2.3 Health-Related Fitness Component

Standard physical fitness evaluations were used in conjunction with routine physical fitness indicators such as sit and reach, 1 min sit up, push up, hand-grip test, predicted VO_{2max} , and medicine ball throw test [8–10]. Before the testing periods, the participants conducted a warm-up that included a 5-to-10-min jog and a series of stretches. The athletes sit on the ground with straight legs, knees on the ground, and feet facing the flat surface of the sit and reach box when conducting sit and reach. The assessor holds the participant's sides of the knees straight against the floor with light pressure. The athletes slowly pushed forward their fingertips while sliding the measuring slide as far as possible across the measuring line with their hands on top of each other. The athletes sustained the position for at least two seconds while recording the distance [11]. The researchers ensured that the athletes' fingertips and legs are in a fixed straight position and that they are not making any jerky

movement. The test results were taken in 0.5 cm increments. The upper muscular strength and core muscle strength were measured using a mat spread on the floor during the push-up and sit-up tests. The athletes performed the test once for each test, within 1 min. A typically adjustable grip strength dynamometer was used to assess the handgrip strength of both hands. The athletes are permitted to stand up straight with the shoulder in adduction and neutral rotation and the elbow fully extended. The predicted $VO_{2\max}$ was measured using a multistage 20-m shuttle run test for the prediction of maximum oxygen uptake [12]. The athlete continued to run for as long as they could until they could no longer keep up with the tape's velocity. All the athlete's test results were expressed as an estimated $VO_{2\max}$, which was determined by examining the last level and shuttle number at the time the individual voluntarily surrender from the test. The overhead throw for distance is a test of upper body strength that involves throwing the ball forwards from over the head. Medicine ball throw tests are used to assess the power of the athletes.

2.4 Skill Related Fitness Component

The 20-m sprint, vertical jump test, standing stork test, standing broad jump test, and T-test are all skill-related fitness component tests. At a 20 m speed, the athlete sprints a set distance while maintaining track of the time. The length between the starting and finishing lines was 20 m [13]. When either foot passed the starting point and either foot approached the endpoint, the period began to be counted [14]. Before beginning to run, the athlete must maintain a fixed stance, starting from a set position and returning the foot to the starting line without making any motions. The assessor used a stopwatch to record the duration of the athletes' run from the starting line to the finish line. Vertec was used to measure the vertical jump ability of the athletes. The height of the colour-coded plastic vanes was modified so that it was in line with the athletes' standing height. The athlete flexed their ankles, knees, and hips, swung their arms upward, and tapped the highest possible vane with their dominant hand's fingers. For statistical analysis, the best of the three trials was chosen. The standing stork test was timed with a stopwatch. The athletes dropped their shoes, placed their hands on their hips, raised their heels to balance on the ball of their foot, and placed their non-supporting foot on the inner part of the supporting leg. The test is terminated if the athletes are unable to maintain the exact position. A standing broad jump requires participants to stand behind a line placed in the area. The athletes are required to land with their arms swinging back and their knees bent rhythmically at roughly 90° forward after a two-foot take-off jump. The athlete must land with both feet and jump as far as possible without falling backwards. If the participant does not follow the procedure, the test will be discarded [15]. Three trials were allowed, with the highest being recorded for analysis. The 'T-test agility test' was used to determine the agility of the athletes. In this test pointers are placed 10-5-5 m apart on a line drawn on the ground in the shape of a 'T.' The participants run from the 10 m marker near the line, through the 5 m markers, turn on the line, and sprint back

over the 5 m markers. The time is recorded using an infrared speed trap from when the athlete first runs through the 5 m back to 10 m markers until they halt. Each athlete attempted two times at their maximum, with the fastest time being recorded for analysis. The participant was advised not to cross the line too far because it would lengthen their time.

2.5 Hierarchical Agglomerative Clustering Analysis (HACA)

HACA (Hierarchical Agglomerative Cluster Analysis) is an unsupervised exploratory method that creates a hierarchy of clusters based on a single observation, consolidating several similar observations into a single observation [16]. The algorithm merges and splits datasets which are commonly done greedily so that related observations are segregated and displayed in a dendrogram [17]. It is worth noting that the dendrogram in HACA displays the number of clusters based on the proximity calculated by a specific or pre-defined cluster. In this investigation, the Cosine's distance was used as the clustering validation approach, and class centroids were used as the clustering validation technique [18]. The data analysis for this study was accomplished using the Orange statistical program version 3.2 for Windows.

3 Results and Discussion

Table 1 shows the descriptive statistics of anthropometric, health, and skill-related fitness characteristics among 218 male athletes examined in the study. The highest and the lowest scores, mean and standard deviations were also tabulated.

In the analysis process, the mean value of each parameter was utilized. This is the type of "average" that gives a better picture of any general tendency in the data when the data the sample are relatively large. Moreover, the mean value is calculated by multiplying all the values by the total number of participants. This strategy works well if the data involves a greater number of observations. It could be seen from the table that the clusters of the athletes in both teams and individual sports have a different score for each performance parameter examined in the study.

Figure 1 shows the study's clustering of the data based on anthropometric performance parameters examined in the study. The clustering technique was able to reasonably divide the athletes into two distinctive groups viz. individual and team sports athletes as shown in the figure. It could be seen that the athletes' performance in individual and team sports differs based on the anthropometric components. The significant parameter for the Individual sports revealed age and arm span, whereas height, weight, and seating height are demonstrated as essential for team sports athletes.

Table 1 Descriptive statistics of individual and team sports athletes

Variable		Types of sports	Min	Max	Mean	Standard deviation
Anthropometric	Age	Individual	13	21	18.44	1.43
		Team	14	20	17.31	2.19
	Weight (kg)	Individual	29.4	165	69.87	18.81
		Team	47.9	129	62.49	15.65
	Height (cm)	Individual	143	182.2	170.1	7.6
		Team	154.9	187	166.9	7.55
	Sitting height (cm)	Individual	35.9	92.2	170.1	5.8
		Team	68.1	104	166.9	6.13
	Arm span (cm)	Individual	140.7	196.5	152.6	33.56
		Team	61.5	185.4	171.7	10.02
Health related	Sit and reach (cm)	Individual	20.3	58	39.17	4.44
		Team	31.2	49.5	39.06	7.61
	1 min sit up (max rep)	Individual	9	80	41.76	6.99
		Team	28	54	39.78	11.91
	Push up (max rep)	Individual	2	203	38.32	15.6
		Team	20	86	59.13	36.89
	Handgrip (kg)	Individual	18.6	62.6	45.06	7.5
		Team	31.7	68.4	42.51	9.06
	Predicted VO 2max	Individual	16	61	45.24	6.31
		Team	31	57	42.02	9.14
Medicine ball throw (m)	Individual	2.6	9.7	7.19	0.97	
	Team	5.75	9.65	6.22	1.45	
Skill related	20 m speed (sec)	Individual	2.78	4.79	3.17	0.24
		Team	2.35	3.8	3.41	0.37
	Vertical jump (inches)	Individual	26.5	66	33.01	11.42
		Team	18	51	45.47	8.27
	Standing broad jump (cm)	Individual	116	271	233.7	32.59
		Team	150	289	213.5	30.02
	Stork stand test (sec)	Individual	0.1	175	29.47	32.79
		Team	1.31	139.6	28.39	30.11
	T test (sec)	Individual	9.1	15.31	11.13	0.87
		Team	10.06	14.15	11.45	1.08

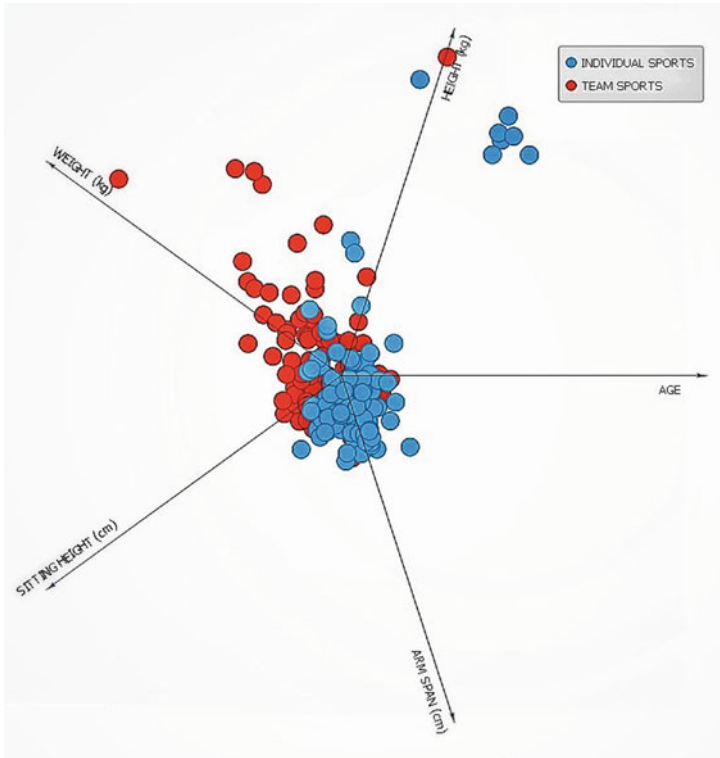


Fig. 1 Anthropometric components clusters for individual and team sports

Figure 2 depicts a health-related component plot for both individual and team sports clusters. Some differences were revealed in individual and team sports parameters. The individuals' sports athletes were more attributed to high flexibility and throwing ability whilst the team sports athletes were more inclined towards higher muscular as well as cardiovascular endurance.

According to the clustering in skill-related components depicted in Fig. 3, team and individual athletes have different abilities. The individual sports athletes recorded higher performances in vertical jump and 20 m speed demonstrating the requirement of leg power and speed for individual sports. On the other hand, the team sports athletes excelled in the performance of standing broad jump, stork stand test, as well as t-test signifying the need for explosive power, balance, and agility in team related sports.

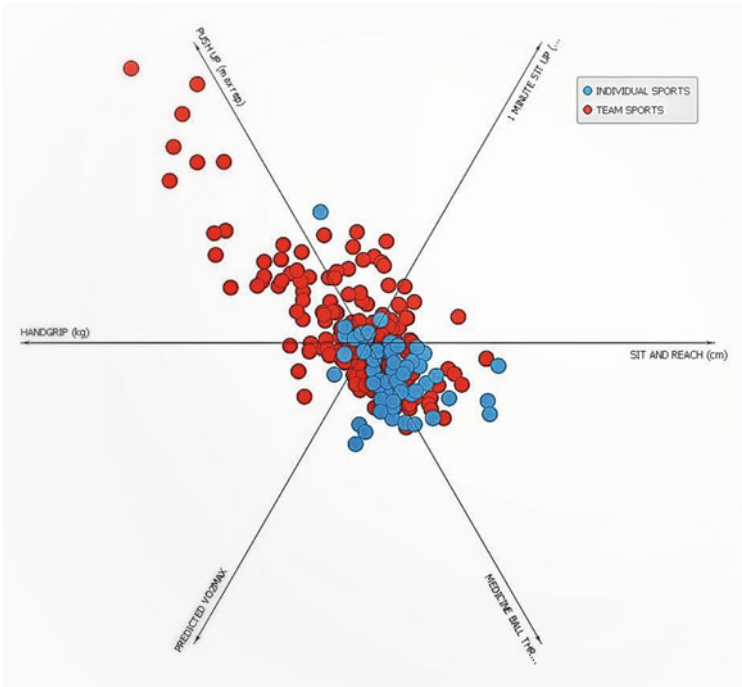


Fig. 2 Health-related fitness components clusters for individual and team sports

The clustering analysis approach can be used to establish athlete the essential performance parameters in individual and team sports. Thus, the segregation of the athletes into a team or individual sports demonstrated the importance of the performance parameters examined in the current study. Moreover, the application of HACA is pivotal towards the appropriate assignment of a group to the athletes involved in the study.

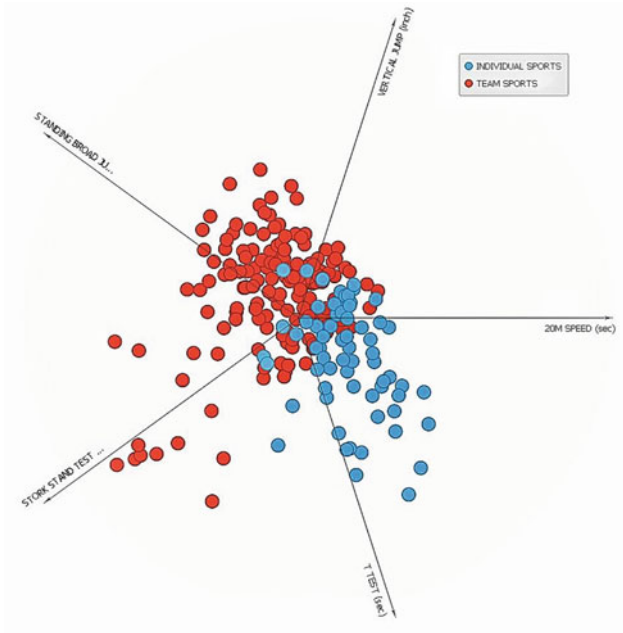


Fig. 3 Skill related fitness components clusters for individual and team sports

4 Conclusion

The current study has successfully established the vital performance measures that can differentiate between individual and team performance in sports using multivariate analysis. Performance parameters which include data from anthropometric components tests such as weight, height, sitting height, and arm span have been demonstrated as essential in discriminating athletes on the type of sport they are involved in. Sit and reach, 1 min sit up, push up, handgrip, projected VO_{2max} , and medicine ball throw as health-related components are also shown to be vital. Skill-related component tests such as the 20-m dash, vertical jump, standing broad jump, stork stand test, and t-test are not trivial in determining which sports an athlete may participate in. The multivariate clustering analysis technique utilized is useful in giving information on relevant performance factors that could enhance individual and team sports performance. The findings of this analysis may be helpful to coaches, team managers, and talent Scouters better understand the parameters that are suitable for both team and individual sports.

Acknowledgements The authors thank the athletes and all the relevant stakeholders for their cooperation throughout the data collection process.

References

1. Pfeffer J (2007) Human resources from an organizational behavior perspective: some paradoxes explained. *J Econ Perspect* 21(4):115–134
2. Nia ME, Ali Besharat M (2010) Comparison of athletes' personality characteristics in individual and team sports. *Procedia Soc Behav Sci* 5:808–812
3. Ortega F, Ruiz, J (2013) Physical activity, physical fitness, and overweight in children and adolescents: evidence from epidemiologic studies 60(8):458–469
4. Rauner A, Mess F, Woll A (2013) The relationship between physical activity, physical fitness and overweight in adolescents: a systematic review of studies published in or after 2000. *BMC Pediatrics* 13(1):1–9
5. Gratton C, Jones DI, Jones I (2014) Research methods for sports studies
6. Caspersen C, Powell K et al (1985) Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Ncbi Nlm Nih Gov* 100:195–202
7. Durandt J, Tee J et al (2006) Physical fitness components associated with performance in a multiple-sprint test. *J Hum Kinetics Com*
8. Noguchi T, Demura S, Physical KT-A et al (2013) Relationships between sit-ups and abdominal flexion strength tests and the thickness of each abdominal muscle. *File Scirp Org* 3(2):84–88
9. Fleishman E (1964) The structure and measurement of physical fitness
10. Taha Z et al (2018) The identification of high potential archers based on fitness and motor ability variables: a support vector machine approach 57:184–193
11. Taha Z et al (2018) The identification of high potential archers based on relative psychological coping skills variables: a support vector machine approach. *IOP Conf Ser Mater Sci Eng* 319(1):012027
12. Leger L, Mercier D et al (1988) The multistage 20 metre shuttle run test for aerobic fitness 6(2):93–101
13. Maliki A, Abdullah M et al (2018) A multilateral modelling of youth soccer performance index (YSPI). *Iopscience.Iop.Org*
14. Musa RM, Majeed APPA, Taha Z, Chang SW, Nasir AFA, Abdullah MR (2019) A machine learning approach of predicting high potential archers by means of physical fitness indicators. *PLoS One* 14(1):e0209638
15. Maliki A et al (2018) The role of anthropometric, growth and maturity index (AGaMI) influencing youth soccer relative performance. *Iopscience.Iop.Org*
16. Maimon O, Rokach L (2005) Data mining and knowledge discovery handbook. Springer, New York. <https://doi.org/10.1007/b107408>
17. Musa RM et al (2018) Supervised pattern recognition of archers' relative psychological coping skills as a component for a better archery performance. *J Fundam Appl Sci* 10(1S):467–484
18. Muazu Musa R, Abdul Majeed APP, Taha Z, Abdullah MR, Husin Musawi Maliki AB, Azura Kosni N (2019) The application of artificial neural network and k-nearest neighbour classification models in the scouting of high-performance archers from a selected fitness and motor skill performance parameters. *Sci Sports* 34(4):e241–e249

Vision-Based Human Presence Detection by Means of Transfer Learning Approach



Jin Cheng Tang, Ahmad Fakhri Ab. Nasir, Anwar P. P. Abdul Majeed, Mohd Azraai Mohd Razman, Ismail Mohd Khairuddin, and Thai Li Lim

Abstract Human–robot interaction (HRI) and human–robot collaboration (HRC) has become more important as the industries are stepping towards the phase of digitalization and Industry 4.0. Indeed, the emphasis is often placed on the safety of the physical well-being of the human workers. To safeguard the human operators from being hurt by the robots or collaborative robots (cobots), a traditional method is to isolate the robots from the human workers by means of fences and sensors. However, the deployment of deep learning models is unknown and shown to be non-trivial in downstream tasks such as image classification and object detection. The present study aimed to exploit the effectiveness of object detection models, particularly EfficientDet models via a transfer learning approach—fine-tuning. A total of 1463 images were obtained from the surveillance cameras from TT Vision Holdings Berhad and split into training, validation, and test by a ratio of 70:20:10. The training images were further augmented using horizontal flip and scale jittering techniques to increase the total training images up to 3072 images. As an outcome, the result revealed that the EfficientDet-D2 fine-tuned model achieved a test AP of 81.70% with an inference speed of 97.06 ms on Tesla T4 while the EfficientDet-D0 fine-tuned model attained a test AP of 69.30% with an inference speed of 30.24 ms on Tesla T4. In comparison between the EfficientDet-D0 fine-tuned model and EfficientDet-D2

J. C. Tang · A. P. P. Abdul Majeed · M. A. Mohd Razman · I. Mohd Khairuddin
Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia

A. F. Ab. Nasir (✉)
Faculty of Computing, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: afakhri@ump.edu.my

A. F. Ab. Nasir · A. P. P. Abdul Majeed
Centre for Software Development and Integrated Computing, Universiti Malaysia Pahang,
26600 Gambang, Pahang, Malaysia

T. L. Lim
TT Vision Holdings Berhad, 11900 Plot 106, Sungai Hilir Keluang 5,
Bayan Lepas FIZ.4, Bayan Lepas, Pulau Pinang, Malaysia

A. P. P. Abdul Majeed
School of Robotics, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool
University, Suzhou 215123, P. R. China

fine-tuned model, the performance improved in terms of AP with the inference speed as the trade-off. The research has shown that it is feasible to detect the presence of human workers and can possibly serve as the visual perception of the robot with regards to human presence detection. Last but not least, the present work proved the applicability of transfer learning methods in human presence detection, specifically fine-tuned object detection models.

Keywords Deep learning · EfficientDet · Transfer learning · Fine-tuning · Human detection

1 Introduction

The terms human–robot interaction (HRI) and human–robot collaboration (HRC) have been introduced into the field and have received an enormous amount of attention from manufacturing enterprises [1]. According to the International Federation of Robotics (IFR), the trend has revealed that the adoption of HRC is escalating, in which the cobot installation has risen by 11% in 2019 [2]. Although this market is still in the infancy phase, HRC has been considered as one of the essential pillars of Industry 4.0 [3]. Among the ASEAN countries, Malaysia is graded as a leading archetype country in terms of readiness for the future of production assessment [4]. HRC is perhaps a fantasy combination of machine productivity and human flexibility, but problems arise when this combination means withdrawing from the conventional paradigm of robots keeping in an isolated workspace within the safety fences. Thereafter, occupational safety has been rated as the most crucial factor to succeed HRI in the future [5].

Research on improving human safety for HRC and HRI has a long tradition. Handful of research has been conducted in relation to the safety aspects of HRC [6]. It is noted that vision-based systems emerged as a famous approach over years, probably due to the uniqueness of this sense. In connection with artificial intelligence, deep learning and transfer learning have been employed in various applications such as tomato classification [7] and electroencephalogram (EEG) signals classification [8]. However, the combination of vision sensors and deep learning object detection approach is rarely seen in previous studies. Mohammed et al. [9] reported an augmented environment-based collision avoidance approach via two Kinect depth cameras by employing a motion feature-based method to simulate the location of the human operators and robots for the implementation of appropriate safety strategies.

Heo et al. [10] claimed that it is the first real-time collision detection implementation that utilized a deep neural network (DNN) and demonstrated solid accuracy with an average delay of only 18.4 ms. Mohammadi Amin et al. [11] combined both visual perception and tactile perception, where a 3D-CNN is used for human action recognition while a 1D-CNN is used for contact detection. It could be clearly seen from previous studies that the exploitation of object detection algorithms in HRC and HRI is somehow limited. On that account, this study aimed to explore the applicability of the object detection models, particularly EfficientDet models, and examine the performance of the models in human presence detection.

2 Methodology

2.1 Image Acquisition

Based on the surveillance cameras in TT Vision Holdings Berhad, the recorded video footage from the database was used for the purpose of image dataset acquisition. The surveillance cameras of the production areas were selected instead of other locations such as the staircase, laboratory, and the office area due to the fact that the people around the production line area tend to move around and appear in different spots of the surveillance cameras, subsequently introducing variability to the dataset. A total of 1463 images were collected on top of the recorded video footage. The dataset was separated into three groups with respect to train, validation, and test by a ratio of 70:20:10. Besides, several data augmentation techniques were used to bring in more variation to the training dataset and increase the size of the dataset. For instance, scale jittering and conventional horizontal flipping were used as the data augmentation techniques. In total, 3072 images were fed into the models during the fine-tuning phase as the training images (Fig. 1).



Fig. 1 Examples of image used to train the model

2.2 *Image Annotation*

The workers appearing in the images were annotated with a bounding box and labeled as “Person”. The occluded body part of the workers was not annotated since the object detection models only intended to detect the presence of human workers in the scene. In specific, a well-known free annotation tool called LabelIMG [12] was used to annotate the images. The annotation files were then outputted in PASCAL VOC format [13] for the following step.

2.3 *TensorFlow Records (TFRecords) Generation and Label Map Preparation*

A deep learning framework named TensorFlow [14] was utilized for training the object detection models. Hence, TFRecords files were generated as this format can be understood by the TensorFlow library. In particular, the TFRecord file is a serialized representation of the entire dataset which contains all images and annotations. A script was used to convert the PASCAL VOC format annotation files and the respected images into TFRecord files. A label map file was created in this case to define a mapping between the labels and the integer class IDs. In this study, one class ID is sufficient for the label map file since only the “person” class was involved.

2.4 *Transfer Learning—Fine Tuning*

TensorFlow Object Detection API [15] was employed in this study, specifically the EfficientDet models in the TensorFlow 2 Detection Model Zoo for human presence detection. The paradigm of the employed models is considered as a one-stage detector. Descriptively, EfficientNet [16] that is pre-trained on the COCO 2017 detection datasets [17] was used as the backbone network. As the neck, a weighted bi-directional feature pyramid network [18] was used in which different layers of semantic information are integrated. For the detection head of the models, the single-shot detector [19] convolutional head was applied to predict the location and the class of the objects with corresponding annotated bounding boxes.

The benefits of transfer learning over training from scratch are that it can save lots of training time and does not need a huge amount of training data [20]. In the present investigation, EfficientDet models were employed as the transfer learning models with only the feature extractor being restored. In consistently, the COCO pre-trained weights were transferred to the fine-tuned EfficientDet models for feature extraction. It is noted that the present study had a different detection task as compared to the pre-trained EfficientDet models, therefore the head of the models was not restored as

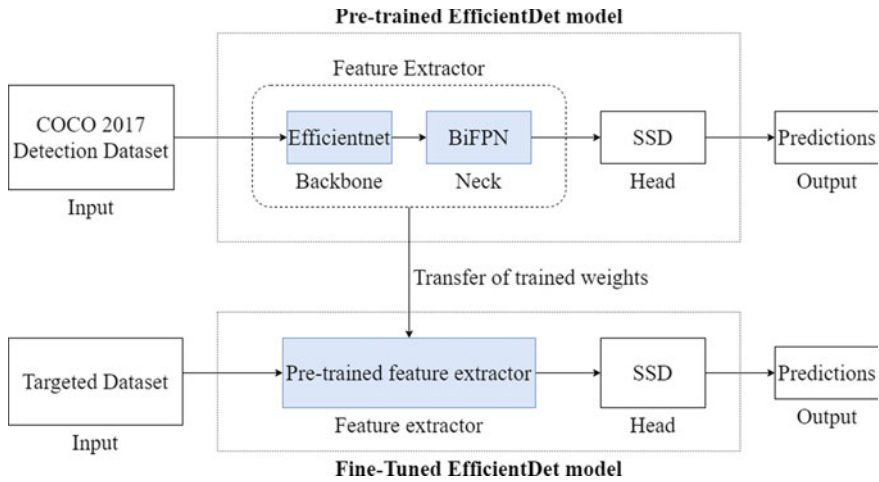


Fig. 2 EfficientDet transfer learning pipeline by means of fine-tuning

Table 1 List of EfficientDet transfer learning models and corresponding network configuration

Transfer learning models	Input image size	Backbone network	BiFPN		Prediction network
			Depth	Width	Depth
Efficientdet-D0	512 × 512	EfficientNet-B0	3	64	3
Efficientdet-D2	768 × 768	EfficientNet-B2	5	112	3

it was required to fine tune the prediction head for the specific task—human presence detection (Fig. 2 and Table 1).

2.5 Performance Evaluation

The mainstream evaluation metrics for the performance of the object detection models is the Average Precision (AP). In computing the Average Precision (AP) for the object detection models, first have to understand the concept of Intersection over Union (IoU). The calculation of IoU is expressed and illustrated as in the figure below (Fig. 3).

Eventually, an IoU threshold was established to justify whether the predicted bounding box is classified as correct detection or false detection. In the present study, an IoU threshold value of 0.5 was used to account for the detection evaluation. In accordance, a precision-recall curve can be depicted, and the general definition of Average Precision (AP) is to determine the area under the precision-recall curve. The formula for AP can be expressed as below:

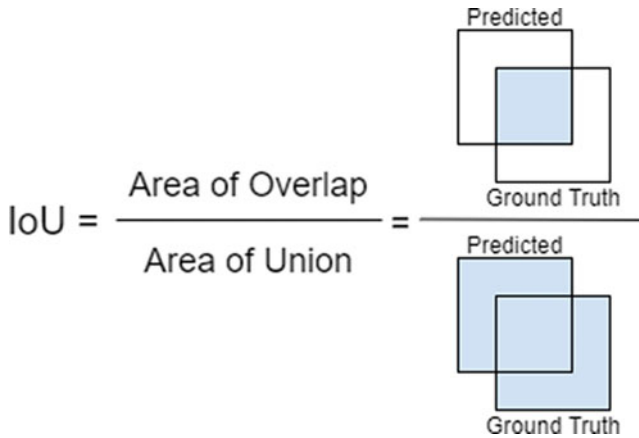


Fig. 3 The computation of IoU by using the predicted bounding box and the ground truth box

$$AP = \int_0^1 p(r)dr \quad (1)$$

3 Result and Discussion

The EfficientDet transfer learning models were trained with a learning rate of 0.01 for a warmup step of 2.5 k and a learning rate of 0.001 for a total training step of 100 k. One of the ways to diagnose the behavior of the models is perhaps through observing the learning curves. Both training loss and validation loss were computed and monitored during the training to prevent the occurrence of overfitting. The learning curves of the transfer learning models are depicted as in Figs. 4 and 5. From the learning curves of the transfer learning models, it can be observed that the learning rate of the models decreased over the learning steps and eventually converged. Nevertheless, the gaps between the training loss and the validation loss were made sure to be sufficiently small and stable over the learning phase. Noting that as seen in the training learning loss of the EfficientDet-D2 fine-tuned model, the initial training loss of the model is suspiciously high at the first 10 k steps due to the 0.01 warmup learning rate. But as long as the training loss decreased afterwards and converged over the learning steps, then this unreasonably high loss can be ignored.

To further evaluate the performance of the fine-tuned models, the evaluation metric known as AP was computed with the train, validation, and test datasets. The AP of different fine-tuned models are shown in Fig. 6. The test AP was reported to be lower than the train AP, indicating that the fine-tuned models do not overfit to the dataset and can generalize well to the data that the models have not seen before.

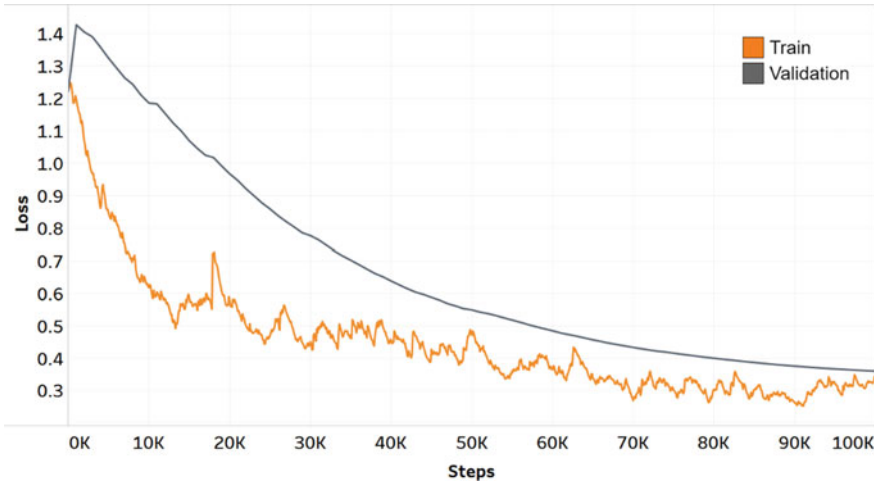


Fig. 4 The train and validation learning curve of EfficientDet-D0 fine-tuned model

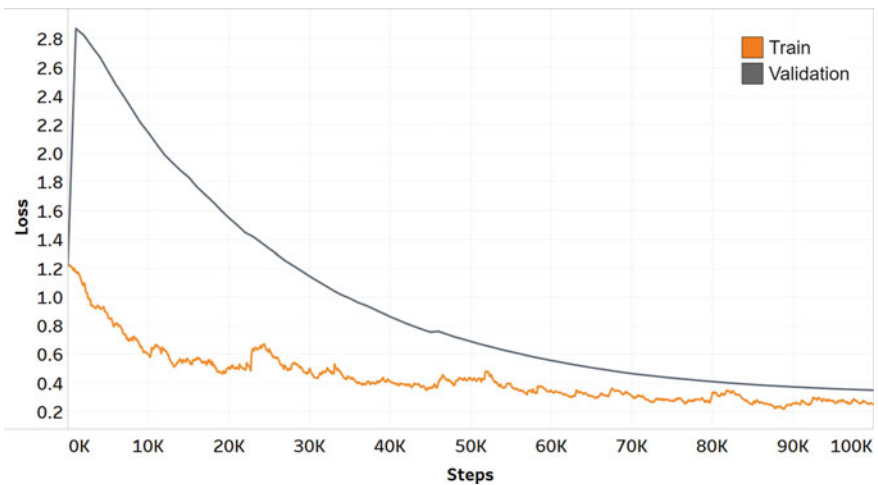


Fig. 5 The train and validation learning curve of EfficientDet-D2 fine-tuned model

Another finding was that the train, validation, and test APs of EfficientDet-D2 fine-tuned model improved as compared to EfficientDet-D0 fine-tuned model. This can be explained by the configuration of the models with regards to the input image size, depth, and width of the BiFPN. It is interpreted that increasing these parameters can improve the models in learning features with higher semantic meaning. Nevertheless, the increase of the parameters was based on the compound scaling method as in [18].

But as a trade-off to the greater AP by means of greater depth and width of BiFPN and the input image size, the fine-tuned model became heavier to load and had slower

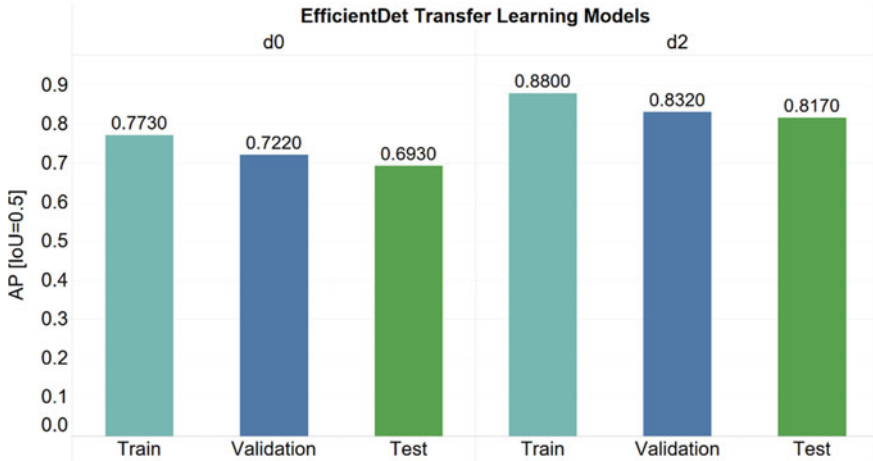


Fig. 6 Train, validation, and test AP metrics of the EfficientDet Fine-tuned models

Table 2 Performance of EfficientDet fine-tuned models based on number of parameters, size of the fine-tuned models and inference speed on Tesla T4

EfficientDet fine-tuned models	Fine-tuned model size (MB)	Number of parameters	Ratio	Inference time on Tesla T4 (ms)	Ratio
Efficientdet-D0	25.5 MB	5.57 M	1×	30.24	1×
Efficientdet-D2	35.9 MB	10.0 M	1.79×	97.06	3.21×

inference speed. The specification of the fine-tuned models in terms of model size and number of parameters is summarized in Table 2. Results indicate that the slower inference speed is most probably due to the higher number of parameters in the fine-tuned models and larger model size.

4 Conclusion

In this study, the recorded video footage was used to obtain the desired image datasets and the persons appearing in the images are annotated with bounding boxes and labels. Consequently, a label map was prepared according to the format required by TensorFlow and TFRecord files were generated from the image datasets to serve as the input images for the models. A transfer learning method known as fine-tuning was used in the present study where the COCO pre-trained weights were transferred, and the head detector was fine-tuned with the custom image dataset. The results demonstrated that, the EfficientDet-D2 fine-tuned model obtained a better accuracy than the EfficientDet-D0 fine-tuned model in terms of AP by sacrificing an adequate amount of inference speed. $n_{D_{bifpn}}$

Acknowledgements The authors would like to thank TT Vision Holdings Berhad for providing the image dataset to make this evaluation possible as well as for funding the study in collaboration with Universiti Malaysia Pahang via UIC200816 and RDU202405.

References

1. Magrini E, Ferraguti F, Ronga AJ, Pini F, De Luca A, Leali F (2020) Human-robot coexistence and interaction in open industrial cells. *Robot Comput Integr Manuf* 61:101846. <https://doi.org/10.1016/j.rcim.2019.101846>
2. International Federation of Robotics (2020) World Robotics Report 2019. https://ifr.org/downloads/press2018/2020-09-24_IFR_press_release_WR_industrial_robots.pdf
3. Evjemo LD, Gjerstad T, Grøtli EI, Sziebig G (2020) Trends in smart manufacturing: role of humans and industrial robots in smart factories. *Curr Robot Reports* 1(2):35–41. <https://doi.org/10.1007/s43154-020-00006-5>
4. World Economic Forum (2018) Insight Report: Readiness for the Future of Production Report 2018. http://www3.weforum.org/docs/FOP_Readiness_Report_2018.pdf
5. Kopp T, Baumgartner M, Kinkel S (2021) Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. *Int J Adv Manuf Technol* 112(3–4):685–704. <https://doi.org/10.1007/s00170-020-06398-0>
6. Matheson E, Minto R, Zampieri EGG, Faccio M, Rosati G (2019) Human–robot collaboration in manufacturing applications: a review. *Robotics* 8(4):100. <https://doi.org/10.3390/robotics8040100>
7. Toon OP, Zakaria MA, Nasir AFA, Majeed APP, Tan CY, Ng LCY (2019) Autonomous tomato harvesting robotic system in greenhouses: deep learning classification. *Mekatronika* 1(1):80–86. <https://doi.org/10.15282/mekatronika.v1i1.1148>
8. Kumar JM et al (2020) An evaluation of different fast fourier transform - transfer learning pipelines for the classification of wink-based EEG signals.pdf. *Mekatronika* 2(1):1–7. <https://doi.org/10.15282/mekatronika.v2i1.4881>
9. Mohammed A, Schmidt B, Wang L (2017) Active collision avoidance for human–robot collaboration driven by vision sensors. *Int J Comput Integr Manuf* 30(9):970–980. <https://doi.org/10.1080/0951192X.2016.1268269>
10. Heo YJ, Kim D, Lee W, Kim H, Park J, Chung WK (2019) Collision detection for industrial collaborative robots: a deep learning approach. *IEEE Robot Autom Lett* 4(2):740–746. <https://doi.org/10.1109/LRA.2019.2893400>
11. Amin FM, Rezayati M, van de Venn HW, Karimpour H (2020) A mixed-perception approach for safe human-robot collaboration in industrial automation. *Sensors* 20(21):6347. <https://doi.org/10.3390/s20216347>
12. Ta LT (2015) LabelImg. Git code. Accessed 28 June 2021, <https://github.com/tzutalin/labelImg>
13. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2010) The pascal visual object classes (VOC) challenge. *Int J Comput Vis* 88(2):303–338. <https://doi.org/10.1007/s11263-009-0275-4>
14. Abadi M, et al (2016) TensorFlow: large-scale machine learning on heterogeneous distributed systems. <http://arxiv.org/abs/1603.04467>
15. Huang J, Sun C, Murphy K, Guadarrama S (2017) Speed accuracy trade-offs for modern convolutional object detectors: supplementary Materials. In: *Cvpr*, pp 7310–7319
16. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: 36th International Conference on Machine Learning, ICML 2019, vol 2019-June, pp 10691–10700. <http://arxiv.org/abs/1905.11946>
17. Lin TY, et al (2014) Microsoft COCO: common objects in context. In: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol 8693 LNCS, no PART 5, pp 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

18. Tan M, Pang R, Le QV (2020) EfficientDet: scalable and efficient object detection. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 10778–10787 (2020). <https://doi.org/10.1109/CVPR42600.2020.01079>
19. Liu W, et al (2016) SSD: single shot multibox detector. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 9905 LNCS, pp 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
20. Zhuang F et al (2021) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76. <https://doi.org/10.1109/JPROC.2020.3004555>