



# **ICAS 2022**

The Eighteenth International Conference on Autonomic and Autonomous Systems

ISBN: 978-1-61208-966-9

May 22nd –26th, 2022

Venice, Italy

## **ICAS 2022 Editors**

Claudius Stern, FOM University of Applied Sciences, Germany

Cosmin Dini, IARIA, USA

# ICAS 2022

## Foreword

The Eighteenth International Conference on Autonomic and Autonomous Systems (ICAS 2022), held between May 22 – 26, 2022, was a multi-track event covering related topics on theory and practice on systems automation, autonomous systems and autonomic computing.

The main tracks referred to the general concepts of systems automation, and methodologies and techniques for designing, implementing and deploying autonomous systems. The next tracks developed around design and deployment of context-aware networks, services and applications, and the design and management of self-behavioral networks and services. We also considered monitoring, control, and management of autonomous self-aware and context-aware systems and topics dedicated to specific autonomous entities, namely, satellite systems, nomadic code systems, mobile networks, and robots. It has been recognized that modeling (in all forms this activity is known) is the fundamental for autonomous subsystems, as both managed and management entities must communicate and understand each other. Small-scale and large-scale virtualization and model-driven architecture, as well as management challenges in such architectures are considered. Autonomic features and autonomy requires a fundamental theory behind and solid control mechanisms. These topics gave credit to specific advanced practical and theoretical aspects that allow subsystem to expose complex behavior. We aimed to expose specific advancements on theory and tool in supporting advanced autonomous systems. Domain case studies (policy, mobility, survivability, privacy, etc.) and specific technology (wireless, wireline, optical, e-commerce, banking, etc.) case studies were targeted. A special track on mobile environments was indented to cover examples and aspects from mobile systems, networks, codes, and robotics.

Pervasive services and mobile computing are emerging as the next computing paradigm in which infrastructure and services are seamlessly available anywhere, anytime, and in any format. This move to a mobile and pervasive environment raises new opportunities and demands on the underlying systems. In particular, they need to be adaptive, self-adaptive, and context-aware.

Adaptive and self-management context-aware systems are difficult to create, they must be able to understand context information and dynamically change their behavior at runtime according to the context. Context information can include the user location, his preferences, his activities, the environmental conditions and the availability of computing and communication resources. Dynamic reconfiguration of the context-aware systems can generate inconsistencies as well as integrity problems, and combinatorial explosion of possible variants of these systems with a high degree of variability can introduce great complexity.

Traditionally, user interface design is a knowledge-intensive task complying with specific domains, yet being user friendly. Besides operational requirements, design recommendations refer to standards of the application domain or corporate guidelines.

Commonly, there is a set of general user interface guidelines; the challenge is due to a need for cross-team expertise. Required knowledge differs from one application domain to another, and the core knowledge is subject to constant changes and to individual perception and skills.

Passive approaches allow designers to initiate the search for information in a knowledge-database to make accessible the design information for designers during the design process. Active approaches, e.g., constraints and critics, have been also developed and tested. These mechanisms deliver information (critics) or restrict the design space (constraints) actively, according to the rules and

guidelines. Active and passive approaches are usually combined to capture a useful user interface design.

We take here the opportunity to warmly thank all the members of the ICAS 2022 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICAS 2022. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICAS 2022 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICAS 2022 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of autonomic and autonomous systems.

We are convinced that the participants found the event useful and communications very open. We also hope that Venice provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

#### **ICAS 2022 Chairs:**

##### **ICAS 2022 Steering Committee**

Roy Sterritt, Ulster University, UK

Mark J. Balas, Texas A&M University, USA

Radu Calinescu, University of York, UK

Karsten Böhm, Fachhochschule Kufstein, Austria

Jacques Malenfant, Sorbonne Université | LIP6 Lab, France

Claudius Stern, biozoom services GmbH - Kassel | FOM University of Applied Sciences – Essen, Germany

Petr Skobelev, Knowledge Genesis Group / Samara Technical University, Russia

##### **ICAS 2022 Publicity Chairs**

Hannah Russell, Universitat Politècnica de València (UPV), Spain

Mar Parra, Universitat Politecnica de Valencia, Spain

## ICAS 2022

### Committee

#### ICAS 2022 Steering Committee

Roy Sterritt, Ulster University, UK  
Mark J. Balas, Texas A&M University, USA  
Radu Calinescu, University of York, UK  
Karsten Böhm, Fachhochschule Kufstein, Austria  
Jacques Malenfant, Sorbonne Université | LIP6 Lab, France  
Claudius Stern, biozoom services GmbH - Kassel | FOM University of Applied Sciences – Essen, Germany  
Petr Skobelev, Knowledge Genesis Group / Samara Technical University, Russia

#### ICAS 2022 Publicity Chairs

Hannah Russell, Universitat Politècnica de València (UPV), Spain  
Mar Parra, Universitat Politecnica de Valencia, Spain

#### ICAS 2022 Technical Program Committee

Mubarak Abdu-Aguye, Mohamed bin Zayed University of Artificial Intelligence, UAE  
Lounis Adouane, Université de Technologie de Compiègne (UTC), France  
Waseem Akram, University of Calabria, Italy  
Alba Amato, Institute for High-Performance Computing and Networking (ICAR), Napoli, Italy  
Vijayan K. Asari, University of Dayton, USA  
Mark Balas, Texas A&M University, USA  
Malek Ben Salem, Accenture Labs, USA  
Julita Bermejo-Alonso, Universidad Politécnica de Madrid (UPM), Spain  
Chiara Bersani, Polytechnic School University of Genova, Italy  
Navneet Bhalla, University College London, UK  
Estela Bicho, University of Minho / Centre Algoritmi / CAR group, Portugal  
Karsten Böhm, Fachhochschule Kufstein, Austria  
Kamel Bouzgou, Université des sciences et de la technologie Oran USTO-MB, Algeria / Université Paris Saclay - Univ. Evry, France  
Kenny Bowers, Georgia Tech Research Institute, USA  
Estelle Bretagne, University of Picardie Jules Verne / lab MIS (modeling, information and systems), France  
Ivan Buzurovic, Harvard University, Boston, USA  
Radu Calinescu, University of York, UK  
Paolo Campegiani, Bit4id, Italy  
Valérie Camps, Paul Sabatier University - IRIT, Toulouse, France  
Elisa Capello, Politecnico di Torino and CNR-IEIT, Italy  
Constantin F. Caruntu, “Gheorghe Asachi” Technical University of Iasi, Romania  
Meghan Chandarana, NASA Ames Research Center, USA  
Wen-Chung Chang, National Taipei University of Technology, Taiwan

Colin Chibaya, Sol Plaatje University, South Africa  
Tawfiq Chowdhury, University of Notre Dame, Indiana, USA  
Stéphanie Combettes, University Paul Sabatier of Toulouse | IRIT Lab, France  
Cosmin Copot, University of Antwerp, Belgium  
Shreyansh Daftry, NASA Jet Propulsion Laboratory | California Institute of Technology, Pasadena, USA  
Giulia De Masi, Zayed University / Rochester Institute of Technology (RIT), Dubai, UAE  
Angel P. del Pobil, Jaume I University, Spain  
Daniel Delgado Bellamy, University of the West of England, Bristol, UK  
Sotirios Diamantas, Tarleton State University | Texas A&M System, USA  
Manuel J. Domínguez Morales, University of Seville, Spain  
Hind Bril El Haouzi, University of Lorraine, France  
Larbi Esmahi, Athabasca University, Canada  
Anna Esposito, Università della Campania “Luigi Vanvitelli”, Italy  
Nicola Fabiano, Studio Legale Fabiano, Italy / International Institute of Informatics and Systemics (IIIS), USA  
Hugo Ferreira, INESC TEC / Porto Polytechnic Institute, Portugal  
Daniel Filipe Albuquerque, Polytechnic Institute of Viseu, Portugal  
Terrence P. Fries, Indiana University of Pennsylvania, USA  
Wai-keung Fung, Robert Gordon University, Aberdeen, UK  
Zhiwei Gao, Northumbria University, UK  
Jemin George, DEVCOM Army Research Laboratory (ARL), USA  
Javad Ghofrani, HTW Dresden University of Applied Sciences, Germany  
Martin Giese, University Clinic of Tuebingen, Germany  
Philippe Giguère, Laval University, Canada  
Jordi Guitart, Universitat Politècnica de Catalunya (UPC), Spain  
Maki Habib, The American University in Cairo, Egypt  
Cédric Herpson, University Pierre and Marie Curie (UPMC) | LIP6, Paris, France  
Randa Herzallah, University of Aston, UK  
Gerold Hölzl, University of Passau, Germany  
Wladyslaw Homenda, Warsaw University of Technology, Poland  
Wei-Chiang Hong, Asia Eastern University of Science and Technology, Taiwan  
Konstantinos Ioannidis, Information Technologies Institute - Centre for Research and Technology Hellas, Thessaloniki, Greece  
Kamran Iqbal, University of Arkansas at Little Rock, USA  
Miquel Kegeleirs, IRIDIA | Université Libre de Bruxelles, Belgium  
Lial Khaluf, Independent Researcher, Germany  
Hasan Ali Khattak, National University of Sciences and Technology (NUST), Islamabad, Pakistan  
Igor Kotenko, SPIIRAS and ITMO University, Russia  
Hak Keung Lam, King's College London, UK  
Charles Lesire, ONERA/DTIS | University of Toulouse, France  
Baoquan Li, Tiangong University, China  
Guoyuan Li, Norwegian University of Science and Technology, Norway  
Hsieh-Yu (Shay) Li, Dyson Singapore, Singapore  
Jiaoyang Li, University of Southern California, USA  
Ji-Hong Li, Korea Institute of Robot and Convergence, Republic of Korea  
Juncheng Li, East China Normal University, Shanghai, China  
Yangmin Li, The Hong Kong Polytechnic University, Hong Kong  
Zhan Li, Swansea University, UK

Enjie Liu, University of Bedfordshire, UK  
Eric Lucet, CEA | LIST | Interactive Robotics Laboratory, Palaiseau, France  
Ren C. Luo, National Taiwan University, Taiwan  
Nacer K M'Sirdi, Aix-Marseille Université, France  
Jacques Malenfant, Sorbonne Université - CNRS, France  
Guilhem Marcillaud, Paul Sabatier University, IRIT, France  
Aurelian Marcu, Center for Advance Laser technology CETAL - National Institute for Laser Plasma and Radiation Physics, Romania  
Raúl Marín Prades, Jaume-I University, Spain  
Philippe Martinet, INRIA Sophia Antipolis, France  
Ignacio Martinez-Alpiste, University of the West of Scotland, UK  
Rajat Mehrotra, Teradata Inc., Santa Clara,, USA  
René Meier, Hochschule Luzern, Switzerland  
Oliver Michler, Technical University Dresden, Germany  
Sérgio Monteiro, Centro Algoritmi | University of Minho, Portugal  
Naghme Moradpoor, Edinburgh Napier University, UK  
Paulo Moura Oliveira, UTAD University, Vila Real, Portugal  
Luca Muratore, Istituto Italiano di Tecnologia, Genova, Italy / The University of Manchester, UK  
Taro Nakamura, Chuo University, Japan  
Roberto Nardone, University of Reggio Calabria, Italy  
Rafael Oliveira Vasconcelos, Federal University of Sergipe (UFS), Brazil  
Flavio Oquendo, IRISA - University of South Brittany, France  
Luigi Palmieri, Robert Bosch GmbH, Corporate Research, Germany  
Eros Pasero, Politecnico di Turin, Italy  
Timothy Patten, University of Technology Sydney, Australia  
Ling Pei, Shanghai Jiao Tong University, China  
Damien Pellier, Université Grenoble Alpes, France  
Van-Toan Pham, National Taipei University of Technology, Taiwan  
Johan Philips, KU Leuven, Belgium  
Agostino Poggi, DII - University of Parma, Italy  
Radu-Emil Precup, Politehnica University of Timisoara, Romania  
José Ragot, Université de Lorraine, France  
Leonīds Ribickis, Riga Technical University, Latvia  
Douglas Rodrigues, Paulista University - UNIP, Brazil  
Oliver Roesler, Vrije Universiteit Brussel, Belgium  
Joerg Roth, Nuremberg Institute of Technology, Germany  
Loris Roveda, Università della Svizzera italiana (USI), Switzerland  
Spandan Roy, International Institute of Information Technology, Hyderabad, India  
Fariba Sadri, Imperial College London, UK  
Mohammad Safeea, Coimbra University, Portugal / ENSAM, Lille, France  
Lakhdar Sais, CNRS | Artois University, Lens, France  
Michael A. Saliba, University of Malta, Malta  
Ivan Samylovskiy, Lomonosov Moscow University, Russia  
Ricardo Sanz, Universidad Politecnica de Madrid, Spain  
Jagannathan Sarangapani, Missouri University of Science and Technology, USA  
André Schneider de Oliveira, Federal University of Technology - Parana, Brazil  
Cornelia Schulz, University of Tübingen, Germany  
Vesna Sesum-Cavic, TU Wien, Austria

Mahmoud Shafiee, University of Kent, Canterbury, UK  
Abdel-Nasser Sharkawy, South Valley University, Qena, Egypt  
Inderjeet Singh, University of Texas at Arlington Research Institute (UTARI), USA  
Edoardo Sinibaldi, Istituto Italiano di Tecnologia (IIT), Italy  
Petr Skobelev, Samara Technical University / Knowledge Genesis Group, Russia  
Mohammad Divband Soorati, University of Southampton, UK  
Bernd Steinbach, University of Mining and Technology, Freiberg, Germany  
Claudius Stern, FOM University of Applied Sciences, Essen, Germany  
Roy Sterritt, Ulster University, UK  
Yun-Hsuan Su, University of Washington, Seattle, USA  
Alireza Taheri, Sharif University of Technology, Tehran, Iran  
Saied Taheri, Virginia Tech, USA  
Omar Tahri, PRISME | INSA Centre Val-de-Loire, France  
Brahim Tamadazte, FEMTO-ST Institute / CNRS, France  
Maryam Tebyani, University of California, Santa Cruz, USA  
Francesco Tedesco, University of Calabria, Italy  
Giorgio Terracina, Università della Calabria, Italy  
Carlos M. Travieso-González, Institute for Technological Development and Innovation in Communications (IDeTIC) | University of Las Palmas de Gran Canaria (ULPGC), Spain  
Xuan-Tung Truong, Le Quy Don Technical University, Vietnam  
Eddie Tunstel, Motiv Space Systems Inc. / Motiv Robotics, USA  
Paulo Urbano, Universidade de Lisboa, Portugal  
Vivek Shankar Varadharajan, École Polytechnique de Montréal, Canada  
Ramon Vilanova, School of Engineering - UAB, Spain  
Nikolaos (Nikos) Vitzilaios, University of South Carolina, USA  
Holger Voos, University of Luxembourg, Luxembourg  
Stefanos Vrochidis, Information Technologies Institute | Centre for Research and Technology Hellas, Greece  
Dingkang Wang, University of Florida, USA  
Mingfeng Wang, The University of Nottingham, UK  
Yin-Tien Wang, Tamkang University, Taipei, Taiwan  
Guowu Wei, University of Salford, UK  
Luke Wood, University of Hertfordshire, UK  
Haiyan Wu, CAS Key Laboratory of Behavioral Science | University of Chinese Academy of Sciences, China  
Yuanlong Xie, Huazhong University of Science and Technology, China  
Jiajun Xu, Nanjing University of Aeronautics and Astronautics, China  
Reuven Yagel, Azrieli - Jerusalem College of Engineering, Israel  
Chenguang Yang, University of the West of England, Bristol, UK  
Linda Yang, University of Portsmouth, UK  
Ali Zemouche, Université de Lorraine, France  
Haichao Zhang, Horizon Robotics, USA  
Vadim Zhmud, Novosibirsk State Technical University, Russia  
Huiyu (Joe) Zhou, University of Leicester, UK

## Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

## Table of Contents

|  |    |
|--|----|
| The Implementation of Disruptive Measures to Enhance Productivity in an Advanced-Manufacturing Environment<br><i>Ganiyat Salawu, Glen Bright, and Chiemela Onunka</i>  | 1  |
| Optimal Control of Unmanned Aerial Vehicles Electric Launcher<br><i>Mohammad H Sadraey</i>   | 8  |
| Traffic Signal Recognition and Application Algorithm for the Autonomous Vehicle in V2X Unable Areas<br><i>Yejin Gu and Daejun Kang</i>   | 14 |
| A Quantitative Measure for the Evaluation of Drone-based Video Quality on a Target<br><i>Daniela Doroftei, Geert De Cubber, and Hans De Smet</i>   | 19 |
| Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations<br><i>Rialda Spahic, Vidar Hepso, and Mary Ann Lundteigen</i>   | 25 |
| A Multimodal AI Approach for Intuitively Instructable Autonomous Systems: A Case Study of an Autonomous Off-Highway Vehicle<br><i>Abdellatif Bey Tamsamani, Anil Kumar Chavali, Ward Vervoort, Tinne Tuytelaars, Gorjan Radevski, Hugo Van Hamme, Kevin Mets, Matthias Hutsebaut-Buysse, Tom De Schepper, and Steven Latre</i> | 31 |
| FPGA Frontend for Highly Efficient Automotive LIDAR Perception<br><i>Sanaz Asgarifar, Pedro Barbosa, Amir Farzamiyan, Marcelo Alvez, Alexandre Correia, and Joao Ferreira</i>  | 40 |
| Self-Aware Industrial Control Systems through Cloud Based Autonomic Computing<br><i>Christopher Rouff, Ali Tekeoglu, Joseph Maurio, and Alexander Beall</i>  | 43 |
| Agility and Semantic Structures to Scaffold Modern Academic Education<br><i>Karsten Bohm</i>   | 45 |

# The Implementation of Disruptive Measures to Enhance Productivity in an Advanced-Manufacturing Environment

Salawu Ganiyat\* Glen Bright\*\*

\*Mechanical Engineering Department  
The University of KwaZulu-Natal,  
Durban, South Africa

\*e-mail ganiatsoliu@gmail.com.

\*\*e-mail: brightg@ukzn.ac.za

Onunka Chiemela

Data Center Engineering Operations,  
AWS,

Virginia, United States

e-mail: onunkac@amazon.com

**Abstract**— The performance of a manufacturing process and the need to increase throughput and reduce the cost of production are of the highest interest among modern manufacturers. Classical mathematical models developed were used to describe the operation of a robot during a complex pick and place task in a virtual manufacturing environment. The design parameters of the conveyor system were examined. Existing designs were studied and modeled to select the best operating speed to optimize throughput during the manufacturing system. The modeled design parameters were analyzed using MATLAB. The results were presented graphically with an optimal throughput obtained at an operating speed of 390m/seconds, operating time of 0.4secs, and power consumption of 12700W. The operation of the robotic arm was manipulated during service to determine the angle of placement that yielded a consistent and efficient throughput during the pick and place task. Consequently, an optimal throughput was reached when setting the manipulator at an angle of 88 degrees.

**Keywords**- Automation of manufacturing process; modelling of robotic arm; simulation of results.

## I. INTRODUCTION

The goals of the fourth industrial technology were to improve the efficiency of the manufacturing process, improve product quality, and enhance safety and security in manufacturing industries. Industry 4.0 involved the integration of intelligent, disruptive technologies into the manufacturing environment by focusing on automation, real-time data, artificial intelligence, machine learning, etc. The new technology uses its automation approach to boost productivity and output, increase efficiency, and create an intelligent manufacturing environment. [1][2]. Industry 4.0 has emerged with various disruptive technologies that can transform multiple manufacturing sectors from labor-intensive processes to a modernized automation process [3][4]. Implementing robotics in an advanced manufacturing environment has been a trending technology that requires further studies, innovations, and adaptations.

Robotics has been used in a wide range of applications. Applications include transporting materials and parts from conveyor systems to various stations, assembling parts,

painting, sorting, packaging and labeling, inspection and testing, picking and place, materials handling, palletizing, etc. [5]. The applications of robotic technology to various systems were achieved with appreciable high speed, precision, and endurance limit [6][7]. The implementation of robotics in the manufacturing sector has improved the economic situation of various industries by enhancing the production process, product quality, and throughput rate. The robotic system uses several innovative sensory devices and control techniques to improve agility and productivity [8][9]. The importance of research in robotics in a recent study is to find solutions to manufacturing problems [10].

Enhancing a manufacturing process has become more apparent due to its technological impact on an advanced manufacturing environment. Recently, there have been various attempts to improve the service of industrial robots to perform some complex and time-consuming tasks in the manufacturing environment [11]. Implementing industrial robots as a disruptive technology in performing tasks has enabled technological growth in the manufacturing sector [12]. The behavior of a manufacturing system can be better expressed with classical models. In this paper, the throughput rate of a manufacturing system was enhanced. Classical mathematical models were used to describe a manufacturing environment, whereas robot was used to perform a complex pick and place task. The scenario was analyzed using various mathematical tools. The manufacturing scenario was validated by manipulating the robotic arm motion to select the best operating angle.

## II. LITERATURE REVIEW

Effective implementation of various automation tools adequately increased production performance and machine up-time. Industry 4.0 introduced various disruptive technologies and has supported manufacturers to have outputs at a reduced cost. Automation of services in an advanced manufacturing environment has reshaped the global market for higher productivity [13]. The uses of disruptive manufacturing technologies have brought about numerous improvements in productivity in the manufacturing process. In recent technological advancements, robots have been disruptive tools and have

introduced tremendous changes in the advanced manufacturing environment. Robots have been widely used among manufacturers due to their exceptional efficiency and ability to perform complex tasks within the shortest period [14][15]. Robotics applications reinforced the manufacturing industries and enabled competitiveness among manufacturers [16].

Various disruptive technologies have been developed to support and enhance the smooth running of multiple stages of the manufacturing processes [17][18]. The use of automation in augmenting human activities has increasingly improved the efficiency of the manufacturing processes. Industrial robots are efficient automation tools with higher flexibility and can be programmed to perform various tasks within a short time [19][20]. The research presented in [20][21][22] showed how robotic manufacturing systems were optimized for optimal efficiency and productivity. The part's positioning to be picked up can affect the robot's performance in an advanced manufacturing environment [23][24]. In this present work, robots were implemented to perform a pick and place task. The scenario was studied to determine the best angle of twist that can produce an optimal throughput rate. The research showed that at an angle of 88 degrees, the manipulator performed excellently as when performing the desired task. This research outcome can be useful among manufacturers in the competitive market.

### III. METHODOLOGICAL APPROACH

A virtual manufacturing scenario was developed and studied to determine the impact of using a robot to perform the complex pick and place task in a manufacturing environment. Classical mathematical models were developed to describe each process involved. The design parameters of an existing conveyor system were studied and simulated to obtain the best design parameter that can yield an optimal throughput rate during a manufacturing process. Also, the waiting time involved during the packaging stage was modeled and studied. The manufacturing process involved parts arriving via a conveyor system from multiple stations to a buffer station. The arriving parts were picked up and placed by a robot.

The arrival behavior followed a Poisson process with varying mean arrival rates. The arrival of the parts from the buffer station took the form of a negative exponential distribution and observed the impatient behavior of customers in the M/G/I queuing system. The M/G/I queuing system follows a Poisson process varying with mean arrival rate  $\lambda$  with a general service time distribution. The repeatability motion of the robotic arm ensures the system has a deterministic feeding time  $\mu$  (mean service time). Parts that were not picked up during the initial cycle were redirected for service in the next cycle.

Mathematical models were developed and analyzed using the queuing mathematical theory. The average queuing time was modeled. This enabled proper optimization of the

performance of each robot with its corresponding queue. Various mathematical expressions were developed and solved using the Newton-Raphson iteration method. Some values were assumed and implemented into the equation to test the efficiency of the models.

#### A. Model analysis for an optimal throughput rate

During the complex pick and place task, the robot does not pick up all arriving parts as the conveyor move past the vision camera. Other parts were redirected to join the following arriving parts. During the traveling period of parts, there existed a continuous movement of parts with a minimum distance/gap  $\alpha$  between the work envelope and the boundaries within which the arriving parts exist. The number of visible parts was fed along with the conveyor system and denoted as  $N_f$ . The center point of the arriving part was determined within the Height and width of the work envelope. The Diameter of the part fed into the process was assumed to be less than 20mm. The velocities of the conveyor belts were the same.

#### Assumptions

- The system was not saturated or starved.
- There exist continuous motion in the conveyor system
- The Height and width of the work envelope were assumed to be equal
- All parts were well guided to avoid slipping off

The following notations and parameters were used to arrive at a suitable expression that was implemented to obtain higher throughput.

#### Notation/parameters

$N_f$ =Number of part fed through the work envelope

$c$  = Centre of work envelope

$d$  = Diameter of fed part

$\alpha$  = Minimum clearance required =  $d/2$

$w$  = Height, and width of the workpiece

$P_b$ = probability that work has been fully cleared

$v$ = Velocity of the belt

$T_r$ = The robot throughput rate

$R_b$ =The arrival rate of parts from the conveyor (parts/mm<sup>2</sup>).

An impatient customer's renewal theorem was implemented to obtain a suitable mathematical expression for optimal throughput. In (1), the Velocity of the conveyor system  $v$  is expressed as:

$$V = \frac{\pi D n}{60} \quad (1)$$

Equation (2) was developed to represent the probability that the work piece has been fully cleared from the work envelope.

$$P_b = \frac{(w-d-2\alpha)^2}{(w-d)^2} \quad (2)$$

In equation(3), the center of the work envelope  $c$  was modeled while considering the width and height of the work

envelope and the Diameter of the conveyor belt with speed (v).

$$c = \frac{w\pi(\alpha+d)^2}{(w-d)^2v} \quad (3)$$

Similarly, in (4), the expression for the arrival rate of parts from the conveyor to the server (robot) was modeled as:

$$r_b = \frac{1}{c} \frac{(w-d)^2v}{w\pi(\alpha+d)^2} \quad (4)$$

In (5), the number of parts fed through the work envelope was obtained by considering the width and height of the work envelope, the Velocity of the conveyor, the probability of cleared work from the work envelope, the center of the work envelope, and the arrival rate of parts from buffer station to the pickup point.

$$[N_f] = \left(\frac{w}{v}\right) r_b p_b e^{-rbc} \quad (5)$$

The throughput rate was also determined by finding the relationship between the number of parts fed through the work envelope, the width of the work envelope, and Velocity of the conveyor system, as indicated in equation (6).

$$T_r = \frac{N_f}{\left[\left(\frac{w}{v}\right) + t_p + [N_f]t_{cl}\right]} \quad (6)$$

### B. Mathematical Model for Queuing Theory and Newton-Rap son.

The model used in the research described the waiting time during the packaging stage in a virtual manufacturing environment. The system performance was modeled using a Poisson distribution function where the service times were exponentially distributed. Classical mathematical models were developed for describing and making a decision in the packaging stage of the virtual manufacturing process.

Parameters were assumed under operating conditions, and values were effectively used to describe the manufacturing system. Parameters that were used within the waiting line model include the cost of waiting per hour ( $C_w$ ), the average number of product in queue ( $L_s$ ), the cost of robot/hour ( $C_p$ ), the average number of product arriving from the machine per unit time ( $\lambda_i$ ), packaging rate which represented the service rate per unit time measured per hour ( $\mu$ ), utility factor of the server which is the robot ( $\rho_n$ ).

The queuing mathematical theory was then used in analyzing the classical models, and the queuing mathematical theory was used in analyzing the classical models. The models that were used in representing the packaging stage is summarized below:

The products arrival rate was presented as

$$\sum_{i=1}^n \lambda_i(t) = n\lambda(t) \quad (7)$$

A discrete value was assumed for the total number of products arriving during the packaging stage and was expressed as:

$$N = n\lambda_x t \quad (8)$$

Total product packed by available robots was expressed as:

$$\mu_j(t) \times t \times m \quad (9)$$

During the packaging phase, the utility factor of robots was expressed as:

$$\rho_n(t) = \frac{\lambda_i(t)}{\mu_j(t)} \quad (10)$$

The mean queue time required by each robot and the production rate of each robot was expressed as:

$$p_j(t) = (\lambda_i + \mu_j)\rho_n(t) + \lambda_i\rho_{n-1}(t) + \rho_{n+1}(t) \quad (11)$$

The overall productivity rate of the system  $\rho_n(t)$  was modeled as:

$$\rho_n(t) = \left(\sum_{i=1}^n \lambda_i + \sum_{j=1}^n \lambda_i \mu_j\right)\rho_n(t) + \sum_{i=1}^n \lambda \rho_{n-1}(t) + \sum_{j=1}^m \mu_j \rho_{n+1}(t) \text{ for } t \rightarrow \infty \quad (12)$$

For a discrete value of time t;

$$\rho_n(t) = \rho_n(t) + \lambda_n \rho_{n-1}(t) + \mu(t) \rho_{n+1}(t) \quad (13)$$

The waiting time denoted by W is expressed in 14:

$$W = \sum_{j=1}^m \mu_j - \sum_{j=1}^m \lambda_i, \quad w = \frac{1}{m\mu_j - n\lambda_i} \quad (14)$$

Waiting time (unit/hour) expressed as:

$$L_s = \frac{n\lambda_i}{n\lambda t - m\mu_j} \quad (15)$$

The cost of production T is given in (16)

$$T = mc_p \mu_j + c_w \rho_n w L_s \quad (16)$$

After substituting the parameters into (16), the production cost gives:

$$T = mC_p \mu_j + C_w \left( \frac{\lambda_i^2}{(\mu_j(m\mu_j - n\lambda_i)^2)} \right) \quad (17)$$

The system's performance was optimized using the waiting time model expressed in (15). The optimization process was achieved by differentiating the model to the service rate of each robot.

$$C_p + \frac{\lambda^2[(\mu-\lambda)^2 + 2\mu(\mu-\lambda)]}{\mu^2(\mu-\lambda)^4} = 0 \quad (18)$$

The model for optimal service was modeled as:

$$C_p \mu^5 - 3C_p \mu^4 \lambda + 3C_p \mu^3 \lambda^2 - C_p \mu^2 \lambda^3 + 3\mu \lambda^2 C_w - C_w \lambda^3 = 0 \quad (19)$$

### C. Simulation Using Newton-Raphson Method.

The Newton-Rap son Iteration method determined the approximate values used in the simulation of the models presented in the thesis. Newton-Rap son iteration method based its strategy on finding an approximate value for the root of a valued function of x [25]. Using the Newton-Rap son equation reduced the errors that were likely to set in when calculating the roots of functions. The efficiency of the Newton-Rap son method was the advantage it has over other methods. This method converges fast when compared with the Gauss Seidel, and other methods used in finding roots of quadratic equations. If it converges we get root (answer) in less number of steps. It requires only one guess.

Formulation of this method is simple. The Newton-Rap son iteration method was used to find the zeros of the arbitrary equations that were implemented during the manufacturing scenario; here, the specific root of a function depended on the initial value. The Newton-Rap son's iteration strategy was utilized in the research as follows:

Given that the root of the derived equation was  $r$ , let  $x_0$  be the estimated value of  $r$ ,  $h$  represents a measure of the approximate value of  $x_0$  from the exact value.

Where

$$r = x_0 + h, h = r - x_0 \quad (20)$$

$h$  was very small and its linear approximation was represented as:

$$0 = f(r) = f(x_0 + h) \approx f(x_0) + hf'(x_0) \quad (21)$$

The mathematical model was valid if,  $f'(x_0)$  was approximately equals to zero.

$$h \approx \frac{f(x_0)}{f'(x_0)} \quad (22)$$

$$r = x_0 + h \approx x_0 - \frac{f(x_0)}{f'(x_0)} \quad (23)$$

Therefore, the estimated value  $x_1$  of  $r$  yielded:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (24)$$

Similarly,  $x_2$  was derived as a function of  $x_1$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \quad (25)$$

For a required number of  $x$ ,  $x_n$  is the next approximate value.

Therefore,  $x_{n+1}$  was modelled as:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (26)$$

The Newton Rap son Iteration was used for analyzing the service rate ( $\mu$ ), and modeled as:

$$\mu_{n+1} = \mu_n - \frac{f(\mu_n)}{f'(\mu_n)} \quad (27)$$

Equations (28) and (29) represented the first and second-order of the Newton-Rap son iteration model of a production system [1].

$$f(\mu) = C_p\mu^5 - 3C_p\mu^4\lambda + 3C_p\mu^3\lambda^2 - C_p\mu^2\lambda^3 + 3\mu\lambda^2C_w - C_w\lambda^3 \quad (28)$$

$$f'(\mu) = 5C_p\mu^4 - 12C_p\mu^3\lambda + 9C_p\mu^2\lambda^2 - 2C_p\mu\lambda^3 + 3\lambda^2C_w \quad (29)$$

The assumed values for the manufacturing scenario were ( $C_p = R2/\text{hour}$ ,  $C_w = R0.1/\text{hour}$ ,  $\lambda_i = 5$  units/hour for 10 machines,  $\mu = 2$  units per hour). The values were implemented into the general equation, solved, and the mathematical models were analyzed. Outcomes were simulated using the Newton-Rap son iteration expression. Close approximates values obtained were varied against each other to obtain the presented results graphically.

#### D. Variation of the Robotic Arm for an Optimal Throughput.

During the complex pick and place task, the robotic arm motion was manipulated to determine the best operating angle that gave an optimal throughput. Three different grasps were involved: the pre-grasp, the grasp, and the post grasp. The pre-grasp provided the suitable positioning of the end-effectors away from the object position and coordinated the trajectory motion to avoid the occurrence of collision during service. Grasping involved positioning the end-effectors with their fingers ready to grip the object for the designated task, while the post-grasp involved moving away of the end-effectors from the point where the object was grasped. The operation was controlled by computing the desired action for the end-effectors to perform the placing down task [25].

Arriving parts from the conveyor were picked up randomly by the manipulator following the first-in, first-out order (FIFO). The vision camera determines the object's position at the robot workspace for immediate pickup. The average operating time was studied and presented in Table 1.

TABLE I. THE AVERAGE TIME SPENT ON THE PICK AND PLACE TASK

| S/no | Time Spent during the pick/place task |   |            |
|------|---------------------------------------|---|------------|
|      | Motion                                | Task  | Left/right |
| 1    | Motion 1                              | The home pose to view an object on the conveyor   | 4.5        |
| 2    | Motion 2                              | Reach object to grasp an object                   | 3.2        |
| 3    | Motion 3                              | Moving to the object drop pose                    | 3.0        |
| 4    | Motion 4                              | Moving from object dropped pose back to home pose | 3.41       |

An operational variation of the manipulator motion was achieved by varying the angle between 81 degrees and 96 degrees. The effects of varying the angle and other operating parameters were studied to determine its impacts on the throughput rate of the manipulator.

#### IV. RESULTS AND DISCUSSION

The performance of the robots was examined to study its effects on the throughput rate during a pick and place task. The result in Fig. 1 shows the relationship between the packaging stages of manufacturing against the arrival rate of the products. Fig. 1 indicates that the packaging rates increased as the arrival rate also increased. The result shown in Fig. 1 implied that the number of robots assumed in the decision-making for the proposed model was suitable to solve the waiting line model.

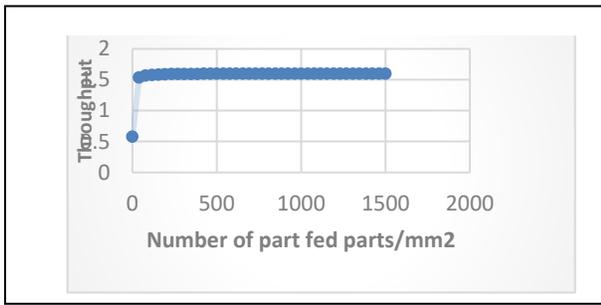


Figure 1. Throughput Rate against Part Fed Through the Conveyor System

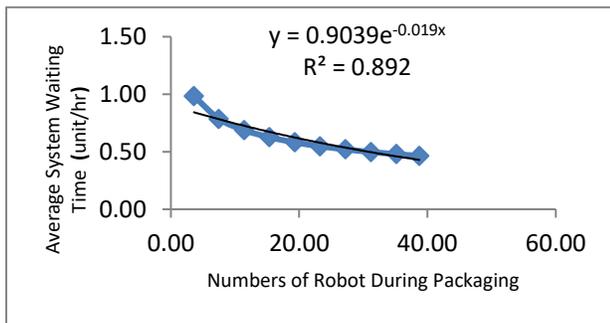


Figure 2. Average Waiting Time Against the Numbers of Robots Used.

During the packaging stage of manufacturing, the average waiting time gradually reduces as the number of robots used to perform the packaging task increases. If more robots were introduced to replace human labor, the lead time would be minimized an efficient production system can be obtained. This is illustrated in Fig 2.

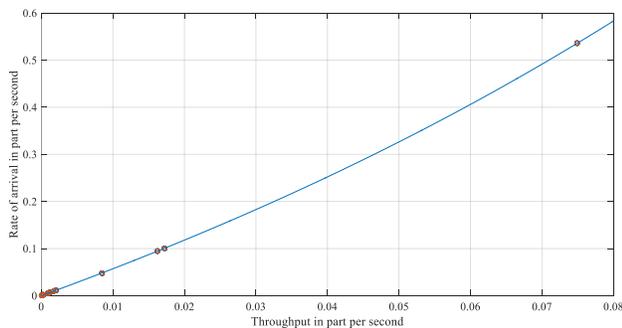


Figure 3. Products Arrival Rate  $R_b$  Against the Throughput  $T_r$  Rate (Parts/Secs).

The result in Fig. 3 shows an increasing throughput rate in correlation with the product arrival rate. The result shown in Figure 3 implies that the selected design parameters for the conveyor system were efficient.

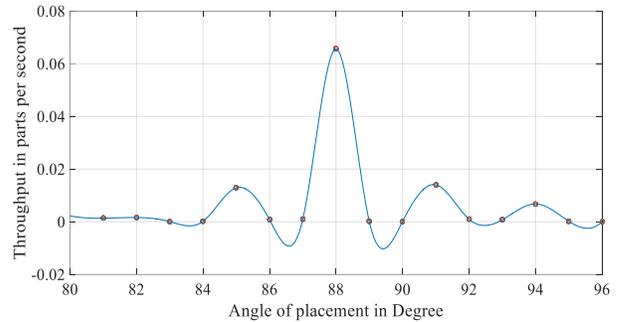


Figure 4. The Average Time Spent for Pick-Place-Task

The robotic arm was varied during the pick and place task to determine the best angle of operation that could produce the highest throughput rate. Result obtained is presented in Fig. 4. This showed that the operational variables of the robotic arm gave an optimal throughput rate at an angle of 88 degrees.

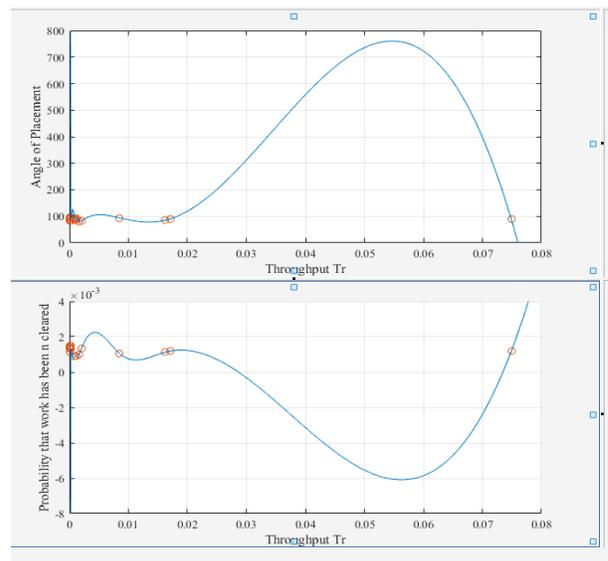


Figure 5. Probability That Work is Being Cleared From the Conveyor Against the Throughput Rates.

The graph in Fig. 5 indicates that the work's probability has been cleared from the conveyor system when measured against the throughput rate. The result showed a closed-looped graph which indicated that as the products arrived, the robots were efficient in picking up from the conveyor faster before it moved past it. The results confirm that the modeled design parameters selected for the conveyor system were efficient for optimal productivity.

### V. CONCLUSIONS AND FUTURE WORKS

Classical mathematical models were used in describing a manufacturing process whereby robots were used to perform a complex pick and place task. The design parameters of a conveyor system were examined, and the best operating

speed that gave optimal throughput during a complex pick and place task was achieved. This was successfully carried out by developing some mathematical models to study the effect of varying the loads below and above the rated speed of the electric motor. Standard equations were further developed and analyzed. MATLAB was used to simulate and analyze the mathematical models and solved using the engineering equation solver (EES). Results were simulated to obtain and select relevant results that gave optimal throughput at an operating speed of 390m/secs, using time at 0.4secs, and at a power consumption of 12700W.

The average waiting time of the robots during the packaging process was studied using the analytical queuing theory. The average queuing time was further differentiated to optimize the performance of each robot with its corresponding queue. A general equation was developed and further analyzed using the Newton-Raphson iteration method. The results confirmed that the suggested model could be suitable for use when the queuing time needs to be controlled during the packaging stage in a real-life manufacturing scenario.

Furthermore, the robotic arm motion was manipulated during the pick and place task to determine the best operating angle for an efficient production system. An optimal throughput rate during the pick and place task was achieved when the robotic arm was positioned at an angle of 88 degrees. This research outcome can be useful among manufacturers in the competitive market. The research is still in progress whereby further studies will be carried out to determine how effective the derived angle will be when the manipulator is slated for heavy task.

#### ACKNOWLEDGMENT

The authors acknowledge the support from the University of KwaZulu-Natal for the funding provided for the research work.

#### REFERENCES

- [1] A. C. Amaral. Using Newton-Raphson Method to Estimate the Condition of Aluminum Electrolytic Capacitors. *34th Annual Conference of IEEE Industrial Electronics*. Coimbra (2008).
- [2] C. Antonelli & F. Quatraro. The effects of biased technological changes on total factor productivity: a rejoinder and new empirical evidence. *J. Tech. Transfer*, 47 (10), pp. 1686-1700, (2014).
- [3] M. Baygin, H. Yetis, M. Karakose, & E. Akin, An effect analysis of industry 4.0 to higher education. *15th international conference on information technology based higher education and training*. Ohrid, Macedonia, (2016).
- [3] P. Bellin, I. Bruno, D. Cenni, & P. Nesi. Managing cloud via Smart Cloud Engine and Knowledge Base. *Future generation and computer system*, 98 (1), pp. 142-154. (2018).
- [4] B. Berman. 3-D printing: The new industrial revolution. *Business horizons*, 55 (2), pp. 155-162, (2012).
- [5] B. Brindle. *howstuffworks.com*. (a division of InfoSpace Holdings, LLC, a System1 Company) Retrieved 4 4, 2020, from <https://electronics.howstuffworks.com/everyday-tech/what-is-disruptive-technology.htm>. (1998).
- [6] F. Christopher. The effects of disruptive innovation on productivity. *Technological Forecasting and Social Change*, 126, pp. 180-193. (2018).
- [7] F, Christopher F. The effects of disruptive innovation on productivity. *Technological Forecasting and Social Change*, 126, pp. 180-193. (2018).
- [8] G. Salawu, B. Glen, C. Onunka,. Impact of disruptive technology on the operational process in an advanced manufacturing environment. *International Journal of Mechanical & Mechatronics*, 2 (3), pp. 47-57. (2020).
- [9] G. Salawu, B. Glen, C. Onunka. Impacts of Disruptive Technology on Operational Process in an Advanced-Manufacturing Environment. *International Journal of Mechanical & Mechatronics Engineering IJMME-IJENS*, 20 (3), pp. 47-58. (2020).
- [10] S. Heydaryan & J. Suaza. (2018). Safety Design and Development of a Human-Robot Process in the Automotive Industry. *Appl. Sci.*, 8 (344), pp. 1-22.
- [11] L. Lin, H. Zhao, & H. Ding. Posture optimization methodology of 6R industrial robots for machining using performance evaluation indexes. *Robot. Comput.-Integr. Manuf*, 48, pp. 59–72. (2017).
- [12] M.W, S. Cross-Validation Optimization for Large Scale Structured Classification Kernel Methods. *Journal of Machine Learning Research*, 9, pp. 1147-1178. (2008).
- [13] X.T. Mathaba, & X. Xia. A Parametric Energy Model for Energy Management of Long Belt Conveyors. *Energies*, 8, pp. 13590–13608. (2015).
- [14] E. Matsas, & G.C., V. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *Int J. International Journal of Interact Design Manufacturing*, pp. 1-15. (2015)
- [15] E. Matsas, G. Christopher, & D. Batras. Effectiveness and Acceptability of a Virtual Environment for Assessing Human - Robot Collaboration in Manufacturing. *International Journal of Advance Manufacturing Technology*, pp. 3903-3917. (2017).
- [16] S. Mousavi, V. Gagnol, B. Bouzgarrou, & P. Ray, Stability Optimization in Robotic Milling Through the Control of Functional Redundancies. *Robot . Comput.-Integr. Manuf*, 50, pp. 181–192. (2017).
- [17] R. Patric, W. Laura, & T. Kirsten, Implementation of virtual reality system for simulation of human - robot collaboration. *Procedia Manufacturing*, 19, pp. 164-174. (2018).
- [18]. Roger, & T.Jeremy. Survey of Research for Performance Measurement of Mobile Manipulators. *Journal of Research of the National Institute of Standards*, pp.121, 15. (2016).
- [19] Salawu, G., Bright, G., & Onunka, C. (2020). Mathematical modeling and simulation of throughput in a robotics. *International Journal of Engineering Research and Technology*, 13 (1), pp. 137-143.
- [20] Salawu, G., Bright, G., & Onunka, C. (2020). Mathematical Modelling and Simulation of Throughput

- in a Robotics Manufacturing System. *International Journal of Engineering Research and Technology*, 13 (1), pp. 137-143.
- [21] Shakir, Muwahida, Zeeshan, Ahsan, & Abdullah, a. (2020). Study of Formation Control of Mobile Robots. *International Journal of Mechanical Engineering and Robotics Research*, 4 (3), pp. 111-116.
- [22] Y. Tian, B. Wang, J. Liu, F. Chen, S. Yang, W. Wang. Research on layout and operational pose optimization of robot grinding system based on optimal stiffness performance. *J. Adv. Mech. Des. Syst. Manuf*, 11. (2017).
- [23] W, S., G, C., T-H, H., K, S., & M., a. S. Performance Evaluation of Human Detection Systems for Robot Safety . *Journal of Intelligent & Robotic Systems*. (2016).
- [24] S. Zhang & Mao, Optimal operation of coal conveying systems assembled with crushers using model predictive control methodology. *Applied Energy*, 198, pp. 65–76. (2017).
- [25] A.C. Amaral. Using Newton-Raphson Method to Estimate the Condition of Aluminum Electrolytic Capacitors. *34th Annual Conference of IEEE Industrial Electronics*. Coimbra. (2008).

# Optimal Control of Unmanned Aerial Vehicles Electric Launcher

Mohammad Hashem Sadraey  
Southern New Hampshire University  
Manchester, USA  
Email: m.sadraey@snhu.edu

**Abstract**— The launch of fixed-wing Unmanned Aerial Vehicles has always been a challenging operation and witnessed various failures in the past decades. The current traditional ramp launchers are employing a rail and are powered by pneumatic/hydraulic systems. This paper discusses a new launch system for fixed-wing UAVs powered by linear synchronous motors and a special track. This novel idea has the potential to be developed to a new technique for launching fixed-wing UAVs. In this paper, linear synchronous motor is recommended as a source of generating launch thrust. The control of launch process is conducted utilizing an optimal controller. Advantages for this method are lower cost, higher reliability, stability, and safety. The verification and validation of the technique are documented using MATLAB simulation. This novel automatic launch system - powered by linear synchronous motors - provides a lower cost and a higher reliability than current launchers.

**Keywords**- optimal control; unmanned aerial vehicles; launcher.

## I. INTRODUCTION

The vast majority of fixed-wing Unmanned Aerial Vehicles (UAV) employ conventional launch and recovery techniques (i.e., conventional takeoff and landing). UAVs without a landing gear do not require a runway and are launched via various techniques, such as rail launchers; rocket launch; hand launch; and vertical takeoff.

A rail/ramp launcher eliminates the need for conventional landing gear for the purposes of takeoff. A ramp launcher is a mechanical device to accelerate a fixed-wing UAV to a minimum controllable airspeed before releasing it from launcher. The current traditional ramp launchers are employing a rail and are powered by pneumatic/hydraulic systems.

The current launch techniques have a few challenges including reliability, cost, stability, and safety. The electric launch technique that has not been employed so far in any unmanned aerial systems, have the potential to solve all these challenges. The novel technique presented in this paper is for the first time for launching an unmanned aircraft.

The pneumatic launching systems are typically designed as a catapult rail launcher. The energy storage is a compressed gas, usually the air is used, due to its availability and no cost. Air is pressurized by a compressor; and is stored in a high-pressure accumulator tank. The launch is performed by applying the compressed air force on UAV through a cylinder/pipe/tube along the ramp. The applied force can be

regulated by adjusting the air pressure in order to support UAVs of different mass.

The hydraulic launching systems are much like the pneumatic ones. To perform a launch, a valve is quickly opened, so that the oil is pumped inside the cylinder to push the piston. Consequently, the gas is compressed on the other side.

The electric launching system is very much different from pneumatic and hydraulic ones. External power is usually used for starting the engine or maintenance activities on the UAV. The external power system allows operation of various electrical systems without discharging the onboard battery. The idea here is to use a special type of electric motors, referred to as the Linear Synchronous Motor (LSM). A LSM is a motor by which the mechanical motion is in synchronism with the magnetic field. A linear electric motor is a motor that has its stator and rotor unrolled. So, instead of producing a torque, it generates a linear force along its length.

To achieve benefits similar to those seen in hybrid-/all-electric ground-based and marine vehicles, large electric machines have been developed for aircraft electric propulsion [1][2]. Subsonic LSMs with high-temperature superconducting magnets are designed [3] to accelerate to a velocity of 1200 km/h in the near-vacuum tubes of 0.001 atm. A method for designing DC-excited linear synchronous motor as a drive system is presented by [4]. Design and characteristic analysis on the short-stator LSM for high-speed maglev propulsion is provided by [5].

The application of LSMs has revolutionized the high-speed passenger transport systems and trains [6][7] in a number of countries including Europe, Japan, and China. Linear synchronous electric motors have been widely employed in industry applications such as in roller coasters. They can drive a linear motion load without gears and mechanical intermediates. For instance, the Incredible Hulk roller coaster [8] was designed in 1999 as part of a 1-billion-dollar park construction. In this roller coaster, 230 electric motors power a series of drive tires which provides 0 to 40 mph in 2 seconds at a 30° incline. Section III provides fundamentals and technical specification of LSMs.

In the past decade, a number of new launch techniques and new launch power sources have been developed. DARPA and Aurora have developed a new runway independent UAV launch and recovery system; called SideArm [9]; which is capable of both land and sea-based operations. It brings

runway independence to a new class of fixed wing UAVs. The SideArm is a self-contained, portable apparatus to launch and retrieve other unmanned aircraft from trucks, ships and fixed bases.

The Insitu [10] trailer-mounted Mark 4 launcher - a self-powered by an onboard diesel fuel generator and air compressor - is compatible with all of Insitu's unmanned aircraft. The setup of this launcher with a weight of 4,200 lb, a deployed length of 22 ft, and with two operators, takes about minutes. The AAI RQ-7 Shadow is also launched from a trailer-mounted pneumatic catapult. Its pneumatic launcher can accelerate the 170 kg air vehicle to 70 knots in 12 m.

The rest of the paper is organized as follow. In Section II, fundamentals of launch technique and governing equations will be described. The launch power system using Linear Synchronous Motors are discussed in Section III. The launch control system employing an optimal controller is presented in Sections IV and V. The verification and validation of the technique are documented using MATLAB simulation in Section VI.

## II. FUNDAMENTALS OF RAIL LAUNCHERS

In a rail launch, the aerial vehicle is placed on the rail; when started, moves in the launcher along the rail for a very short time. In an ideal situation, a straight motion is desired for the UAV along the launcher. Major elements of a launcher (Figure 1) are ramp/slipway, elevation platform/mechanism, mobility wheels, transportation truck, power source, and push mechanism. Unlike missiles, there is often no rush in launching a UAV; thus, the azimuth and elevation platforms are adjusted by human operators, rather than through a closed-loop control system.

In the design of a launcher, parameters such as launcher length, launcher weight, launch angle, and the required force and power must be determined. Moreover, the type of source of launcher power (here, electric) needs to be formulated.

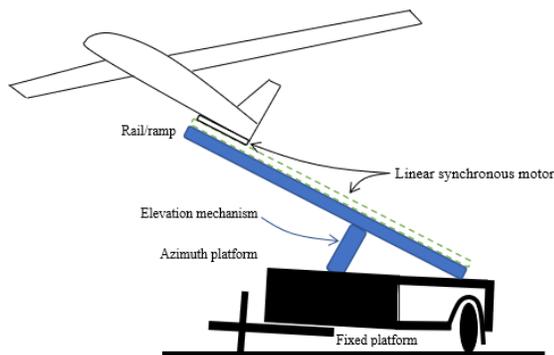


Figure 1. Major elements of a launcher

The launch process is basically a linear accelerated motion, where the UAV is accelerating along the ramp. The relationship between acceleration, launch speed, and launcher length has significant implications for launcher design. In launch operation, a large structural load on the UAV is often imposed, due to a high linear acceleration.

When a moving object with an initial velocity of  $V_1$  accelerates to a new velocity of  $V_2$ , the distance ( $x$ ) covered is governed by the following equation:

$$V_2^2 - V_1^2 = 2aL_R \quad (1)$$

where “a” represents the linear acceleration. For an UAV launcher, the initial velocity is often zero, and the distance traveled is equal to the length of ramp launcher ( $L_R$ ). In order for a launcher to create such acceleration, it must provide a sufficient launch force ( $F_L$ ). Along the launcher ramp, sum of the forces along x-axis creates the acceleration. The contributing forces in this accelerated launching motion are launch force (including UAV engine thrust);  $F_L$ ; UAV-ramp nonlinear friction force;  $F_f$ ; UAV weight ( $W = mg$ ), UAV drag ( $D$ ), UAV lift ( $L$ ), where  $m$  denotes the UAV mass. In the  $x$  direction (along the ramp), the momentum equation is:

$$F_L - F_f - D = ma \quad (2)$$

The aerodynamic forces of lift and drag [11] are:

$$L = \frac{1}{2} \rho V^2 S C_L \quad (3)$$

$$D = \frac{1}{2} \rho V^2 S C_D \quad (4)$$

where  $\rho$  is air density  $V$  is velocity,  $S$  is wing planform area, and  $C_L$  and  $C_D$  are the lift and drag coefficients respectively. The friction force is proportional to the UAV normal force ( $x$ -component of the weight minus lift),  $N$ . It is the product of the coefficient of friction  $\mu$  with the normal force  $N$ , and acts to oppose the motion. Hence, the UAV-Ramp friction force ( $F_f$ ) is:

$$F_f = \mu N \quad (5)$$

where  $\mu$  is the friction coefficient between launcher rails and the UAV (indeed, two metals), and  $N$  denotes the normal force on the ramp. UAV weight and UAV lift are two contributing forces to the normal force. The ramp (Figure 2) has frequently a launch (i.e., climb) angle,  $\theta$ .

$$N = W \cos \theta - L \quad (6)$$

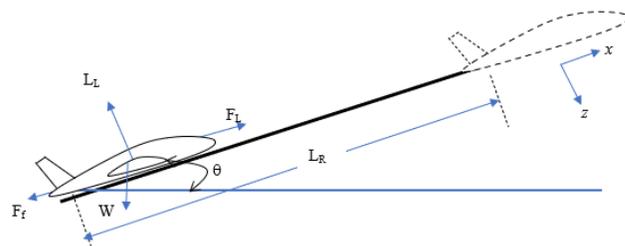


Figure 2. Contributing forces on a launcher

The friction coefficient ( $\mu$ ) between a UAV and the launcher rails is typically about 0.05-0.12. When the friction force from equations 5 and 6 is substituted into equation 4, we obtain:

$$F_L = W \cos \theta (1 + \mu) + ma \quad (7)$$

This is the force ( $F_L$ ) that the launcher (indeed, the LSM) and UAV engine must provide to generate the desired acceleration during the launch operation. The engine thrust is contributing to the launch process; it will create a portion of the required launch force.

### III. LINEAR SYNCHRONOUS MOTOR PROPULSION SYSTEM

A launcher requires a power supply for the activation of mechanisms such as the elevation platform. This paper introduces a new launch system for fixed-wing UAVs powered by LSM and a special track. Basically, the idea is that by launching the UAV along the rails at an angle, using LSMs and an electric power supply.

In general, a linear synchronous motor propulsion system is composed of two components, one is the LSM and the other is the power supply. A linear synchronous motor [12] is a motor by which the mechanical motion is in synchronism with the magnetic field. The force is generated as an action of traveling magnetic field produced by a poly-phase winding and an array of magnetic poles or a variable reluctance ferromagnetic rail. To reduce cost, synchronous linear motors rarely use commutators, so the rotor often contains permanent magnets, or soft iron.

To minimize the vehicle's weight, the electric system interconnections should be optimized (e.g., use smaller diameter wires). In wiring modules, it is recommended to place the wiring to one side of the module rack, so that, the other side will be free and available for temporary hookups. LSMs are currently in use in amusement parks around the world as a method to launch roller coasters, and they have been proven to be reliable at quickly and smoothly accelerating large payloads hundreds of times a day.

A linear electric motor is a motor that has its stator and rotor unrolled (one part over the launcher aril, and one part attached under the UAV). So, instead of producing a torque, it generates a linear force along its length. In a linear synchronous motor, the mechanical motion is in synchronism with the magnetic field. LSM drives a load (here, a UAV) linearly without a need to gears and mechanical intermediates. Linear synchronous motors are the low-acceleration, high speed and high-power motors with an active winding on one side of the air-gap and an array of alternate-pole magnets on the other side. Figure 3 illustrates the free-body diagram of a synchronous linear motor.

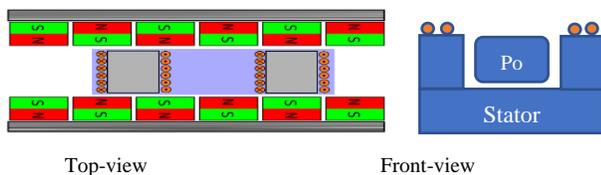


Figure 3. Free-body diagram of a linear synchronous motor

The UAV is lay attached to a cart on the rails and will accelerate to the desired speed. The track (rail) contains on-board exciting magnets for LSM. Flux from the exciting magnet interacts with the traveling magnetic wave from the stator to generate launch force. The launch force is generated as an action of traveling magnetic field produced by a poly-phase winding and an array of magnetic poles or a variable reluctance ferromagnetic rail. The part that generates the magnetic flux or variable reluctance is referred to as salient

pole rail. The part that generates the traveling magnetic field is referred to as the armature.

The length of required track is mainly a function of UAV weight, launch angle and launch power. LSMs benefit from high efficiency due to low magnetizing current and zero slip. This leads to a significant reduction of inverter rating, resulting in a substantial cost saving.

The LSM is selected as a source of generating launch thrust along the track. The model modeling of LSMs nonlinear, long, and complicated. The electromagnetic launch force developed by a LSM is obtained by:

$$F = \frac{P_{elm}}{u_s} \quad (8)$$

where  $P_{elm}$  is the electromagnetic power and  $u_s$  is the synchronous speed. The  $P_{elm}$  and  $u_s$  are functions of frequency of primary supply, the number of armature phase, rms value of the input voltage, and the rms value of armature current. The electromagnetic power and the synchronous speed are obtained by:

$$P_{elm} = mV_1 I_a \cos(\phi) - mR_1 I_a^2 \quad (9)$$

$$u_s = 2f\tau \quad (10)$$

where  $m$  represents the number of armature phase number,  $\cos(\phi)$  is the input power factor,  $I_a$  is the rms value of armature current,  $R_1$  is armature winding resistance, and  $V_1$  is the rms value of the input voltage,  $f$  is the frequency of primary supply, and  $\tau$  is the mover (pole) pitch. The armature current is a function of desired launch force ( $F_L$ ):

$$I_a = \frac{F_L u_s (\frac{1}{\eta} - 1)}{\rho_w l_w j_c} \quad (11)$$

where  $j_c$  is the amplitude of linear current density,  $\rho_w$  is the electrical resistivity of the primary windings, and  $l_w$  is the primary windings length. The parameter  $\eta$  is estimated in 0 - 1 range and will be finalized during the design. The two most generally used conductors (for wire windings) are copper and aluminum. Copper has a higher conductivity (about 40% more); is more ductile; has relatively high tensile strength; and can be easily soldered. Copper is more expensive and heavier than aluminum.

This launch system has a number of significant advantages as compared with conventional launchers. Loading UAV and setting them up would also take much less time and effort than other power systems.

### IV. OPTIMAL CONTROL

To optimize the launch operation, a closed-loop optimal control system is employed. The optimal control - Linear Quadratic Regulator (LQR) [13] - is based on the optimization of some specific performance criterion; or Performance Index;  $J$ . In this technique, no disturbance, noise, or uncertainty is considered.

We are interested in minimizing the error of the system; any deviation from equilibrium point is considered an error. To this end, an error-squared performance index is defined. For a system with one state variable,  $x_1$ , we have

$$J = \int_0^t [x_1(t)]^2 dt \quad (12)$$

An optimization technique for a dynamic system in state-space format is employed. The LQR is an optimal controller and is defined as follows. The system of interest is of the form:

$$\dot{x} = Ax + Bu \quad (13a)$$

$$y = Cx + Du \quad (13b)$$

Given the *weighting matrices* Q and R, the design task is to find the optimal control signal  $u(t)$  such that the quadratic cost function:

$$J = \frac{1}{2} \int_0^{\infty} (x^T Q x + u^T R u) dt \quad (14)$$

is minimized. The solution to this problem is:

$$u = -Kx \quad (15)$$

where

$$K = R^{-1} B^T P \quad (16)$$

and P is the unique, positive semi-definite solution to the Algebraic Riccati Equation (ARE):

$$PA + A^T P + Q - P B R^{-1} B^T P = 0 \quad (17)$$

Based on this technique, the LQR gains are calculated using a MATLAB code, and then a control system is designed. The engineering judgment skill must be utilized in the selection of Q and R. Tuning techniques are recommended in the determination of design parameters. For instance, Q and R must be such that the detectability (i.e.,  $(\sqrt{Q}, A)$  must be detectable) and observability requirements are met.

The LQR is a popular optimal control technique that has been successfully applied to control several UAV configurations. The LQR was successfully used [14] in stabilizing a Yamaha RMAX helicopter.

## V. LAUNCH CONTROL SYSTEM

The electromechanical mathematical model of the stage with loads is formulated by employing the governing equation presented in Section III. The equation of motion can be stated in terms of electric current (i) and distance travelled (x) as:

$$m\ddot{x}(t) = F_e - F_f - D \quad (18)$$

By inserting models of electromagnetic force, friction force, and drag into equation 18, we obtain:

$$m\dot{V}(t) = K_e i(t) - K_f \dot{x}(t) - K_D \dot{x}(t) \quad (19)$$

or,

$$\dot{V}(t) = \frac{K_e}{m} i(t) - \frac{K_f + K_D}{m} \dot{x}(t) \quad (20)$$

where  $F_e$  denotes the electromagnetic force (here, the launch force,  $F_L$ ),  $K_e$  represents thrust coefficient and the  $K_f$  and  $K_D$  denote friction and drag factors respectively. The inherent force ripple and the effects of the magnetic flux distortions and saturations are neglected for simplicity. The launch angle is assumed to be constant, and two parameters of velocity (V) and acceleration (a) are controlled. By reformatting equation 20, the state space representation of the launcher from is obtained:

$$\begin{bmatrix} \dot{x} \\ \dot{V} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{K_f + K_D}{m} \end{bmatrix} \begin{bmatrix} x \\ V \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{K_e}{m} \end{bmatrix} i \quad (21)$$

According to Equation (21), and the relation between linear acceleration and linear speed, the block diagram of the linear motor positioning system is illustrated in Figure 4.

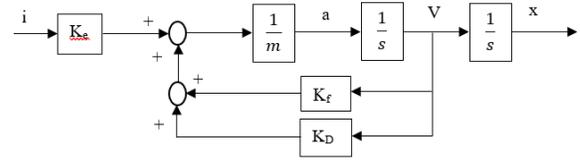


Figure 4. Block diagram of the electric launcher control system

In literature, various aspects of LSMs including their modeling, analysis, design, and control have been discussed. For instance, Ref. [5] has presented the design and characteristic analysis on the short-stator LSM for high-speed Maglev propulsion. Here, we use a linear model, since the objective of this paper is mainly to provide the effectiveness of hybrid launch system.

Two important motion parameters need to be controlled during a launch operation: 1. Velocity (V), 2. Acceleration (a). This objective requires a closed-loop feedback control system using an appropriate control law. The goal of control law is to have a near-constant acceleration at start of the launch but ease off slightly at the end. Proximity sensors should be installed at increments along the track to measure and report UAV position and speed to calculate the linear acceleration. For safety reasons, some other parameters such as the armature current (i) of LSMs may be controlled too.

The block diagram of closed-loop control system of the launch operation is shown in Figures 5. There are motion sensors for three outputs: 1. Position is measured, 2. Velocity is calculated by differentiation of position, and incorporating time of measurement, 3. Acceleration is calculated by differentiation of velocity, and incorporating time of measurement.

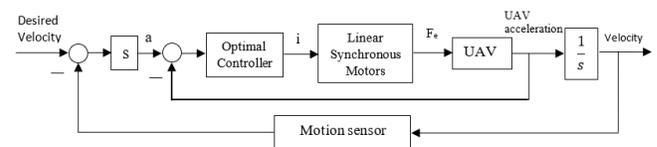


Figure 5. Block diagram of control system of the launch operation

The controller sends full-power signal to the motor when triggered. Due to the linear motion of the UAV along the track, an optimal control law will suffice in effectively controlling the launch operation. This dynamic system is modeled with two state variables, one input, and two outputs. The state space model is:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{K_f + K_D}{m} \end{bmatrix}, B = \begin{bmatrix} 0 \\ \frac{K_e}{m} \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (22)$$

The state and output variables are:

$$x = \begin{bmatrix} x \\ V \end{bmatrix} \text{ and } y = \begin{bmatrix} x \\ V \end{bmatrix}$$

The control signal  $u$  is given (Figure 6) by

$$u = k_1(r - x_1) - k_2 x_2 = k_1 r - (k_1 x_1 + k_2 x_2) \quad (23)$$

For a zero-reference input (i.e.,  $r = 0$ ):

$$u = -k_1 x_1 - k_2 x_2 = -Kx \tag{24}$$

To determine the state-feedback gain matrix  $K$ , where  $K = [k_1 \ k_2]$  (25)

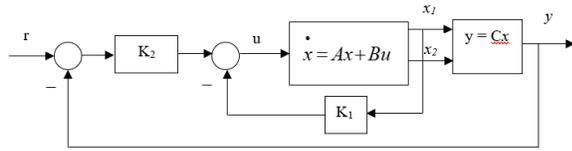


Figure 6. Quadratic optimal regulator system for the launcher

The modified state-space representation is developed by incorporating two feedbacks into the original plant:

$$\dot{x} = Ax + B(-Kx + k_1 r) = (A - BK)x + Bk_1 r \tag{26}$$

Therefore, the revised state-space model matrices are:

$$A1 = A - B*K; B1 = B*K(1); C1 = C; D1 = D \tag{27}$$

The new state-space model can be used in the simulation to determine the behavior of the system. Two objectives of control are: 1. Keep the acceleration below 7g, 2. Keep the launch speed to follow the desired velocity profile. UAV weight plays a major role in control parameters due to incline angle. The controller is generating a signal to the actuators (i.e., LSM) based on the control law. The LQR gains are functions of LSM features and UAV weight. Tabular values for LQR gains may be determined for various UAVs weights.

The main output of the LQR controller is the rms value of armature current ( $I_a$  or just  $i$ ), which will be the input to the LSM. The main output of the LSM and the track is the launch force ( $F$ ), which will be the input to the UAV. To reduce “jerk”, we need to have an overdamped velocity profile. To have a successful launch, the wind speed and direction should be measured at the launch site and incorporated in the control process.

## VI. SIMULATION AND RESULTS

In order to validate the design outcome, a launch system including a linear model of LSM for launching a UAV with a mass of 20 kg has been simulated by matlab Simulink. It is desired that the UAV reaches the velocity of 10 m/sec along the track before the end of launch operation.

The simulation is presented to demonstrate the efficacy of the proposed launch system with the control algorithm. Figures 7 through 11 illustrate the simulation results for: 1. LSMs current in Amps, 2. Velocity of UAV in m/sec, 3. Vehicle non-dimensionalized acceleration, 4. UAV displacement in meter, 5. Force generated by LSM in N, and 6. Electric power provided to LSMs in W, respectively. It is assumed that voltage for LSMs is 120 Volts.

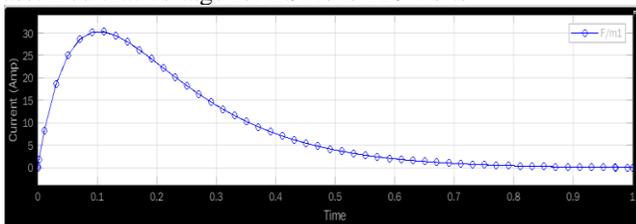


Figure 7. Variations of LSM current (in Amps)

From Figure 7, the maximum current is about 30 Amps at 0.1 seconds to the launch. When the UAV reaches the desired velocity, the LSM current is reduced to almost zero.

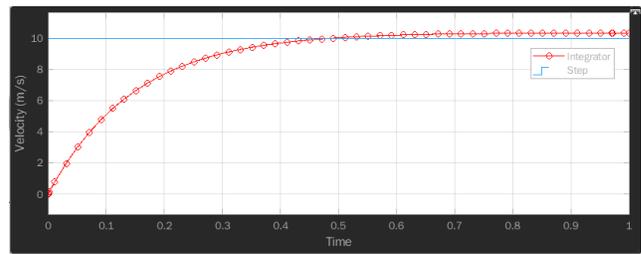


Figure 8. Variations of UAV velocity (in m/sec)

In Figure 8, the desired velocity of 10 m/sec (blue line) is a given value and shown. It can be seen that the UAV reaches this velocity in about 0.5 seconds (intersection of red graph with blue line) and continue to increase to about 10.5 m/sec due to the UAV linear momentum. Since no brake is considered in the launch system, the velocity is not reduced back to 10 m/sec during launch.

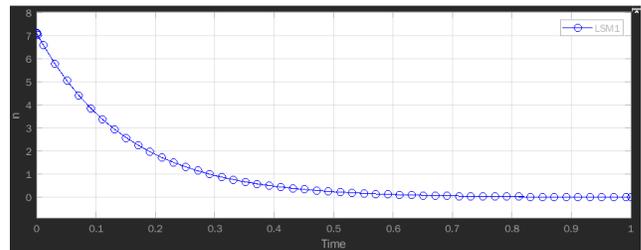


Figure 9. Variations of normalized acceleration

As shown in Figure 9, the maximum non-dimensionalized acceleration, i.e., (in g) is about 7, it happens in the beginning of launch. When the vehicle reaches the desired velocity, the acceleration will become zero.

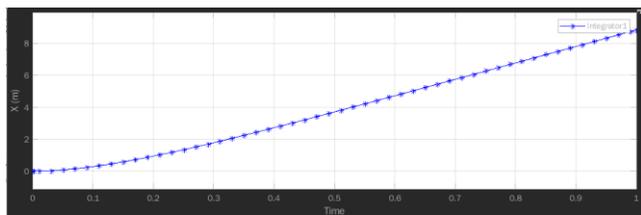


Figure 10. Variations of vehicle displacement (in m)

As shown in Figure 10, after 1 second, the vehicle has displaced about 9 m. From this part of the simulation, it is concluded that a launch track with the length of 9 meters is required to launch a UAV with a mass of 20 kg. To have a launch site for UAVs with various masses, a longer launch track should be constructed.

As Figure 11 indicates, the maximum force generated by LSM is 1400 N at the beginning of launch. When the vehicle reaches the desired velocity, this force will become almost zero. Afterward, a small amount of force is needed to cover the track friction and UAV drag.

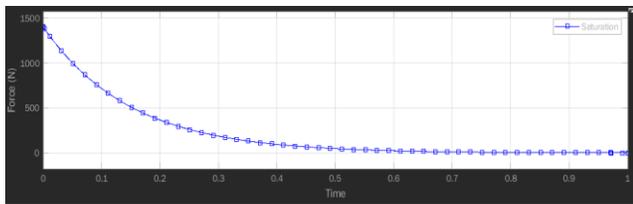


Figure 11. Variations of the LSM generated force

The maximum electric power provided to LSMs is 3.5 kW after about 0.1 seconds. The engine of the vehicle should be started when the vehicle reaches the desired velocity (i.e., before UAV reaches the end of launch track). However, if the UAV own engine is operating along the track (concurrent with LSM), the vehicle will have a much faster velocity at the end of the launch operation.

By examining the simulation results, one can conclude that launch operation utilizing electric power is feasible. Using the LQR controller, the UAV is tracking and following the desired velocity on the track. For other UAVs with different weights, the LQR gains should be adjusted/changed. As the simulation results indicate, the launch operations using LSM is successful. Since the vehicle is along the ramp and uses a special rail connection, the UAV is stable and reliable.

Moreover, this system can be employed over and over again for various UAVs, so the overall launch cost is much lower than a vertical launch. The investment for the track and LSM is for a long run application. Furthermore, there is no possibility of explosion by UAV engines, since the launch power system is of electric type.

## VII. CONCLUSION AND FUTURE WORK

This paper presented a new launch system for fixed-wing UAVs powered by linear synchronous motors and a special track. This novel idea has the potential to be developed to a new technique for launching fixed-wing UAVs. In this paper, LSM was recommended as the source of generating launch force. The control of launch process was conducted utilizing an optimal controller. The most notable advantages of this technique are lower cost, higher reliability, stability, and safety. A difficulty in this technique is a relatively heavy equipment. For the future work, the next technical step would be to focus on reducing weight of the power system and improving the efficiency of this electric launcher.

## REFERENCES

[1] Z. Xiaolong, B. Cheryl, T. O'Connell, and K. Haran, "Large electric machines for aircraft electric Propulsion", *IET Hybrid Propulsion for Aviation*, Vol. 12 Issue 6, 2018, pp 767-779.

[2] B. Sarlioglu and C. T. Morris, "More electric aircraft: review, challenges, and opportunities for commercial transport aircraft", *IEEE Transactions on Transportation Electrification*, 2015, 1, pp. 54-64

[3] J. F. Gieras, J. P. Zbigniew, and T. Bronislaw, "Linear synchronous motors: transportation and automation systems", CRC Press, 2018

[4] R. S. Fattahpour and A. Shiri, "A New Method for Designing DC-Excited Linear Synchronous Motor", *11<sup>th</sup> Power Electronics, Drive Systems, and Technologies Conference*, 2020, IEEE Xplore, pp 1-6

[5] H. Cho, H. Sung, S. Sung, D. You, and S. Jang, "Design and Characteristic Analysis on the Short-Stator Linear Synchronous Motor for High-Speed Maglev Propulsion", *IEEE Transactions on Magnetics*, Vol. 44, No. 11, pp. 4369-4372, Nov 2008

[6] H. Ohsaki, "Linear Drive Systems for Urban Transportation in Japan, Proceedings of the Maglev Conference", Yamanashi, Japan, April 1998, pp. 29

[7] M. Mossi and P. S. Rossel, "A revolution in the high-speed passenger transport systems", In *Proceedings of the 1<sup>st</sup> Transport Research Conference*, pp 1-16, Ascona, Switzerland, 1-3 March 2001

[8] A. Stilwell, "19 For 99: Incredible Hulk Coaster at Universal's Islands", July 4, 2019

[9] M. D. Adamski, R. Root Jr., and A. M. Watts, "UAV recovery system", US Patent US7219856 B2, 2007

[10] Scan Eagle fact sheet and specification, Insitu.com, Insitu, A Boeing company, 2020

[11] M. H. Sadraey, "Aircraft Performance Analysis: An Engineering Approach", CRC Press, 2017

[12] R. J. Kaye and E. Masada, "Comparison of Linear Synchronous and Induction Motors", *Urban Maglev Technology Development Program*, Sandia National Laboratories, FTA-DC-26-7002.2004.01, June 2004

[13] B. D. O. Anderson and J. B. Moore, "Optimal Control - Linear Quadratic Methods"; Dover, 2007

[14] F. Kendoul, "Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems", *Journal of Field Robotics*, vol. 29, no. 2, 2012, pp 315-378

# Traffic Signal Recognition and Application Algorithm for the Autonomous Vehicle in V2X Unavailable Areas

Yejin Gu

Department of ICT Convergence R&D  
Korea Automotive Technology Institute  
Cheonan, Republic of Korea  
e-mail: yjgu@katech.re.kr

Daejun Kang

Department of ICT Convergence R&D  
Korea Automotive Technology Institute  
Cheonan, Republic of Korea  
e-mail: dj kang@katech.re.kr

**Abstract**— Recognizing traffic signals is essential to operate autonomous driving on urban roads. The V2I is the priority in recognizing traffic signals in autonomous vehicles because it can send the exact Signal Phase and Timing (SPaT) of the traffic signal. However, there are many V2I unable areas currently, and even if it is available, it might have trouble due to communication delay or fail-operation. Accordingly, we present the traffic signals recognition framework using the convergence of Vehicle-to-Infrastructure(V2I) and camera detection to complement V2I. This study uses the signal recognized by the camera sensor in the area when the V2I signal is unable. By changing the existing one-way decision-making method, we implemented a customized communication system that dynamically changes in real time and fits the infrastructure situation. We identified that the proposed method works well by deploying an autonomous driving pilot system on the designated segment in Sejong-si, South Korea, where the V2I is partially available.

**Keywords**—safety and traffic efficiency applications; object detection; C-V2X; image processing; CNN; deep learning

## I. INTRODUCTION

Recognizing the driving environment is critical autonomous vehicles [1][2]. In particular, to operate autonomous vehicles on city roads, it is essential to recognize signals such as intersections. One method for recognizing the traffic signal is the vehicle to infrastructure (V2I) communication to receive traffic signals. Since accurate information is required for autonomous driving, Signal Phase and Timing (SPaT) information is preferentially used through V2I. However, there are V2I unable areas due to communication errors or insufficient system construction. Furthermore, though V2I is available, autonomous vehicles might have trouble recognizing traffic signals via V2I due to communication delay or breakdown. Accordingly, it is essential to complement recognizing traffic signals by using other sensors because of fail-safety.

In this paper, a camera is mainly used to complement V2I. In many studies, traffic signal recognition generally defines pre-segment a scene to find a region of interest. Ruta et al. [3] propose a general detector purification procedure based on mean moving clustering. Heuristics, in which traffic signal recognition utilizes prior knowledge of color and shape, have also been developed to recognize traffic signals [4][5].

However, it is challenging to process real-time camera-based signal recognition in real environments. Therefore, real-time image processing was performed using TenorRT [6], which can achieve fast computation speed with excellent performance. In addition, it is necessary to transmit accurate signal information to autonomous vehicles in a section where there is no V2I communication. Existing research showed the accuracy of signal recognition based on dataset, but it is not possible to show test results that are processed in real-time by being mounted on an actual autonomous vehicle [7][8]. In addition, research has not been conducted on the application of autonomous driving systems by fusion with V2I information in actual road areas.

In Section 1, a real-time signal recognition inference engine was implemented engine, and in Section 2, it was mounted on the system and fused with V2I in the real driving environment and applied to the autonomous driving system. The real-time video was taken with one camera, and the traffic light area was decided using computer vision-based technology.

## II. METHODOLOGY

The dataset used in this study was collected from Sejong City, South Korea. Therefore, we propose an algorithm to detect a traffic signal in an image frame through a real-road environment as input, as shown in Fig.1 The camera was installed in the center of the vehicle's dashboard. The image is saved as a JPEG with RGB values of 2048x1536 pixels using a Blackfly S USB3. All images were captured under natural light conditions, such as variations in sunlight and complexity of the background. In particular, these conditions greatly increase the difficulty of detecting traffic lights in the field.



Figure 1. An example from the real-world road dataset. (a) The input camera image (b) The traffic light detection result.

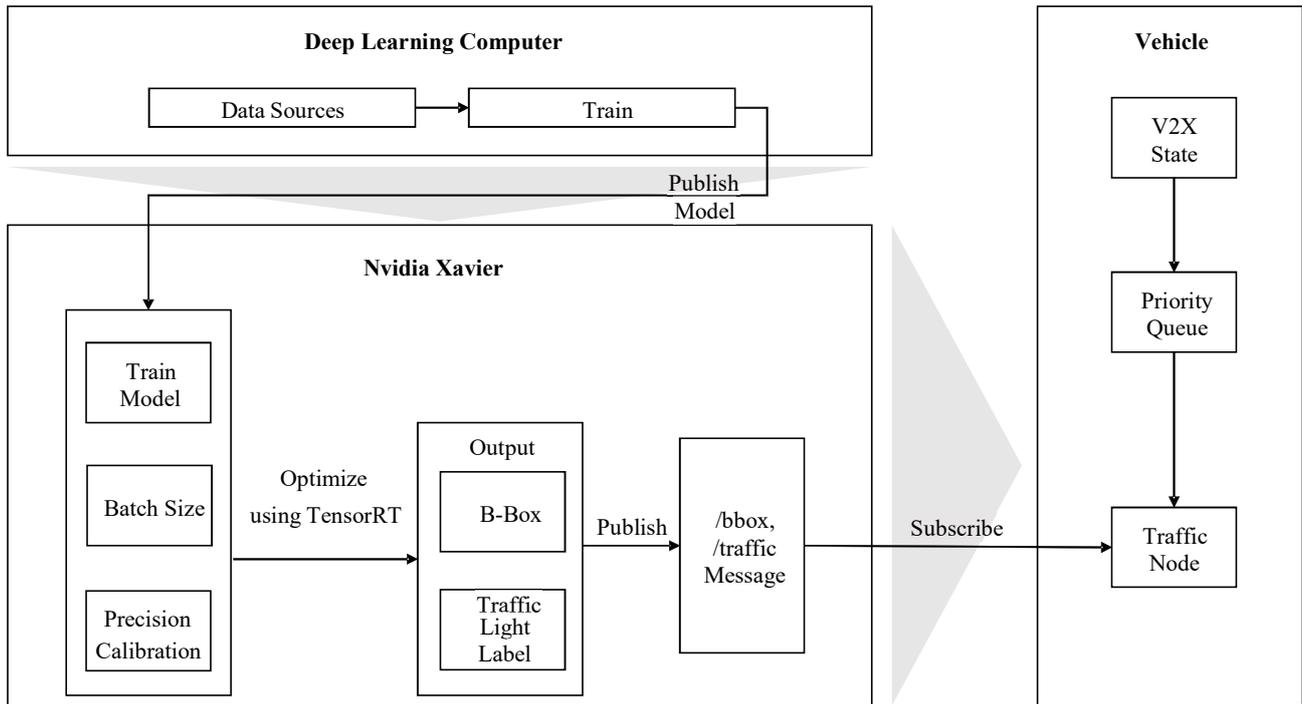


Figure 2. The overall pipeline of traffic light recognition system. Batch size : Total number of training examples present in a single batch. Precision : Precision calibration. B-Box : bounding box center offset and scales

The Figure 2 shows the overall flowchart of the training and detection process for constructing an applicable detection model in this study. Our training model is reliable, fast, capable of detecting signals of various scales, and generalizes to new data sets obtained from the training set in the real world.

#### A. Data Construction

Labeled data requires class labels and locations for all ground-truth bounding boxes in the training image. In order to apply deep learning, traffic lights in the real driving environment were manually extracted and identified. Therefore, the dataset we collected using the camera is 2M images annotated with labels and bounding boxes. The model used in this study can detect four object classes {stop, go, turn, turn&go}. The trained model was tested using an image resolution of 832x832 pixels with the batch size set to 32 for consistency with the training image resolution. Because of the small size of the traffic lights in the whole image, it can be difficult to recognize the correct category, so you need to resize it to the appropriate image size.

#### B. Traffic recognition system

We next describe the model used for traffic light recognition and how it fuses with V2I in real-time.

##### 1) Detection model

To detect object in an image, we use an algorithm that detects in a specific area instead of processing the entire image in high resolution. Each grid cell is an image divided into squares. Slide a rectangular cell over the image to detect and classify objects. Then, each value calculated in the cell is analyzed for significant differences between the different cells, and the bounding area is calculated, outputting a bounding box. Each box is given as a tuple  $(x, y, w, h)$  and a confidence measure, where  $(x, y)$  is the center of the predicted box relative to the cell boundary, and  $(w, h)$  is the width and height of the bounding box relative to the image size. The neural network computes the features of the extracted object regions and classifies them into categories. This classification is scored on a predefined scale in the network based on the scores of the pre-classified objects. We use a model architecture that

directly predicts the object bounding box of an image in a one-stage model.

Raw images are sent from the camera to the computing system NVIDIA Xavier for processing. During training, the input to the model is a 3 channel color image with an annotated bounding box around each traffic light. Image data is propagated through multiple convolutional layers. In order to define the boundary of the object in the learning model, a grid size 832x832 is first defined. Then an object prediction is performed for each cell. A vector of size  $(C+5) \cdot B$  is created for each grid element, where  $B$  defines the number of masks in one layer and  $C$  defines the number of classes. Parameter 5 shows the center coordinates of the grid cell inner boundary, the height and width, and the probability that the boundary is defined correctly. This results in a prediction bounding box of

size 106x106xB. The framework for the traffic light detection can be seen in Figure 3.

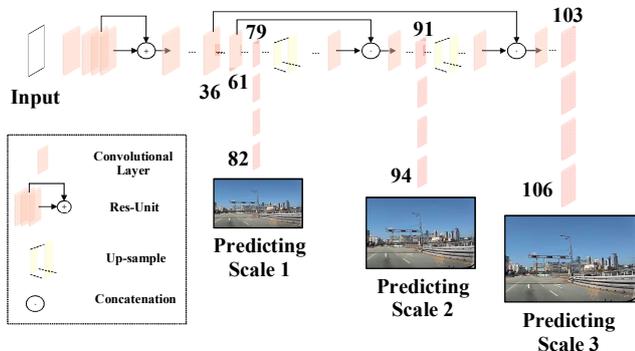


Figure 3. Framework for the traffic light detection

### 2) Real-time Inference Engine (model acceleration)

Optimization of the neural network architecture can be used to speed up processing and keep the accuracy at the same level. There are ways to speed up neural network processing without spending a lot of time changing programs. TensorRT is a platform that uses algorithms to optimize architectures and speed up neural network processing using algorithms that use the power of NVIDIA GPUs to increase computation. This is convenient in many cases because it can speed up programs without spending a lot of resources changing code. The process of using TensorRT to create a network for Xavier is shown in Figure 4.

TensorRT analyzes graphs that represent network models. If there are repeating elements in the graph, TensorRT merges them. As a result, the network size becomes smaller.

### 3) Traffic Signal Recognition using the Convergence of V2I and Camera Detection

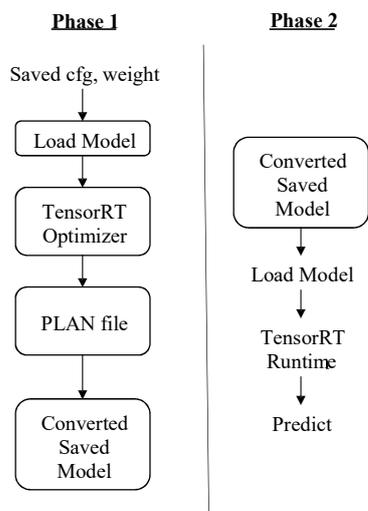


Figure 4. A flow of the network creating on TensorRT

In the proposed system, traffic lights on the actual road were captured by a camera attached to the vehicle. For object detection, a camera captures an object and then extracts features and learns it from a deep learning model. When a traffic light is identified, it is waiting in the form of a Robot Operating System (ROS) topic message so that data can be transmitted via ROS. In general, the braking distance information consists of distance-speed calculation, and the existing braking distance calculation map is presented in Table 1 [9]. Between 30 and 40 km/h, the braking distance is 10 to 18m, and according to equation (1), the response time is 1.38 to 2.5 seconds. In this study, 1.5 seconds is used as the required response time because the vehicle is traveling at 30 km/h approximately. If the V2I data is blank for 1.5 seconds, the priority is switched to the image detector, and the current signal state can be determined by receiving the image detector's ROS topic.

TABLE I. BRAKING DISTANCE CALCULATION MAP

| Speed (km/h) | Braking distance (m) | Needed Response Time (s) |
|--------------|----------------------|--------------------------|
| 30           | 10                   | 2.5                      |
| 40           | 18                   | 1.38                     |

$$Needed\ Response\ Time = \frac{Braking\ distance}{Speed\ of\ Vehicle} \quad (1)$$

The raw image of camera array C, the distance D from the vehicle to the traffic light transmitted via V2I communication, is given as input; the data transmission period is represented by the tick variable. As long as image data exists, the algorithm iterates. Traffic lights are detected using a pre-trained model from the input image data. Traffic light information including the location and label of the bounding box is created and published in the form of a ROS topic. If data is not input for more than 1.5 seconds, the result derived from the learning model is subscribed to the vehicle.

The traffic light state communication algorithm is shown in Figure 5.

### III. IMPLEMENTATION DETAILS

In this study, a TensorRT model was loaded on NVIDIA Jetson AGX Xavier, a powerful inference engine with JetPack 4.6 provided by NVIDIA. CUDA 10.2, cuDNN 8.2.1, TensorRT 8.0.1, and OpenCV 3.4.0 are required for TensorRT to work properly. The system uses the Python programming language. Also, a file containing the trained model weights (weights) and network configuration(cfg) is needed to create the plan file. We found that there is a significant difference between the number of frames in Xavier and the number of frames when TensorRT is applied.

**Algorithm 1** Traffic Light State Communication Algorithm

```

1: Data: raw camera image in  $C$ ;
2:   stand-by time  $tick$ ;
3: Result: traffic light state topic
4: initialization;
5: while  $C$  is not empty do
6:   detect traffic light;
7:   publish traffic light state topic (Xavier);
8:   Calculate ticktime;
9:   if  $tick > 1.5$  then
10:    subscribe traffic light state topic (Xavier  $\rightarrow$  Vehicle);
11:  end
12: end
    
```

Figure 5. Traffic Light State Communication Algorithm

Table 1 compares the hardware specifications of the NVIDIA developer kit. NVIDIA AGX Xavier is powered by the new NVIDIA Xavier processor and delivers over 20x performance and 10x energy efficiency, especially over the NVIDIA Jetson TX2.

As shown in Table 1, the number of real-time video frames was measured. The number of video frames is one of the important factors that can affect real-time image processing. According to the test results, when TensorRT is applied, 14.9 frames are compared to 6.6 frames when TensorRT is not applied. When TensorRT is a difference of about 2.25 times compared to when it is not applied. The accuracy of detection did not decrease after using TensorRT.

TABLE II. HARDWARE SPECIFICATION

| specification | Jetson AGX Xavier                                     | Jetson Nano                                | Jetson TX2  |
|---------------|---|--|---|
| GPU           | 512 NVIDIA CUDA Cores and 64 Tensor Cores             | 128 NVIDIA CUDA Cores                      | 256 cores NVIDIA Pascal GPU   |
| CPU           | 8 cores NVIDIA Carmel Armv8.2 64bit CPU 8MB L2+4MB L3 | Quad cores ARM Cortex-A57 MPCore Processor | Dual cores Denver 2 64 bit CPU & Quad cores Arm Cortex-A57 MPCore Processor |
| Memory        | 32GB 256bit LPDDR4x 136.5GB/s                         | 4GB 64bit LPDDR4                           | 8GB 128bit LPDDR4 59.7GB/s  |

TABLE III. FRAME DIFFERENCE BETWEEN STANDARD AND TENSORRT

| type             | FPS  |
|------------------|------|
| Without tensorRT | 6.6  |
| Adjust tensorRT  | 14.9 |

An accelerated engine model is generated and analyzed to determine one of four classes to quickly and accurately

determine traffic light object recognition in a vehicle in real-time. A NVIDIA Xavier equipped with an algorithm is implemented in the vehicle and transmits the traffic light recognition result to the main server that controls the entire vehicle using ROS topic message. In this way, it can be easily applied to other vehicles, resulting in high scalability. As a result of the experiment, it was found that the proposed system provides accurate traffic signal information in the unabled section of V2I in the city image with a complex background in Figure 6.

## IV. CONCLUSION

In this study, we proposed a traffic light recognition algorithm that applied TensorRT to minimize image resource usage in the section where V2I is unabled. Performance was analyzed based on real-world scenarios tested in neural networks. To compensate for sections without V2I using camera sensors, we used a deep learning architecture based on Convolution Neural Network (CNN) to label traffic lights. As a result of our experiments, our algorithm was able to accelerate in terms of speed with the same accuracy. This function was available in real-time driving, where processing speed is an important value. Our method is simple but effective, and yields significant improvements in demanding road traffic object detection and image classification datasets.

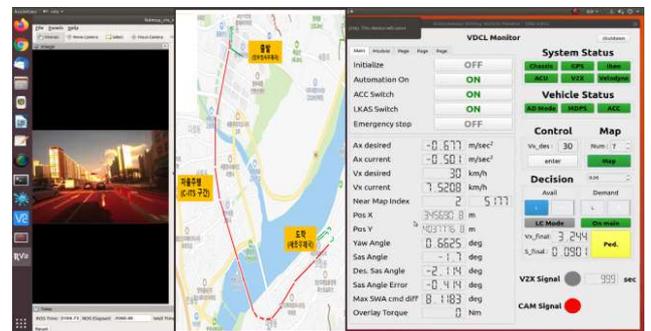


Figure 6. Experiment driving and Trajectories

## ACKNOWLEDGMENT

This work was supported in part by the Development of chip level solid-state sensor using nano-photonics structure for autonomous driving under grant 2020485, funded by the Ministry of Trade, Industry and Energy.

## REFERENCES

- [1] M. Priisalu, A. Pirinen, C. Paduraru, and C. Sminchisescu, "Generating scenarios with diverse pedestrian behaviors for autonomous vehicle testing," *Conference on Robot Learning*, PMLR, pp. 1247-1258, 2022.
- [2] R. Greer, J. Isa, N. Deo, A. Rangesh, and M. M. Trivedi, "On Saliency-Sensitive Sign Classification in Autonomous Vehicle Path Planning: Experimental Explorations with a Novel

- Dataset,” *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 636-644, 2022.
- [3] A. Ruta, F. Porikli, S. Watanabe, and Y. Li, “In-vehicle camera traffic sign detection and recognition,” *Machine Vision and Applications*, vol.22.2, pp. 359-375, 2011.
- [4] G. Piccioli, E. De Micheli, P. Parodi, and M. Campani, “Robust method for road sign detection and recognition,” *Image and Vision Computing*, vol. 14.3, pp. 209-223, 1996.
- [5] G. Loy, N. Barnes, D. Shaw, and A. Robles-Kelly, “Regular polygon detection,” *Proc. of the 10th IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 778-785, 2005.
- [6] H. Vanholder, “Efficient inference with tensorsrt,” *GPU Technology Conference*, vol. 1, pp. 2, 2016.
- [7] M. Omachi and S. Omachi, “Traffic light detection with color and edge information,” *2009 2nd IEEE International Conference on Computer Science and Information Technology*, IEEE, pp. 284-287, 2009.
- [8] M. P. Philipsen, M. B. Jensen, A. Møgelmoose, T. B. Moeslund, and M. M. Trivedi, “Traffic light detection: A learning algorithm and evaluations on challenging dataset,” *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE, pp. 2341-2345, 2015.
- [9] P. Greibe, “Braking distance, friction and behaviour,” *Trafitec, Scion-DTU*, 2007.

# A Quantitative Measure for the Evaluation of Drone-Based Video Quality on a Target

Daniela Doroftei  
*Department of Mechanics*  
*Royal Military Academy*  
 Brussels, Belgium  
 daniela.doroftei@rma.ac.be

Geert De Cubber  
*Department of Mechanics*  
*Royal Military Academy*  
 Brussels, Belgium  
 geert.de.cubber@rma.ac.be

Hans De Smet  
*Department of Economy, Management and Leadership*  
*Royal Military Academy*  
 Brussels, Belgium  
 hans.de.smet@rma.ac.be

**Abstract**—For the evaluation of drone operators, it is important to assess their capability to produce high-quality video of a certain object. However, traditional video quality assessment methodologies are in general more geared towards video compression and thus focus on the correct image representation, and not on the real content of the produced data. In this paper, we therefore propose a methodology to define a video quality metric, specifically geared towards drone operations. Using this quantitative measure as a baseline, we also propose a methodology which proposes optimal drone trajectories for obtaining a maximum amount of qualitative video data about a given object or target in a minimum amount of time. The proposed methodologies are validated within a virtual pilot training environment.

**Index Terms**—video production; quality assessment; drones; unmanned aircraft systems; situational awareness

## I. INTRODUCTION

Drones are used nowadays for a plenitude of tasks, including often missions where the drone operator is required to produce a high-quality video of a certain object or target. These kind of operations can range from wedding photography over search and rescue [1] or bridge inspection [2] to military operations [3]. Obviously, in order to fulfil such tasks in a proper way, the drone operator requires a specific form of training and skills development. Moreover, for mission-critical applications, it is essential to assess on beforehand that the drone operator has a sufficient level of skills with respect to this task. However, to date the quantitative evaluation of this drone pilot training expertise remains problematic, as it is not really possible to quantify the quality of the piloting skills, notably related to the capability of producing high-quality video data of a given target.

Indeed, as we will develop in Section 2 on the state of the art, there does not really exist a tool which allows to tell whether a video produced by a certain drone operator contains enough information about a certain target or not. Therefore, we propose in this paper a methodology to quantitatively assess the content of drone-based video data.

It should be noted that the methodology, which is explained in detail in Section 3 of this paper, is *not* performed by an analysis of the video signal, as this would render the approach very difficult to port from one type of application or mission

scenario to another. Instead, the methodology is based on the analysis of the position data, which drones typically receive via their positioning sensors. As such, the methodology is task-agnostic and can be applied to a wide range of applications.

We do focus in this research study mostly on military operations, where aim is to gather a maximum amount of data about a target in a minimum amount of time. A drawback of this choice as an application is that the proposed approach ignores cinematographic constraints (rule of thirds, etc.) as they are commonly used for professional video photography, which makes it less useful for these kinds of applications.

In the fourth section of this paper, we show how the proposed evaluation metric can be used inside an optimization scheme in order to automatically generate drone trajectories that maximize the amount of high-quality video data obtained from a certain target.

In the fifth section of this paper, we validate the proposed methodologies in two use cases, highlighting the novel contributions of this paper:

- A methodology for content-based video-quality analysis, used for the assessment of the performance of drone operators
- A methodology for the automatic generation of optimal drone trajectories for maximizing the information gathered of a certain subject

In relation to these validation experiments, it should be noted that the quantitative drone-based video quality assessment methodology presented in this paper is not an isolated development. It is developed within the framework of the ALPHONSE project by Belgian Defence [4], which has as a goal to develop a virtual training environment for the training of drone pilots of security services (e.g., police, firefighters, civil protection, military) and to study the human factors that influence the performance of these operatives. Within the ALPHONSE training environment, drone operators perform regular virtual training missions and the goal is to track their performance related to high-quality video production with the tools presented in this paper.

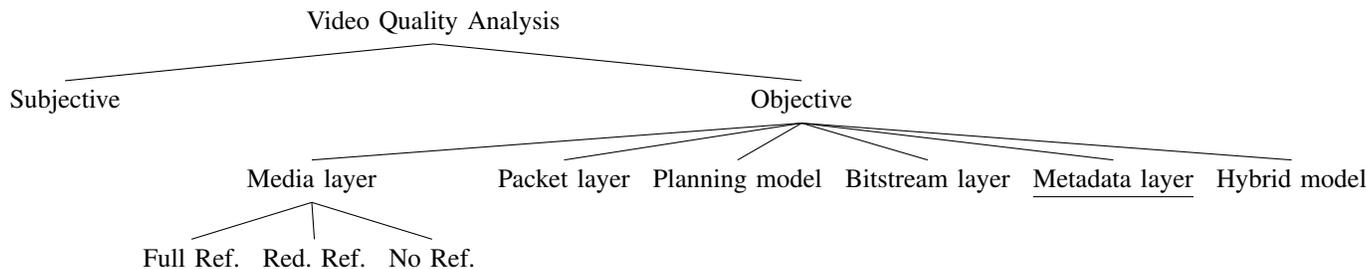


Fig. 1: Taxonomy of video quality analysis methodologies.

## II. RELATED WORK

Video quality analysis methodologies can be generically classified into two different categories. On the one hand there are the subjective video quality analysis methodologies [5], which directly express the video quality as experienced by humans. In order to do so, subjective video analysis methodologies require the video sequences to be shown to groups of viewers. The subjective opinion of the audience is recorded and averaged into a so-called Mean Opinion Score (MOS) to evaluate the quality of the video sequence. While subjective video analysis methods render excellent results, the problem is that they are extremely labour-intensive and therefore difficult to deploy in a practical context. As a result, more objective - and therefore also more automated - have been developed.

Objective video quality analysis methods have been classified by the International Telecommunications Union (ITU), according to the input data employed by the algorithms [6].

*Media layer models* directly use the video signal to define a quality measure. These methods do not require a priori information about the system under test and are therefore often used for the comparison of different compression methodologies. Depending on the type of source data the processed video is compared with, 3 types of submethods can be further identified:

- Full-reference methods extract data from high-quality undegraded source signals. This is the type of methodology (often a PSNR-derivative [7]), which is used for video codec evaluation.
- Reduced-reference methods extract data from a side-channel with signal parameter data.
- No-reference methods do not employ any source information.

*Parametric packet-layer models* estimate the video quality using only information from the signal packet header.

*Parametric planning models* use quality planning parameters for networks as an input in order to estimate a video quality measure.

*Bitstream layer models* estimate the video quality by combining encoded bitstream information with packet-layer information.

*Hybrid models* use a combination of the previously discussed methodologies in order to estimate a measure for the video quality.

All these traditional methodologies for video quality analysis estimate the video quality based upon the assumption that there exists a perfect input signal, which is then degraded, due to encoding, network transmission, decoding, display constraints, etc. In our application, this is not really the case: the question we are interested in is more whether a certain subject has been perceived sufficiently within the video material. As such, it is more a content-based analysis which is required.

Video content-based planning for drones has been shown before by Hulens and Goedemé in [8]. Within this paper, an autonomous drone is presented that automatically adjusts its position in order to keep a subject (in this application an interviewee) within view under certain cinematographic constraints. In comparison with our research question, where it is the aim to maximize the information gain about generic subjects, this approach is very much geared towards one application (the subjects are always human faces).

Building upon the ITU taxonomy for objective video quality analysis methods (see Fig. 1), we therefore introduce a new category of models which are based upon a processing of the *metadata layer*. Indeed, drones typically have accurate GPS sensors on-board, enabling to geo-localize all image and video data produced by these systems. As we will develop in the following section, we propose a methodology which uses this meta-data to estimate a content-based video quality metric.

Obviously, the new metadata-layer ignores important aspects of the video quality analysis paradigm, as it does not consider any errors that may be induced in the *encoding - transmission - decoding - display* - pipeline. We acknowledge this and advocate that it is - in a realistic deployment - probably the best idea to incorporate the proposed metadata layer model together with another model in a hybrid architecture. However, in this paper, we will focus fully on the elaboration and validation of the metadata layer model by itself.

Drone trajectory optimization is a research field that has received a lot of attention in recent years, as researchers have started employing drones for a wide range of applications [9], [10]. In its essence, this problem boils down to a constrained optimisation problem, as there is an objective function (e.g., a number of targets that need to be reached) that is to be minimized, while taking into consideration the constraints imposed by the flight dynamics of the drone. This is the same in this paper, where we employ in Section 4 the proposed video quality metric as a basis for the optimization function.

### III. METHODOLOGY TOWARDS THE QUANTITATIVE EVALUATION OF DRONE-BASED VIDEOS

The video quality analysis methodology presented here is developed to be as task-agnostic as possible. However, this makes it necessary to define some key basic assumptions made by the algorithm.

- We assume that the drone camera is always looking straight at the target. This assumption is made in order not to over-complicate the algorithm with (dynamic) viewpoint changes in function of the drone movement. This is a realistic assumption, as in real operations, it is very often the case that there is a separate camera gimbal operator, whose task it is to point the camera at the target. Obviously, this is a task that can be automated as well using visual servoing methodologies [11]. In this paper, we assume that such a method has been implemented.
- In order to ensure that the target object is perceived equally from any different viewing angle, we assume that the target has a perfect spherical shape. This will be obviously an approximation, and for realistic objects with a very different shape, it may lead to a different evaluation. However, this is the most generic assumption which can be made, and it can - if required - be further refined if certain specific target shapes are better suited for specific applications.
- As the zoom factor is a piece of information which is not dynamically available to the algorithm, we assume here a static zoom factor.
- The sole input parameters used by the video quality assessment algorithm are the position of the drone at a certain time instance  $\mathbf{x}_i = (x_i, y_i, z_i)$  and the position of the target  $\mathbf{x}_t = (x_t, y_t, z_t)$ , which is assumed to be static during the entire video sequence.

The methodology proposed here towards quantitative video quality analysis considers three sub-criteria, which determine together the overall measure of video quality. These three metrics are:

- 1) *The number of pixels on target*  $\phi_p$ . It is well-known that for machine vision image interpretation algorithms (e.g., human detection [12], vessel detection [13]), the number of pixels on target is a key parameter to predict the success of the image interpretation algorithm. Also, for human image interpretation, the so-called Johnson's criteria [14] state clearly that the ability of human observers to perform visual tasks (detection, recognition, identification) is a function of the image resolution on a target.

The number of pixels on target is obviously correlated to the zoom factor of the camera. However, as the zoom factor is assumed to be constant, the number of pixels on target is inversely proportional to the distance between the target and the drone, such that:

$$\phi_p = \frac{\lambda}{|\mathbf{x}_i \mathbf{x}_t|}, \quad (1)$$

where  $\lambda$  is a constant parameter ensuring that  $0 \leq \phi_p \leq 1$ . The parameter  $\lambda$  is dependent on the minimum distance between the drone and the target, the resolution of the camera and the focal length.

- 2) *The data innovation*  $\phi_d$ . As expressed in the introduction, we want to assess the capability of drone operators to obtain a maximum amount of information about a given target in a minimum amount of time. What is therefore very important is that the operators are able to produce high-quality *new* video data of a target. The data innovation metric is there to evaluate this innovation quality.

This is performed by building up a viewpoint history memory  $\theta_j$  with  $j = 1 \dots i-1$ , which contains a memory of all normalized incident angles of previous viewpoints. The current incident angle  $\theta_i$  is then compared to this memory. In practice this is done by taking the norm of the difference between the current incident angle and each of the previous incident angles. The data innovation is then equal to the smallest of these norms, as this represents the distance to the closest viewpoint on a unit sphere:

$$\phi_d = \min_{j=1}^{i-1} (|\theta_i - \theta_j|) \quad (2)$$

As the idea is to generate as much as possible new data, the new viewpoint  $\theta_i$  should be as far away as possible from existing viewpoints, which is expressed by (2).

- 3) *The trajectory smoothness*  $\phi_t$ . In order to achieve a high-quality video, it is important that the trajectory of the drone is smooth over time. Indeed, if the drone follows an irregular motion pattern, then the resulting video signal would be hard to interpret by human operators or by machine vision algorithms. The metric  $\phi_t$  therefore evaluates the trajectory smoothness, by building up a velocity profile  $\dot{\mathbf{x}}_j$  with  $j = 1 \dots i-1$ , which contains a memory of all velocities at previous time instances. The current drone velocity  $\dot{\mathbf{x}}_i$  is then compared to the  $n$  most recent iterations in this velocity memory. This is done by taking the norm of the difference between the current velocity and each of the previous velocities. In order to make more recent data count more in the evaluation, this norm is weighted according to the recency of the information. The weighted and normed sum of the  $n$  most recent velocity differences is a measure for the changes in the motion profile and is thus inversely proportional to the trajectory smoothness, as expressed by (3).

$$\phi_t = \frac{1}{\sum_{j=i-n}^{i-1} \frac{1}{i-n} |\dot{\mathbf{x}}_j - \dot{\mathbf{x}}_i|} \quad (3)$$

All 3 video quality subcriteria have been constructed such that they produce values between 0 and 1. According an equal importance to each of these subcriteria, the overall proposed measure for drone-based video quality can be written as:

$$\phi = |(\phi_p, \phi_d, \phi_t)| \quad (4)$$

Obviously, it is possible to attach weights to this global metric in order to prioritize one or two of the subcriteria, in function of the requirements of a given application. In this paper, we study the generic case and do not apply any of such weights.

#### IV. METHODOLOGY FOR GENERATING OPTIMAL DRONE TRAJECTORIES FOR TARGET OBSERVATION

As advocated in Section 2, the drone trajectory optimization problem is essentially a constrained optimization problem, where the constraints are given by the flight dynamics of the drone and where the optimization function expresses some application-specific goal. Therefore, we need to define in the first place the drone model and the application scenario.

As described in the introduction, one of the goals of this project is to develop a methodology to automatically generate drone trajectories such that a maximum amount of information can be gathered about a subject in a minimum amount of time. The application scenario is thus clearly a target observation mission.

In this research work, we assume to be dealing with a rotorcraft drone. This is a reasonable assumption, as rotorcraft are in most cases also the types of unmanned aircraft that would be used for short inspection or target observation tasks. While the execution of complex dynamic flight behaviours with rotorcraft drones requires a complex motion model and control architecture [15], this motion model can be quite well simplified for low-speed and quite static observation applications as is the case in the context of this paper. Therefore, we adopt a very simple motion model [16] for the drone to generate possible locations to move to.

Another assumption that we make is that we do not account for weather effects such as wind. Obviously, such external influencing factors can be incorporated into the system later, but here we wanted to validate in the first place the effectiveness of the proposed trajectory generation approach.

A pseudo-code representation of the general framework of the algorithm for generating drone trajectories is given in Algorithm 1. We will now explain this methodology line by line:

- *Line 2:* As stated above, the algorithm starts from a simple drone motion model, which proposes a number of possible discrete locations where the drone can move to, taking into account the flight dynamics constraints. In a first step, we perform a search over all possible new locations in order to assess which one is the best to move to. This means that a brute brute-force search is followed for searching for the optimal position. This is a quite simplistic approach, but we have opted for this option as the number of possible locations is not so enormous and it is therefore not required to incorporate some advanced optimization scheme.

- *Line 3:* In a second step, the safety of the proposed new drone location is assessed. This analysis considers in fact two different aspects:
  - The physical safety of the drone, which is in jeopardy if the drone comes too close to the ground. Therefore, a minimal distance from the ground will be imposed and proposed locations too close to the ground are disregarded.
  - The safety of the (stealth) observation operation, which is in jeopardy if the drone comes too close to the target, which means that the target (in a military context often an enemy) could hear / perceive the drone and the stealthiness of the operation would thus be violated. Therefore, a minimal distance between the drone and the target will be imposed and proposed locations too close to the target will be disregarded.
- *Lines 4-6:* The different sub-criteria are assessed, following equations (1), (2) and (3).
- *Line 7:* The global objective video quality measure  $\phi$  at the newly proposed location is calculated, following the equation (4).
- *Line 8:* The point with the highest video quality score  $\phi$  is recorded.
- *Line 9-10:* At this point, an optimal point for the drone to move to has been selected ( $\mathbf{x}_b$ ). The viewpoint history memory  $\theta_j$  and the velocity history memory  $\dot{\mathbf{x}}_j$  are updated to include this new point.
- *Line 11:* The drone is moved to the new point  $\mathbf{x}_b$ , in order to prepare for the next iteration.
- *Line 12:* The point  $\mathbf{x}_b$  is appended to the drone trajectory profile.

---

#### Algorithm 1: Trajectory generation algorithm.

---

**Input:** drone position  $\mathbf{x}_i = (x_i, y_i, z_i)$   
 target position  $\mathbf{x}_t = (x_t, y_t, z_t)$

**Output:** Drone trajectory  $\mathbf{Y}$

```

1 while not at end of the iteration do
2    $\mathbf{x}_n \leftarrow \text{CalculatePossibleNewPositions}(\mathbf{x}_i)$ 
3   forall proposed positions  $\mathbf{x}_n$  do
4     if EvaluateSafety( $\mathbf{x}_n$ ) then
5        $\phi_t \leftarrow \text{CalculatePixelsOnTarget}(\mathbf{x}_n, \mathbf{x}_t)$ 
6        $\phi_d \leftarrow \text{CalculateInnovation}(\mathbf{x}_n, \mathbf{x}_t, \theta_j)$ 
7        $\phi_t \leftarrow \text{CalculateSmoothness}(\mathbf{x}_n, \mathbf{x}_t, \dot{\mathbf{x}}_j)$ 
8        $\phi \leftarrow |(\phi_p, \phi_d, \phi_t)|$ 
9        $\mathbf{x}_b \leftarrow \text{RecordBestPoint}(\mathbf{x}_n, \phi)$ 
9    $\theta_j \leftarrow \text{UpdateDataInnovation}(\theta_j, \mathbf{x}_b)$ 
10   $\dot{\mathbf{x}}_j \leftarrow \text{UpdateTrajectorySmoothness}(\dot{\mathbf{x}}_j, \mathbf{x}_b)$ 
11   $\mathbf{x}_i \leftarrow \text{AdvanceDrone}(\mathbf{x}_b)$ 
12   $\mathbf{Y} \leftarrow \text{RecordDronePosition}(\mathbf{x}_b)$ 
    
```

---

## V. RESULTS & DISCUSSION

For the validation of the proposed methodologies, we started with the assessment of the performance of drone operators in a simulation environment [4]. Multiple simulated operators were asked to produce a high-quality video of a target within the simulation environment and the resulting total  $\phi$  scores they obtained were recorded, as shown in Fig. 2. The figure shows that the algorithm is capable of discriminating between proficient users (e.g., number 3) and less proficient users. However, it is certainly the case that more research is still required in order to validate the relationship between the subjective quality assessment and this objective metric.

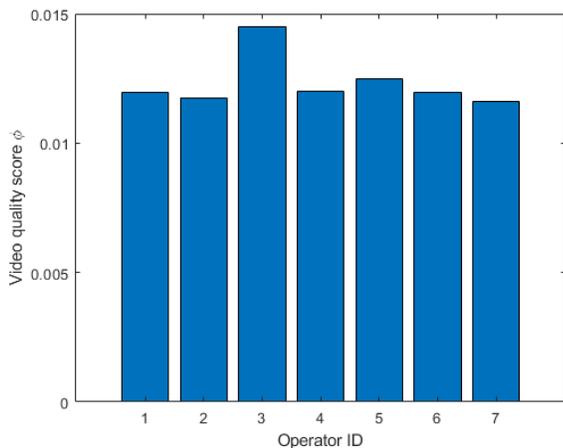


Fig. 2: Video quality scores obtained by seven operators.

In a second phase of evaluation, we validated the proposed automated drone trajectory generation, as presented in Section 3. For this validation process, we let the drone start from a random location and evaluate the optimal trajectories estimated by the algorithm.

An example of this analysis is shown in Fig. 3. In the case of the experiment depicted in Fig. 3, the drone starts from a location which is high above the target. The solution proposed by the proposed automatic trajectory generation methodology is shown in Fig. 3e, where the target position is depicted by the large sphere on the bottom. As can be noted, the proposed solution in this case consists of a spiraling downwards movement, which ensures that the target is well-perceived from all angles. Once the safety distance (from the ground and from the target) is reached, the movement pattern switches more towards an outwards extending rectangular pattern. This movement pattern is both economic for the drone and ensures that the target is perceived from ever more oblique angles.

Figs. 3a-3c show the evolution of the subcriteria  $\phi_p$ ,  $\phi_d$  and  $\phi_t$  during different steps of the drone trajectory. As can be noted, the algorithm achieves attaining a relatively high amount of pixels on target in the first half of the trajectory, while the drone is spiraling downwards. In the second half,

the number of pixels on target decreases as the drone goes further away to obtain more oblique views.

The data innovation  $\phi_d$ , shown on Fig. 3b shows a mostly decreasing evolution, which is due to the fact that it becomes ever more difficult to find new information.

As can be seen on Fig. 3c, the trajectory smoothness is quite constant during the majority of the trajectory, which means that the algorithm achieves in choosing smooth trajectories. Only near the end, there are peaks and valleys, which are related to the rectangular pattern where 90° turns are interspersed with straight lines.

Summing up the data innovation  $\phi_d$  over time allows to define a so-called scan completeness measure, shown in Fig. 3d. This metric gives an idea of the amount of new data which is gathered per step in the trajectory. In all experiments we have conducted, this scan completeness metric shows an asymptotic evolution, as shown in Fig. 3d. This is also to be expected, as it becomes after some time harder and harder to obtain new data. This metric can therefore be very useful for drone operatives to evaluate in real-time whether it has sense to continue the observation task or whether it is more sensible to stop the mission.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have developed a novel metric for assessing the video quality for drone-based observation or inspection tasks. This measure is based upon an analysis of the GNSS positioning metadata embedded in the signal. The metric is essentially based upon three criteria: the number of pixels on target, the data innovation and the trajectory smoothness. The metric was embedded in an automated trajectory generation approach, which finds optimal trajectories for maximizing the amount of information perceived from a given target. The metric and the trajectory optimization methodology were validated in the framework of a drone training and simulation environment.

The validation showed that the proposed video quality metric is capable of discriminating between different levels of users. However, more research is certainly required in this domain to assess the viability of the proposed metric. As the metric is now incorporated in the drone training environment [4], it is now the idea to start larger-scale user-testing to address this issue.

An obvious shortcoming of the proposed metric, is that it only takes into account (GNSS) positioning metadata. We will therefore in a future iteration integrate this approach in a hybrid video quality analysis model, in order to come to a more comprehensive metric. Furthermore, we aim to address some of the assumptions made in this work, making the approach work also for non-spherical objects and while also accounting for weather effects.

As discussed in Section 5, the proposed trajectory generation approach succeeds in finding optimal trajectories for target observation and inspection. Moreover, having a real-time view on the scan completeness allows to know when it is the best time to end a mission. Both of these contributions of this paper

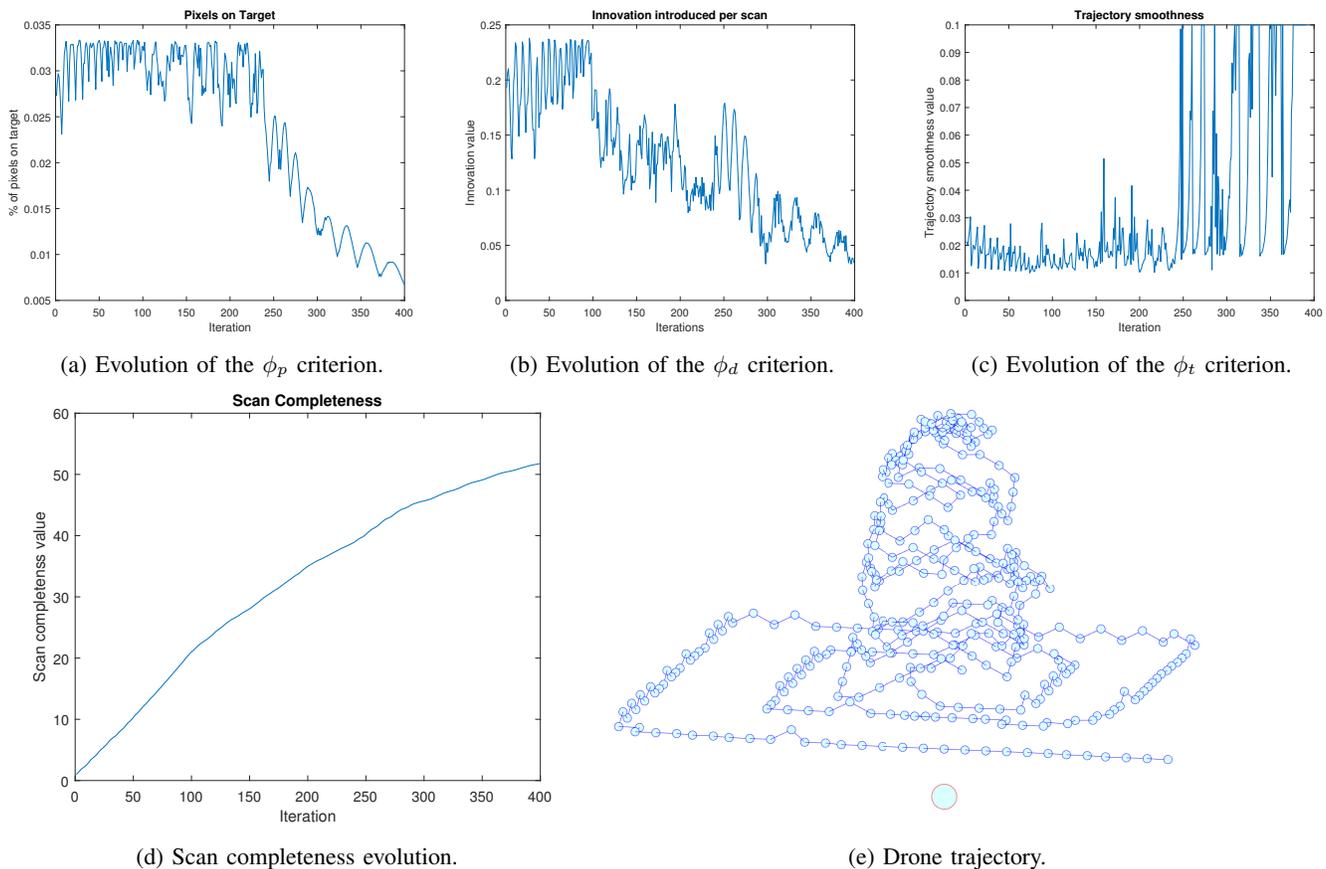


Fig. 3: Drone trajectory generation validation.

can be important time-savers for drone operators, or they can form the basis for automated target observation missions.

#### ACKNOWLEDGMENTS

This research was funded by the Belgian Royal Higher Institute for Defense and the VIAS Institute in the framework of the research study HFM19/05 (ALPHONSE).

#### REFERENCES

- [1] G. De Cubber, D. Doroftei, K. Rudin, K. Berns, D. Serrano, J. Sanchez, S. Govindaraj, J. Bedkowski, and R. Roda, *Search and rescue robotics- from theory to practice*. InTechOpen, 2017.
- [2] J. Seo, L. Duque, and J. Wacker, "Drone-enabled bridge inspection methodology and application," *Automation in Construction*, vol. 94, pp. 112–126, 2018.
- [3] I. Lahouli, E. Karakasis, R. Haelterman, Z. Chtourou, G. De Cubber, A. Gasteratos, and R. Attia, "Hot spot method for pedestrian detection using saliency maps, discrete chebyshev moments and support vector machine," *IET Image Processing*, vol. 12, no. 7, pp. 1284–1291, 2018.
- [4] D. Doroftei, G. De Cubber, and H. De Smet, "Reducing drone incidents by incorporating human factors in the drone and drone pilot accreditation process," in *11th International Conference on Applied Human Factors and Ergonomics (AHFE 2020)*, vol. 2020-July, Springer, 2020.
- [5] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, pp. 541–558, Feb 2014.
- [6] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the itu for a qoe assessment of iptv," *IEEE Communications Magazine*, vol. 46, pp. 78–84, February 2008.
- [7] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From psnr to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, pp. 660–668, Sep. 2008.
- [8] D. Hulens and T. Goedemé, "Autonomous flying cameraman with embedded person detection and tracking while applying cinematographic rules," in *2017 14th Conference on computer and robot vision (CRV 2017)*, vol. 2018-January, pp. 56–63, IEEE, 2017.
- [9] P. Perazzo, F. Sorbelli, M. Conti, G. Dini, and C. M. Pinotti, "Drone path planning for secure positioning and secure position verification," *IEEE Transactions Mob. Computing*, vol. 16, no. 9, pp. 2478–2493, 2017.
- [10] K. S. Yakovlev, D. A. Makarov, and E. S. Baskin, "Automatic path planning for an unmanned drone with constrained flight dynamics," *Scientific and Technical Information Processing*, vol. 42, no. 5, pp. 347–358, 2015.
- [11] G. De Cubber, S. A. Berrabah, and H. Sahli, "Color-based visual servoing under varying illumination conditions," *Robotics and Autonomous Systems*, vol. 47, no. 4, pp. 225–249, 2004.
- [12] G. De Cubber and G. Marton, "Human victim detection," in *Third International Workshop on Robotics for risky interventions and Environmental Surveillance-Maintenance, RISE*, 2009.
- [13] J. S. Marques, A. Bernardino, G. Cruz, and M. Bento, "An algorithm for the detection of vessels in aerial images," in *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 295–300, Aug 2014.
- [14] J. Johnson, "Analysis of image forming systems," in *Image Intensifier Symposium*, p. 244–273, 1958.
- [15] M. Kamel, K. Alexis, M. Achtelik, and R. Siegwart, "Fast nonlinear model predictive control for multicopter attitude tracking on so(3)," in *2015 IEEE Conference on Control Applications (CCA)*, pp. 1160–1166, Sep. 2015.
- [16] C. Powers, D. Mellinger, and V. Kumar, "Quadrotor kinematics and dynamics," in *Handbook of Unmanned Aerial Vehicles*, pp. 307–328, Springer, 2014.

# Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations

Rialda Spahic  
*Engineering Cybernetics*  
*Norwegian University of*  
*Science and Technology*  
 Trondheim, Norway  
 email: rialda.spahic@ntnu.no

Vidar Hepsø  
*Geoscience and Petroleum*  
*Norwegian University of*  
*Science and Technology*  
 Trondheim, Norway  
 email: vidar.hepso@ntnu.no

Mary Ann Lundteigen  
*Engineering Cybernetics*  
*Norwegian University of*  
*Science and Technology*  
 Trondheim, Norway  
 email: mary.a.lundteigen@ntnu.no

**Abstract**—The Unmanned Autonomous Systems (UAS) are anticipated to have a permanent role in offshore operations, enhancing personnel, environmental, and asset safety. These systems can alert onshore operators of hazardous occurrences in the environment, in the form of anomalies in data, during real-time inspections, enabling early prevention of hazardous events. Time series data, collected by sensors that detect environmental phenomena, enables the observation of anomalous data as dynamic instances of the dataset. Recent research characterizes anomalies in terms of their patterns of occurrence in data. However, there is insufficient research on anomalous temporal change patterns. In this paper, we examine anomalies in relation to one another and propose a conceptual categorization system for anomalies based on their temporal changes. We demonstrate the categorization through a case study of potentially hazardous occurrences observed by UAS during underwater pipeline inspection. Analyzing anomalies based on their behavior can provide further information about current environmental changes and enable the early discovery of unwanted events, simultaneously minimizing false alarms that overwhelm the systems with low-significance information in real-time.

**Keywords**—anomalies, anomalous change detection, anomaly detection, time-series analysis, autonomous systems

## I. INTRODUCTION

Sensors integrated into Unmanned Autonomous Systems (UAS), such as underwater autonomous vehicles, are reshaping our perception of the world by detecting environmental phenomena and responding to them through inputs such as graphics, motion, pressure, and heat. Underwater UAS, particularly in the offshore industry, are intended to replace operators in remote and potentially dangerous locations by residing on the seabed, collecting the data, and continuously monitoring and inspecting assets and the environment. In crucial situations, real-time data collection and analysis of the environment or assets can provide critical information, signaling us of potentially harmful deviations within the data, known as anomalies. Failure to capture anomalies effectively can have a devastating effect on the environment and result in severe financial loss.

Despite their ample presence in research and industry, anomaly detection methods have not yet matured as they are frequently too specialized or complex to evaluate [1]. Detecting anomalies, particularly for time-series data, is a challenging task that needs real-time processing while learning from analyzed data and making predictions [2]. Most anomaly detection methods are based on statistical samples of some data regions col-

lected over time [3]. When the input data for these data regions changes, it becomes challenging to select the most appropriate strategy for detecting anomalies [3]. More compellingly, it becomes challenging to detect anomalies and capture their changing nature in real-time. The anomalous change detection method searches for unusual discrepancies between measurements taken at the same site at various periods [4]. These discrepancies may be due to harmless changes in atmosphere or sensor equipment. However, they may also be pervasive and potentially indicative of something hazardous evolving at the monitored site, i.e., a deteriorating material of a pipeline surface at the offshore oil and gas platform. Unfortunately, anomaly detection methods can have two significant drawbacks: they can ignore anomalies for the sake of efficiency as tolerable collateral damage [5], or they can overload the system with low-significance data, referred to as false alarms or noise [6]. The ideal outcome of anomaly detection is to alert operators of anomalous occurrences as soon as they are detected while minimizing false alarms [2].

Historically, anomalies have been defined primarily by their pattern of occurrence in data. However, there is insufficient investigation and categorization of anomalies based on how they relate to one another, particularly by the patterns of their temporal change. The time-series data enables the collection and observation of anomalies as dynamic instances of data that alter, evolve, disappear, and reappear. Therefore, this paper's contributions is a conceptual categorization of anomalies according to patterns of their temporal change, through an overview of the identification of anomalies during time-series change detection. Analyzing anomalies based on their behavior can provide more information about current environmental changes and allow for the early detection of anomalous, potentially hazardous occurrences in real-time. Consequentially, analyzing anomalies by their behavior can assist in minimizing false alarms by allowing for the more certain elimination of noisy data.

This paper is structured as follows: Section II discusses related work exploring anomalies' characteristics and categorization, anomalous change detection methods, and real-time anomaly detection. In Section III, we describe the proposed anomaly categorization according to their temporal changes. Section IV summarizes the findings and concludes the paper. Finally, Section V discusses future research.

II. RELATED WORK

A. Anomaly Characteristics and Categorization

Anomalies are instances in a dataset that are unusual in some way and deviate from the dataset’s overall or predicted trend [7]. There have been numerous attempts in the literature to categorize anomalies based on their presence in data, the data structures in which they arise, or even application-specific high-level categorization.

1) *Anomalies by Data Structure*: In a recent review on the nature and categories of anomalies, Foorthius [1] presents an overview of anomaly categories from a data-centric perspective. Because most datasets follow a well-defined, organized format, the author [1] describes the anomalies by examining the data structures that include them: cross-sectional, time-series, time-oriented, sequence, graph, tree, spatial, and spatio-temporal data structures. The author [1] then divides anomalies into univariate, multivariate, and multivariate aggregate anomalies, each of which includes numerical, class, or categorical anomalies and mixed data anomalies.

2) *Anomalies by Occurrence in Data*: While categorizing anomalies according to the data structure in which they occur simplifies their detection, the literature most often refers to a more general approach to anomaly categorization [8]:

- Global anomaly - one or more independent data points that deviate from the rest of the data. Global anomalies are alternatively referred to as point, and content anomalies [9] [10].
- Collective anomalies - a group of data points that differ from the rest of the data. When observed individually, these points often do not constitute an anomaly. Collective anomalies are alternatively referred to as group or aggregate anomalies.
- Contextual anomalies - anomalies that deviate when an intentionally chosen context is considered, i.e., weather, season, or location. Contextual anomalies are alternatively referred to as conditional anomalies [11].

3) *Anomalies by Data Source*: According to Erhan et al., [12], sensor systems have become the primary source of data. Therefore, the authors [12] categorize anomalies according to their origins and potential causes (see Table I). Sensor data frequently deviate from predicted behavior. The authors [12] underline the importance of evaluating the performance of anomaly detection systems using physical world data, as opposed to virtual testing with simulators. Since anomalies occur suddenly and are frequently unusual in physical world data, artificially manufacturing them through simulations or data extrapolation can be challenging.

TABLE I  
ANOMALY CATEGORIZATION BY ORIGIN, ADAPTED FROM ERHAN ET AL. [12]

| Anomaly origin | Potential cause   |
|----------------|---|
| Environment    | Unusual events, disasters, weather changes, new objects or compounds                        |
| System         | Hardware limitations, system malfunctions   |
| Communication  | Network loss or delay   |
| Attacks        | Malevolent attacks on the physical components, malevolent interference or attack in network |
| Spike          | Short peak in measured values, distinct deviation from common measurements                  |
| Noise          | Increase in the variance in successive data samples   |
| Constant       | A constant neutral value reported by sensor   |
| Drift          | Off-set in the measurements   |

4) *Application-Defined and Specific Anomaly Types*: Ragozin et al. [13] approached forecasting complex time-series within an automated industrial system by *basing anomalies on their distinct dynamic characteristics* to increase the efficiency of information security management within the observed system. The authors [13] developed a method based on structural analysis of multi-component time series and digital signal processing technology for decomposing complex multi-component time series into several essential components for further real-time monitoring of the industrial information system and detecting any component-specific behavior anomaly event or proximity to such event.

Lutz et al. [14] analyzed operational safety-critical anomalies. The authors [14] argue that despite the widely-established benefits of anomaly analysis for operational software, research on anomaly analysis for safety-critical systems has been sparse. Patterns of software anomaly data for operational, safety-critical systems, in particular, are poorly known [14]. The authors [14] describe the findings of two hundred abnormalities on seven spacecraft systems using classification methods. The results of their study demonstrated various classification patterns, including the causal significance of data access and delivery issues, hardware degradation, and unusual incidents. Anomalies frequently revealed hidden software needs critical for the system’s robust, accurate operation [14].

B. Anomalous Change Detection

In a recent review of change detection, Liu et al. [15] classify change detection methods based on their application purpose, data availability, and automation degree. The authors [15] describe anomalous change detection, and time-series change detection as application-specific methods most frequently used in image analysis. By suppressing background and emphasizing alterations, anomalous change detection finds anomalous changes between images. Anomalous change detection is typically focused on detecting minor changes caused by the insertion, deletion, or movement of produced small items and on small stationary objects that exhibit spectrum shifts between images, as with camouflage concealment and deception [15]. The authors [15] argue that the critical point is to examine the image statistics, increase the likelihood of detecting changes

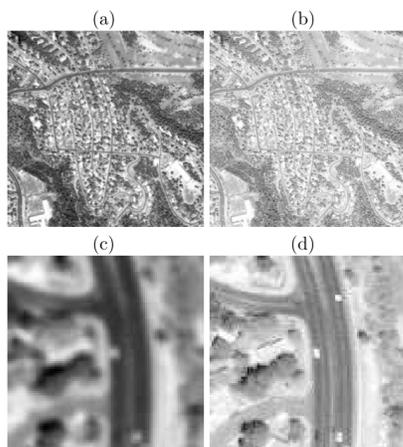


Fig. 1. (a,b) Predictable change in image contrast and brightness; (c,d) Interesting change with (artificially) added vehicle, adapted from [4]

induced by human activity, and suppress background in image scene sequences.

Theiler et al. [4] employed anomaly detection to identify uncommon changes in images of the same scene captured at various periods and often under varying viewing conditions (see Figure 1). The detection of anomalous changes in imaging is of broad general interest and is particularly useful in remote sensing [4]. The authors [4] emphasize that anomalous change is distinct from and more unusual than changes across an entire scene. The authors [4] propose a framework based on a non-flat background distribution stated in terms of data distribution, with anomaly detection treated as a classification problem. The proposed framework identifies anomalous changes capturing meaningful differences between images while avoiding predictable noisy information caused by the camera's focus, contrast, or brightness.

### C. Time-Series Anomaly Detection

Although many organizations collect time-series data, Feremans et al. [16] contend that automatically analyzing them and extracting valuable knowledge, such as a comprehensible model that flags critical anomalies, remains a complex problem, despite decades of effort. After examining various benchmark datasets for time series anomaly detection, the authors [16] discovered that these datasets frequently contain univariate time series with local or global extrema or point anomalies. By contrast, their research concentrated on collective and contextual anomalies, requiring data analysis from multiple sources to detect anomalies successfully. As a result, the authors [16] proposed a method for detecting anomalies in mixed-type time series. The method uses frequent pattern mining methods to create an embedding of mixed-type time series to train a prevalent anomaly detection method, isolation forest. Assuming that the anomalies are infrequent in the data, the isolation forest isolates them by continually splitting the data with low computational costs [17]. Experiments on multiple real-world univariate and multivariate time series and a synthetic mixed-type time series

demonstrate that the proposed method outperforms established anomaly detection methods such as MatrixProfile, Pav, Mifpod, and Fpof [16].

Hannon et al. [18] used anomaly detection on streaming data to explain a power-grid system's real-time behavior and provide insight to system operators. The authors examined a real-time anomaly detection followed by a data-driven framework based on the statistical machine learning methods (decision trees and k-nearest neighbors) to enable the remote analysis of individual grid components for monitoring, detecting, and classifying anomalies that generate warnings of possible shortcomings in the system. They [18] concluded that classification of identified anomalies using well-defined probabilistic scores and classification of detected anomalies using interpretable decision trees demonstrates a high level of accuracy, as a result enabling operators to take corrective action to avert cascading blackouts and prevent system failures.

Previous research has established a variety of applications for anomaly detection and a need for a more profound comprehension of anomalies. In a discussion paper *Anomalousness: How to measure what you can't define*, Theiler [19] describes anomaly detection as target detection with unknown targets and with the objective to differentiate anomalies (unknown targets with stubbornly undefined attributes) from a background that is generally too cluttered to support an explicit model. Despite the challenges in defining and categorizing anomalies, the outcomes and discussions of previous studies demonstrate a promising direction in application-specific and dynamic-oriented anomaly categorization.

### III. CATEGORIZATION OF ANOMALIES BASED ON THEIR TEMPORAL CHANGES

After decades of research on anomaly detection, selecting anomalies to investigate and those to disregard as noise continues to be a complex problem, particularly with the pressure of a growing need for autonomous systems. Given the poor camera vision and ambiguous sensor inputs in the subsea environment [20], it is only natural to assume that strange phenomena, such as biological growth or misplaced objects, are frequently misinterpreted. This misinterpretation can further result in the misallocation of resources or the omission of signs indicating a more hazardous occurrence. Using inspiration from prior research on grouping time-series data [21] and integrating time-series and event logs into itemsets [16], we open opportunities to investigate prospects for isolating and analyzing changes in anomalies based on their geospatial context. By combining insights from time-series change detection on dynamic data points [21]–[23] with application-specific anomalies [14] [24], we observe that anomalies can display behavioral patterns such as frequent or reoccurring, disappearing and reappearing, and expanding.

*a) Frequent or Recurring Anomalies:* Feremans et al. [16] discuss frequent patterns in data, assuming that because

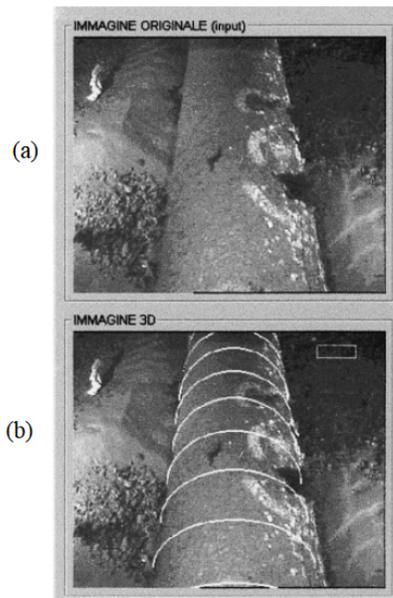


Fig. 2. (a) Visual inspection of underwater pipeline, images taken by autonomous underwater vehicle, adapted from [20]; (b) 3D scan over the underwater pipeline, adapted from [20]

anomalous activity infrequently occurs in time series, the frequent patterns represent frequently seen normal behavior. The main advantage of frequent pattern extraction is that the extracted patterns are easily interpretable and aid classifiers and anomaly detection methods in differentiating between normal and anomalous behavior in data. However, it might quickly become problematic if an anomalous event occurs repeatedly or in patterns. Anomalies that reoccur in patterns, hence generating a recurrent pattern in obtained data, present a concern because they can be difficult to spot or even mistaken as part of the normal dataset. Normal data can mask these anomalies, making it particularly difficult to detect when using unsupervised methods.

A practical example, seen on Figure 2, is the pipeline with unclear surface material, provided by images collected during a visual inspection of sea bottom infrastructure by an autonomous underwater vehicle. Visually inspecting structures can detect various phenomena, from object detection to material degradation such as corrosion monitoring [25]. However, a less intrusive process, such as biological growth, happens frequently and can readily obscure a more intrusive process, corrosion. Although additional measurements like ultrasonic testing and electromagnetic mapping are used to identify additional information about the corrosion process, the pace of corrosion (spread over time), the exact location, and even plausible causes [25], relying on unsupervised visual inspection of anomalies may not be sufficient.

*b) Disappearing and Reappearing Anomalies:* Although disappearing anomalies are not usually mentioned in industrial anomaly detection applications, they are a fairly common topic in stock market anomaly detection. During the analysis of

the dynamic persistence of anomalies, Marquering et al. [26] highlighted the occurrences of disappearing and reappearing anomalies. Since most seasonal or predictable anomalies are well-known, they should not persist [26]. However, the authors [26] question the persistence of such anomalies as a source of contention. They highlight essential questions on disappearing and reappearing anomalies in data: *Are there still anomalies in recent data? Are they just existent during specific periods, or did they completely vanish? What is the immediate cause of the endurance of the anomaly?* The occurrence of disappearing and reappearing anomalies may be of interest in time-series change detection for various applications.

During a real-time inspection of an underwater pipeline, as depicted in Figure 3, recordings of fading unusual events may represent a low-importance environmental phenomenon that does not require comprehensive inspection, thus saving additional resource allocation. However, the persistence of such occurrences may represent something of more profound research interest [26].

*c) Expanding Anomalies:* As the environment evolves and changes over time, assuming that anomalous occurrences will exhibit similar changes is natural. Despite anomalies' dynamic and evolving nature being frequently discussed in sensor networks, it is not often discussed in other applications. What appears to be an innocuous anomaly may grow to affect various regions of the inspected structure. The purpose is to identify the onset of the anomaly as fast as feasible while maintaining a low false alarm rate [23]. This detection problem is formulated as a stochastic optimization problem utilizing a delay metric based on the anomaly's worst-case path [23]. In Figure 4, we illustrate a point anomaly (Figure 4 (a)) expanding into a collective anomaly (Figure 4 (b-j)). At an early stage (Figure 4 (a)), the detected point anomaly or a smaller collection of anomalies may not yet indicate a high-significance unusual occurrence. However, if unexplored, the anomalous collection may develop into a possibly hazardous state (Figure 4 (j)), leaving less time for a reactive response. Detecting anomalies early enables preventative measures. Expanding fractures of the pipeline surface material are a practical example of expanding anomalies during an underwater pipeline inspection.

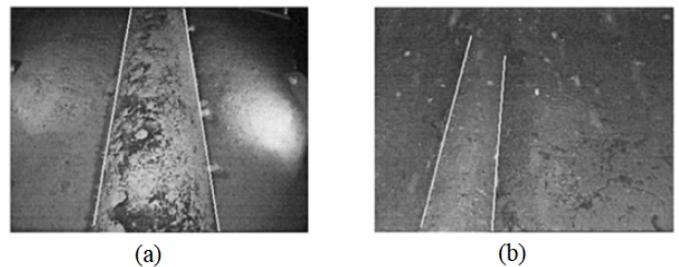


Fig. 3. (a) Visual inspection of underwater pipeline, images taken by autonomous underwater vehicle: Possible material degradation or biological growth?, adapted from [20]; (b) 3D scan over the underwater pipeline, adapted from [20]

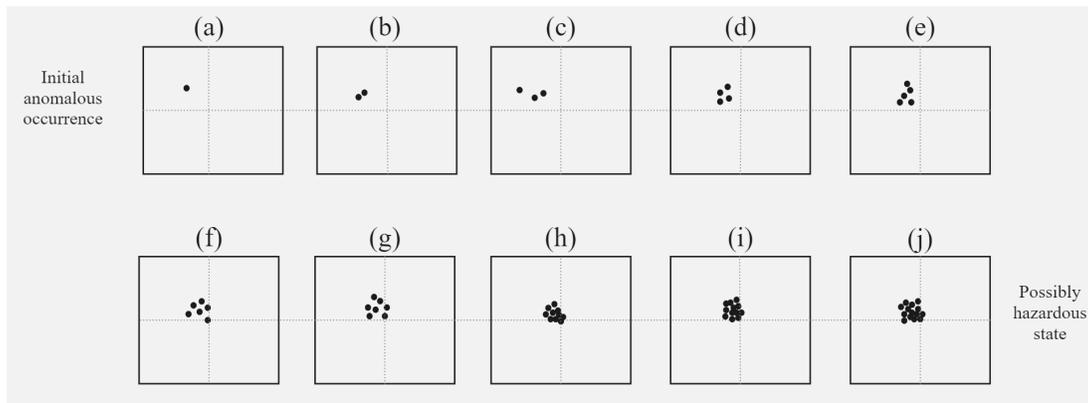


Fig. 4. Anomalies that expand over time

TABLE II  
DESCRIBING ANOMALIES BY TEMPORAL CHANGE

| <i>Anomaly Type</i> | <b>Frequent / Recurring</b>  | <b>Disappearing and Reappearing</b>  | <b>Expanding</b>  |
|---------------------|--|--|---|
| <b>Point</b>        | Frequently occurring point anomaly.  | Disappearing and reappearing point anomaly may be a sign of pervasive environmental phenomena.   | Point anomaly may evolve into a collective anomaly of larger size and impact.                         |
| <b>Collective</b>   | Frequently occurring collection of anomalies with similar properties (i.e., geospatial context).   | Disappearing and reappearing collective anomaly may be a sign of pervasive environmental phenomena.  | Collective anomalies may evolve into a more intrusive anomalous occurrence of larger size and impact. |
| <b>Contextual</b>   | Anomalous depending on the context due to a potential risk of being misinterpreted as normal and left unexposed or a frequent anomaly collection obscuring more intrusive processes. | Context (i.e., geospatial, seasonal, weather) aids in determining the anomalousness of the disappearing/reappearing phenomena and finding the causes of their persistence. | Anomalous depending on the context.   |

The proposed conceptual categorization of anomalies according to their temporal changes does not impede their occurrence in data as point, collective, and contextual anomalies. Table II summarizes the two categories that are intended to complement one another, aiding in our comprehension of unusual events occurring during autonomous operations. Anomalies’ behavior is highly dependent on context, not just on their occurrence as a single point or collection of anomalies. The criticality of frequently occurring point and collective anomalies varies by context, as they may be seen as normal and therefore obscure more intrusive processes. This increases the likelihood that the unexposed anomaly may develop into a potentially hazardous event that could have been discovered earlier. Similarly, the context (i.e., seasonal, weather) of disappearing and reappearing anomalies can aid in identifying the cause of their pervasiveness and provide additional reasoning for unanticipated environmental phenomena. Additionally, the point anomaly may expand creating a collective anomaly of more impactable volume and intrusiveness. Contextual information (e.g., changed material properties due to chemical or temperature variations) can assist in determining the criticality and anomaly of observed unanticipated changes. Observing and categorizing anomalies according to their temporal changes adds context to our understanding of how anomalies relate to one another and evolve in a normal and predictable data environment. This knowledge

enables the UAS to perceive environmental phenomena and anomalous events in their geospatial and temporal context, improving understanding of the significance and criticality of anomalous occurrences.

#### IV. CONCLUSION

The research on time-series anomaly detection has been application-oriented and vague. Despite decades of research and categorization approaches, persistent obstacles prevent anomaly detection from maturing and becoming a dependable component of autonomous systems. While an unsupervised and data-driven strategy is common in industry and research, it is insufficient to achieve reliable autonomy. Therefore, this paper proposes a fundamentally different perspective of anomalies via a conceptual categorization of anomalies according to their temporal changes. Frequent or recurrent, disappearing and reappearing, and expanding anomalies describe the behavior of anomalies and provide context for their dynamics observed through time-series data analysis. Observing anomalies as they evolve through time enables us to deduce the underlying causes of anomalous occurrences, focusing on more pertinent data from the vast collections of sensor measurements, thus allowing the UAS to react if and when the situation requires it during real-time operations.

## V. FUTURE WORK

We regard our approach of categorizing anomalies according to their temporal change as a starting point for future research to construct a framework for detecting anomalous change in real-time by identifying practical time-series anomaly detection methods. Thus, future work involves simulating streaming data and analyzing images collected by the UAS during visual inspection of an underwater pipeline to validate the proposed temporal categorization of anomalies and identify potential shortcomings.

## ACKNOWLEDGMENT

This research is a part of BRU21 – NTNU Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and supported by Equinor.

## REFERENCES

- [1] R. Foorhuis, "On the nature and types of anomalies: a review of deviations in data," *Int. J. Data Sci. Anal.*, vol. 12, no. 4, pp. 297–331, 2021.
- [2] A. Lavin and S. Ahmad, "Evaluating Real-time Anomaly Detection Algorithms - the Numenta Anomaly Benchmark," in *IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015*. Miami, Florida, USA: Institute of Electrical and Electronics Engineers Inc., 2015, pp. 38–44.
- [3] A. S. Alghawli, "Complex methods detect anomalies in real time based on time series analysis," *Alexandria Eng. J.*, vol. 61, no. 1, pp. 549–561, 2022.
- [4] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," in *ICML Work. Mach. Learn. Algorithms Surveill. Event Detect.*, 2006, pp. 7–14.
- [5] K. Makhoul, S. Zhioua, and C. Palamidessi, "On the applicability of ML fairness notions," *arXiv*, pp. 1–32, 2020.
- [6] R. Sekar et al., "Specification-based Anomaly Detection: A New Approach for Detecting Network Intrusions," in *Proc. 9th ACM Conf. Comput. Commun. Secur. - CCS '02*. New York, NY, USA: Association for Computing Machinery, 2002, pp. 265–274.
- [7] C. C. Aggarwal, "An Introduction to Outlier Analysis," in *Outlier Anal.* Springer, Cham, 2017, ch. 1, pp. 1–34.
- [8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," *ACM Comput. Surv.*, vol. 14, no. 1, pp. 1–22, jul 2009.
- [9] A. Fisch, I. Eckley, and P. Fearnhead, "Subset Multivariate Collective And Point Anomaly Detection," *J. Comput. Graph. Stat.*, pp. 1–51, 2019.
- [10] M. A. Hayes and M. A. Capretz, "Contextual anomaly detection in big sensor data," in *Proc. - 2014 IEEE Int. Congr. Big Data, BigData Congr. 2014*. Institute of Electrical and Electronics Engineers Inc., sep 2014, pp. 64–71.
- [11] S. Xiuyao, W. Mingxi, C. Jermaine, and S. Ranka, "Conditional anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 631–644, may 2007.
- [12] L. Erhan et al., "Smart anomaly detection in sensor systems: A multi-perspective review," *Inf. Fusion*, vol. 67, no. September 2020, pp. 64–79, 2021.
- [13] A. N. Ragozin, V. F. Telezhkin, and P. S. Podkorytov, "Forecasting complex multi-component time series within systems designed to detect anomalies in dataflows of industrial automated systems," *ACM Int. Conf. Proceeding Ser.*, pp. 1–5, 2019.
- [14] R. R. Lutz and I. C. Mikulski, "Empirical analysis of safety-critical anomalies during operations," *IEEE Trans. Softw. Eng.*, vol. 30, no. 3, pp. 172–180, mar 2004.
- [15] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, 2019.
- [16] L. Feremans et al., "Pattern-Based Anomaly Detection in Mixed-Type Time Series," *Mach. Learn. Knowl. Discov. Databases*, vol. 11906, pp. 240–256, 2020.
- [17] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proc. - IEEE Int. Conf. Data Mining, ICDM*. IEEE, 2008, pp. 413–422.
- [18] C. Hannon, D. Deka, D. Jin, M. Vuffray, and A. Y. Likhov, "Real-time Anomaly Detection and Classification in Streaming PMU Data," in *2021 IEEE Madrid PowerTech*. Madrid: IEEE, 2021, pp. 1–6.
- [19] J. Theiler, "Anomalousness: how to measure what you can't define," *Fourier Transform Spectrosc. Hyperspectral Imaging Sound. Environ.*, p. JT1A.2, 2015.
- [20] G. L. Foresti, "Visual inspection of sea bottom structures by an autonomous underwater vehicle," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 31, no. 5, pp. 691–705, oct 2001.
- [21] T. Rakthanmanon, E. J. Keogh, S. Lonardi, and S. Evans, "Time series epenthesis: Clustering time series streams requires ignoring some data," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 547–556, 2011.
- [22] S. Guggilam, V. Chandola, and A. Patra, "Tracking clusters and anomalies in evolving data streams," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 15, no. 2, pp. 156–178, 2021.
- [23] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami, "Quickest Detection of Growing Dynamic Anomalies in Networks," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2020-May, pp. 8926–8930, may 2020.
- [24] T. Wang, C. Fang, D. Lin, and S. F. Wu, *Localizing temporal anomalies in large evolving graphs*. Society for Industrial and Applied Mathematics Publications, 2015.
- [25] Y. T. Al-Janabi, "Monitoring of Downhole Corrosion: An Overview," *Soc. Pet. Eng. - SPE Saudi Arab. Sect. Tech. Symp. Exhib. 2013*, pp. 108–118, may 2013.
- [26] W. Marquering, J. Nisser, and T. Valla, "Disappearing anomalies: A dynamic analysis of the persistence of anomalies," *Appl. Financ. Econ.*, vol. 16, no. 4, pp. 291–302, 2006.

# A Multimodal AI Approach for Intuitively Instructable Autonomous Systems: A Case Study of an Autonomous Off-Highway Vehicle

Abdellatif Bey Temsamani<sup>1</sup>, Anil Kumar Chavali,  
Ward Vervoort  
Flanders Make  
Lommel, Belgium

<sup>1</sup>abdellatif.bey-temsamani@flandersmake.be

Tinne Tuytelaars<sup>2</sup>, Gorjan Radevski  
KU Leuven, ESAT  
Leuven, Belgium

<sup>2</sup>tinne.tuytelaars@esat.kuleuven.be

Hugo Van hamme<sup>3</sup>  
KU Leuven, PSI  
Leuven, Belgium

<sup>3</sup>hugo.vanhamme@esat.kuleuven.be

Kevin Mets<sup>4</sup>, Matthias Hutsebaut-Buysse, Tom De  
Schepper, Steven Latré  
University Of Antwerp - imec  
Antwerpen, Belgium

<sup>4</sup>kevin.mets@uantwerpen.be

**Abstract**— In current production shop floors, a fleet of production machines and AGVs form a full manufacturing system with a high degree of automation. These current manufacturing systems need to deal with high variability of products and production tasks. Every task, however, requires a proper reconfiguration & control that is often done manually requiring complex settings and long configuration time. With AI techniques the reconfiguration of these systems to deal with a new task can be made more intuitive. In this paper we present the upgrade of an autonomous system, used for manufacturing assets handling and transportation, with AI features that make it easy to reconfigure in order to deal with high variability of assets and missions. Visual and spoken information is used to instruct and guide the autonomous vehicle using an AI multimodal framework where first, spoken language, with different local dialects, is translated to digital instructions, that can be associated to visual information to form control instructions to the autonomous vehicle. Different AI models, respectively for spoken language understanding, visual perception, vision based navigation are associated through a multimodal AI framework to intuitively control the AGV to perform a specific task. Beside the challenges related to the integration of these models in the AGV platform, other challenges related to dealing with variabilities of dialects, objects, surroundings and ambient conditions are partly tackled in this research.

**Keywords**-AI based autonomous systems; Multimodal AI; Natural language processing (NLP); deep learning; neural networks; reinforcement learning

## I. INTRODUCTION

Autonomous Guided Vehicles (AGVs) are becoming more and more popular in industrial applications. They can pick up and deliver materials around a manufacturing facility or warehouse [1]. However, with the continuously increase of mass customization [2], a return on investment of production AGVs can only be obtained if these AGVs can easily perform large variability of tasks and / or deal with large variability of products.

Tasks scheduling and allocations have been done by a central entity for a fleet of AGVs following predefined configurations. Driven by flexibility, robustness and scalability requirements, the current trends in AGV systems are customization and decentralization [3]. In a decentralized architecture, an AGV broadcasts the information about its states in a local way and decides which actions to take [4].

Although new generation of AGVs are highly instrumented with different guidance systems (optical, magnetic, laser, etc.), they are more optimized and suited for long-distance transportation of materials from / to multiple destinations, and / or tuned for repetitive and predictable tasks [5].

(Re-)configuring AGVs to perform multiple tasks in a non-predictable environments remains, however, a challenge today in industrial floors due to dynamically changing environments. Literature in this research remains very limited and focuses more on path planning methods in unknown environments, yet using references (e.g. markers, identifiers, etc.) [26]. Research on voice controlled AGV remains in the level of performing basic operations (e.g. moving with constant speed) in a prescribed path [27].

In this paper, we propose a Multimodal Artificial Intelligence (AI) framework that allows to intuitively and easily (re-)configure an AGV to perform different and variable tasks. In this framework, an operator can instruct the AGV by speech interaction that can be done locally or remotely. The operator can intuitively instruct the vehicle. This instruction is then decoded through different interpretation layers that make respectively use of (i) natural language processing, (ii) association with vision deep learning for objects recognition and localization and (iii) association with reinforcement learning for navigation.

(i) Spoken interaction offers fast and natural interaction with machines and AGVs, while operators keep their hands and eyes free for other tasks. The task of a Spoken Language Understanding (SLU) component is to map speech onto an interpretation of the meaning of a command, while taking the variability in the input signal into account: differences in voice, dialect, language, acoustic environment (noise, reverberation), hesitation, filled pauses and pure linguistic variation. Traditionally, SLU is approached as a cascade of Automatic Speech Recognition (ASR) mapping speech into text followed by Natural Language Understanding (NLU) mapping text onto meaning. This cascaded approach tends to propagate and inflate ASR errors and requires application-specific textual data, which is unnatural to acquire. Instead, this work uses End-to-End SLU (E2E SLU), where spoken instructions are directly mapped onto meaning without textual intermediate representations.

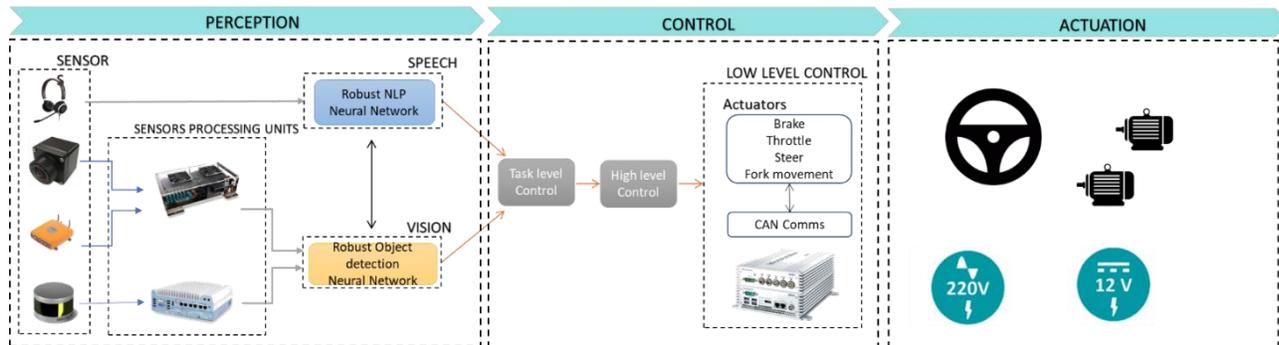


Figure 1. Autonomous platform around the autonomous off-highway vehicle

(ii) For the agents to interact with the environment, they must process and understand visual input, i.e., extract the semantically relevant cues from the environment in order to execute the desired task. Should the input be provided from an RGB camera, a plethora of Deep Learning techniques could be leveraged to achieve visual understanding. Deep Learning techniques rely on Neural Networks, commonly (pre-)trained on large-scale general-purpose datasets, e.g., for visual recognition [6] such as object detection [7]. Since our goal is to interpret a language-based instruction we need to locate the object(s) referred to by the person in the environment. To this end, we build on a state-of-the-art object detection method. Given an RGB input, the object detector's role is to locate (detect) the relevant objects, i.e., potential objects that the person instructing the AGV might refer to. This serves as a backbone to perform multimodal interaction — associating the representation of the language-based instruction with the representation of the spatial layout of the scene (2D location and categories of the detected objects).

(iii) Egocentric navigation is one of the core problems intelligent systems need to master. An agent needs this skill not only to execute the task at hand, but also to navigate, in order to collect experience that can be used to learn from. Navigation is typically done by either using expensive specialized lidar and radar sensors, or by relying on visual inputs. In this paper we examine the performance of utilizing an RGB camera, a depth camera or a combination of both for navigation. In the presented approach we have chosen for an end-to-end learning-based navigation approach. Such an approach is able to outperform Simultaneous Localization and Mapping (SLAM) based approaches [8], it doesn't suffer from propagation errors due to mapping errors and excels in visually sparse environments [9]. As we need to train our navigation system in simulation due to the large amount of required interactions with the environment, we also propose a digital-twin based solution to utilize a navigation model trained in simulation in the real world. Through our multimodal speech and vision system, combined with learned navigation, we demonstrate an intuitively instructible autonomous system, which can act as a platform for various tasks.

## II. CASE STUDY – AUTONOMOUS OFF-HIGHWAY VEHICLE

The architecture of the autonomous vehicle, its HW / SW components and upgrades with the Multimodal AI framework are described in this section.

### A. Autonomous vehicle architecture

The AGV used in this paper consists of the off-highway tractor developed at Flanders Make [10]. The architecture of the AGV consists of a perception, control, and actuation frameworks (Figure 1). To perceive the environment we use cameras, lidars, a GNSS system and a microphone. The sensors data is then processed in separate computing platforms and stored on middleware (ROS), from where the Speech and Vision units send the information to the control block. This later is divided in two levels, (i) a High-level controller that controls the tractor via a state machine and (ii) a Low-level controller, built in a dSpace platform [11], that controls the trajectory such that velocity and heading can be followed. The output signals are sent to different actuators that consist of the brakes, throttle, steering and fork implement that are controlled via servo motors. Autonomous vehicle upgrades to deal with Multimodal AI

An example of intuitive instructions given by an operator to the AGV to execute a task and their high level interpretations by the Multimodal AI framework, described in this paper, is illustrated in Figure 2.

The instruction: *'Pick up the red pallet and put it on the truck'*, needs first to be communicated to the computer that runs the speech AI module (described in Section III). In the next level, a vision module, where real time 2d vision data is processed and fed to a pretrained NN, allows objects classification and their association to different attributes such as object's type, color, etc. (as described in Section IV). The AGV should then move towards the recognized object. This step is supported by the association made so far between speech and vision data as well as the navigation data. This later makes use of the cartesian coordinates of the AGV in the navigation space and the reinforcement learning module (as described in Section V) that allows to estimate the optimal trajectory between the AGV and the object of interest.

In order to implement and demonstrate the Multimodal AI framework, The AGV is updated by a newly installed system for interfacing through speech with a dedicated PC. This PC is also used for developing and testing the neural

networks. It is equipped with a powerful Nvidia GPU and a new headset microphone for giving audio commands. The autonomous tractor internally uses ROS to communicate between the different sub-systems. Originally it was only used sparingly in the autonomous tractor, mainly to communicate lidar sensor data. After the system upgrade, also the control unit, the dedicated PC and the Nvidia Drive platform have a ROS interface. While the Nvidia Drive could technically runs the neural networks, for more convenience, during testing we installed the neural networks on the dedicated PC. Data from the cameras on the Nvidia Drive, LiDAR and navigation all come in as ROS messages while for speech a simple microphone is connected to the PC. The output of the multi-modal setup is the location of a specific object together with the task the tractor must complete. This information can be communicated through ROS to the navigation module.

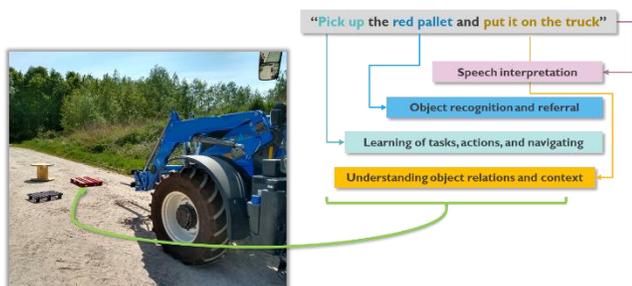


Figure 2. Example of speech-based instruction and multimodal mapping

### III. SPOKEN LANGUAGE UNDERSTANDING

This section summarizes the speech data generation, speech model training and testing as well as the architecture and the validation of the spoken language understanding.

#### A. Speech data generation

To train the SLU model, training dataset with audio fragments is made. It is important that the recorded speech seems natural, as if the participants are really interacting with the AGV. To this end, we believe that a visual feedback to the participant would be very useful. Therefore, a simple automotive simulator called Webots [12] was used and a set of API calls were written in order to control the simulated tractor in the simulated environment (Figure 3).

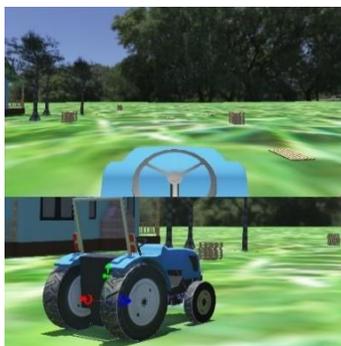


Figure 3. Simulator that provides visual feedbacks to participants for speech recording

The participants are given some high-level objectives and it is up to them to control the tractor with speech

commands in order to fulfil these tasks. With the ‘high-level’ objectives (in contrast with explicitly providing the primitive commands to the participants) we aim to improve the variability of commands that participant’s would naturally choose to control the tractor. Every time the participants speak a relevant command, the experiment supervisor presses a button to invoke the correct API call. This way, we already have some automatically generated annotations linking the participant’s speech command to the supervisor’s API call invocation. We recorded the audio in Audacity in WAV format using a headset microphone and a separate standalone microphone. The commands were mainly basic control commands like turning a direction or driving speed. A total of 14 people who speak Dutch language (different dialects) were recorded with mixed female and male voices.

#### B. SLU model architecture & training

Classical semantic frames are used for representing the semantics of an utterance. A semantic frame is composed of slots (e.g. “direction”) that take one of multiple slot values (e.g. “forward” or “backward”). This encoding represents the affordances of the AGV and corresponds to API calls with parameters filled in. The task of the SLU component is to map an utterance (spoken command) to a completed semantic frame. The SLU architecture follows the encoder-decoder structure first described in [13] and later refined in [14] to allow for encoder pretraining for ASR targets on generic Dutch data. The decoder is trained on the task-specific data. The encoder encodes an utterance in a single high-dimensional embedding in two steps. The first step maps MEL-filterbank speech representations to letter probabilities using a transformer network [15] preceded by a down sampling CNN, trained maximal cross-entropy between predicted and ground truth transcriptions in a 37-letter vocabulary. The training data consist of 200 hours of Flemish speech with its textual transcription from the CGN corpus [16], fourfold augmented with noise (0-15 dB) and reverberation (sampled from [28]) to achieve acoustic robustness. The second step counts bigram occurrence frequencies of all letter pairs across the utterance and repeats the same while skipping one position in the bigram, resulting in a  $2(37^2) = 2738$  dimensional utterance embedding.

The decoder maps the utterance embedding onto a multi-hot encoding of the slot values via non-negative matrix factorization (NMF) [17] as described in [13]. Other than in the pretraining stage, the training pairs here do not require textual transcription, but are pairs of speech with the completed semantic frame. Here, a neural network could be taken as well, but the chosen decoder has several advantages: (1) it requires few training data, (2) it retrains in a fraction of a second when user interaction data becomes available and (3) it establishes a bag-of-words model making the SLU system less sensitive to the rather free word order in Dutch (at least compared to English). Learning a stricter word order would require more task-specific training data exhibiting the word order variability.

The approach is evaluated on the Grabo corpus [18], which contains a total of 6000 commands to a robot spoken by ten Flemish speakers and one English speaker. The

commands were recorded with the participants' own hardware in a quiet room at their homes. The semantics are described in eight different semantic frames describing driving, turning, grabbing, pointing, ... using one (e.g. "close gripper") to three (e.g. "quickly drive forward a little bit") of ten slots (e.g. angle, direction, ...), which can take between two and four different values. In total, 33 different meanings occur in the data. The accuracy is evaluated as the F1-score for slot values as a function of the number of task-specific training examples. The trained decoder is speaker-specific. The average accuracy over speakers is plotted in Figure 4 and shows that with the minimal of 33 training utterances, i.e. one example per meaning, an accuracy of over 98.5% is reached. The performance saturates around 180 task-specific utterances.

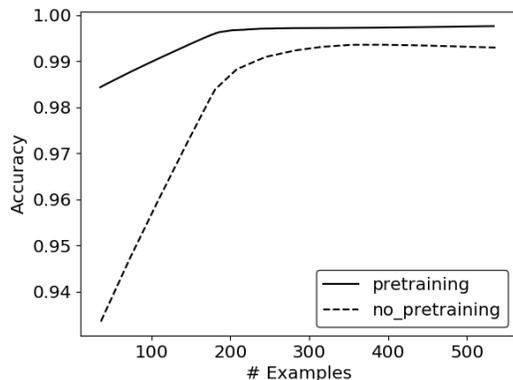


Figure 4. F1-score as a function of the number of task-specific training examples.

### C. SLU model validation

For deployment we set up a docker container to run all the code. We developed a user interface to be able to easily visualize the results of the SLU model and provide training examples for training the decoder. In this interface, it is possible to record samples, open the microphone so the tractor can listen, give feedback to the model and retrain the model. After each command is given the confidence value of the prediction is estimated. Commands with sufficient confidence are forwarded to the tractor through ROS to the control PC.

The initial accuracy of the model depends a lot on the person giving the commands and their accent. But we were able to achieve high levels of accuracy of more than 90 percent in the noisy tractor environment using an active learning approach. In this approach, the operator can give feedback samples to retrain the model. In this experimental set-up, repeating an instruction in 5 instances proved to achieve high accuracy (90%). The retraining flow is quite time-efficient and takes less than a second to retrain.

## IV. VISUAL PERCEPTION AND ASSOCIATION WITH SPEECH

The visual perception of the scene, including the scene data generation, the AI model architecture, it's training and testing as well as its association with speech in the Multimodal AI framework are described in this section.

### A. Vision AI Objects detection and classification

#### 1) Vision data generation

The dataset for training the vision model contains images with mostly objects that the AGV can pick up. This means mostly pallets and boxes of varied materials, shape and sizes containing materials like bobbins and wooden planks. This data was recorded on the Flanders Make local site, spread over two occasions: once on an early cloudy morning in spring and one just after noon in summer with sunny weather. Every image was recorded with a resolution of 960 x 608 pixels. The entire dataset contained 1100 images, derived from 9 videos. These videos each recorded one configuration of objects from many angles.

#### 2) Vision NN architecture & training

The main building block of the vision pipeline is the object detector. It gets an RGB image  $I$  as input, where  $I \in \mathbb{R}^{3 \times H \times W}$  and  $H$  and  $W$  are the image height and width respectively. The model we use is a state-of-the-art two-stage object detector, where in the first stage, a region proposal network generates regions of interest for the image, and in the second stage, bounding boxes and object classes are predicted for each proposal, which exhibits an objectness score above a certain threshold. The region proposal network generates region proposals by sliding a spatial window over a feature map obtained from a Convolutional Neural Network (CNN), i.e., a backbone. Additionally, the object detector includes a Feature Pyramid Network [19], a fully-convolutional module, which generates feature maps at different levels, thus enabling the model to recognize objects at different scales. The object detector we use is a Faster R-CNN [20], with a ResNet101 backbone [21], pre-trained for general purpose object detection on COCO [7]. Even though less resource intensive Faster R-CNN backbones exist, such as MobileNets [29], given our computational budget, we find the Faster R-CNN variant we use to yield the best tradeoff between detection performance and speed (near real-time).

The model's outputs are object bounding boxes and classes with a confidence score for each. The confidence score for the predicted class is obtained as the Softmax probability of the highest scoring class.

We perform fine-tuning of the Faster R-CNN on images consisting of scenes from the environment, where the objects of interest are annotated with bounding boxes and classes. The images we use are video frames, extracted from 9 videos of the AGV navigating the environment while encountering the objects. Considering that the amount of data at our disposal is limited, we have to ensure that the model does not overfit on some, irrelevant properties of the data, e.g., the weather, the relation between objects' position in the frame and their categories, etc. To decrease the influence of these components, and to attempt simulating a diverse evaluation environment to a certain extent, we determine the optimal hyperparameters by training the object detector in a leave-one-out fashion. Namely, we train on a subset of 8 videos and perform evaluation on the remaining one. We iterate this process until we train a separate model on all unique subsets. The final model performance is averaged over each of the videos. We evaluate the model's performance using the

standard COCO [7] mean average precision (mAP). The final model, i.e., the model used in the AGV, is trained on all 9 videos using the hyperparameters determined during the leave-one-out training/evaluation process.

We train the model for 5 epochs with a learning rate of  $1e-4$ . We perform random horizontal flip data augmentation, enabling us to synthetically increase the dataset size and make the detector invariant to such data transformations. We sample a subset of 128 region proposals to estimate the regression and classification loss of the region proposal network.

We quantitatively evaluate each of the trained models on the videos, which were held-out during training (Table 1). The lowest score is highlighted in red, while the highest scoring one in green. Overall, taking the current state-of-the-art of COCO as a reference point (~60% mAP when writing this paper), we observe that the performance is relatively high across all different videos (57.3 mAP). We further observe that the performance on Video 2 (Vid. 2), is significantly lower compared to the average performance. To inspect the reason for the lower performance, we qualitatively inspect the samples from Video 2 as discussed below.

TABLE 1. QUANTITATIVE EVALUATION OF THE VISION AI TRAINED MODEL

|     | Vid. 1 | Vid. 2 | Vid. 3 | Vid. 4 | Vid. 5 | Vid. 6 | Vid. 7 | Vid. 8 | Vid. 9 | Avg.  |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|
| mAP | 55.04  | 40.90  | 56.03  | 66.42  | 68.35  | 50.50  | 65.25  | 61.9   | 51.42  | 57.30 |



Figure 5. (left) all objects are correctly classified, (right) some objects are not detected

We qualitatively evaluate the object detector’s performance by visualizing the predictions on the held-out videos during training. In Figure 5 (left), we observe that the model correctly predicts all objects, which is in line with our expectations as the objects are fully visible and of a reasonable size. On the other hand, in Figure 5 (right) we observe several mis-detected objects of a frame from Video 2. We conclude that even though the model performs well, it struggles to recognize objects, which are (1) far from the camera (small size), and (2) occluded in the environment – both of which are active areas of object detection research.

*B. Visual grounded SLU*

To deal with the data sparsity, and to be able to ground (localize) the speech model output in the image, we perform discretization of the spatial layout (the bounding boxes and classes obtained as output from the object detector). To be specific, we perform mapping of both modalities to a canonical space, where we later measure the similarity between the output of the speech model and each of the detected objects in the image. To that end, we encode each detected object as a collection of one-out-of-k encodings of its category (box, pallet, etc.), material (wooden, plastic, etc.), size (regular or small), and location in the image. Note

that the object category, material and size are jointly predicted by the object detector as the object class. Namely, since the number of possible category + material + size combinations is limited to 6 in our use-case, we encode each combination as a separate class. Lastly, we want to emphasize that such approach does not scale well as the number of possible object categories, materials and sizes increases, however, we leave the decoupling as future work.

Lastly, we quantize the location of the object, i.e., we represent the object’s location based on the object’s horizontal and the lower vertical position. We showcase the grid over the image including the spatial references according to the x and y axis in Figure 6.

Finally, we represent each detected object as a vector of size 12, where we allocate 3, 2, 3, 3, 1 indices for the object’s class, material, x-location, y-location and size respectively. When measuring the similarity between the speech model output (a vector of size 12 as well) and each encoded object detection, we explore different weighting strategies for each object attributes which we discuss next.

*C. Adding Spatial relations*

We evaluate different strategies for measuring the similarity between each (discretized) object detection and the

speech model output. The output is a bounding box, which represents the grounding location of the instruction. We evaluate each grounding strategy on two variants of the dataset, namely (1) a descriptive variant, where the objects are commonly described based on their attributes, e.g., pick up the wooden box, and (2) a spatial variant, where the referred object is described based on its location in the frame, e.g., pick up the box furthest on the left.

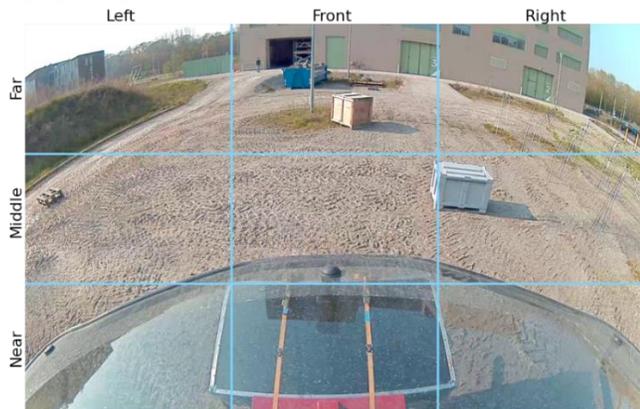


Figure 6. Grid over image with object’s spatial reference

The grounding strategies we evaluate are:

1. Random matching (RM): A naïve baseline, where we ground the speech given instruction to a randomly selected bounding box. We establish a lower bound on the grounding performance with this baseline.

2. Basic matching (BM): We obtain the dot-product between the one-hot encodings of speech instruction and each object detection, representing the similarity.
3. Weighted matching (WM): We (re-)scale the contributions of the individual elements in the dot-product with pre-defined weights.
4. Confidence matching (CM): We represent the speech model with the confidence scores.
5. Weighted confidence matching (WCM): We use confidence scores for the speech model output and additionally weight the individual contributions using the pre-determined weights.

We perform evaluation using the standard grounding accuracy metric, where we score a hit if the predicted grounding bounding box has intersection over union (IoU)>0.5 with the ground truth box. For the random baseline, we perform inference 5 times and report the average performance. For the grounding strategies which weight the importance of each attribute (WM, WCM), we perform grid search over various weight combinations, and establish a weight of 0.1, 0.7, 0.2, and 0.05 for the spatial indicators, the object class, the object material and the object size respectively. We hypothesize that the success of this particular weight combination is a result of (1) the object class and material are essential to ground/locate the referred object, (2) the spatial indicators are somewhat imprecise, but still indicative of the location of the object, and (3) the object size mostly depends on the distance between the AGV camera and the object, which makes it noisy, and should be down weighted. We report the results in Table 2.

TABLE 2. EVALUATION OF THE AI MODEL WITH SPATIAL RELATIONS

| Method     | Dataset type |              |
|------------|--------------|--------------|
|            | Descriptive  | Spatial      |
| RM         | 25.91        | 17.14        |
| BM         | 65.91        | 59.52        |
| WM         | 70.45        | 57.94        |
| CM         | 76.14        | 62.70        |
| <b>WCM</b> | <b>79.55</b> | <b>65.87</b> |

We observe consistent gains when we weigh (WM) or use the speech model confidence scores (CM) in the grounding, compared to the baseline basic matching (BM) method. Additionally, a combination of the weight and confidence matching (WCM) yields superior results across the different data (descriptive, spatial) and significantly outperforms the other methods. Lastly, even though the spatial data is more challenging than the descriptive data, the WCM module performs well, indicating that by re-weighting and adding confidence scores, we can ground spatial speech data reasonably well.

#### V. REINFORCEMENT LEARNING BASED NAVIGATION & ASSOCIATION WITH SPEECH-VISION DATA

In this section, the navigation part of the Multimodal AI and it’s association with the speech-vision data is described. The currently developed proof of concept consists of a simulation environment with the hardware in the loop.

To make this simulator as close to real life as possible, a 3-D scan of the test environment by using an aerial scanning using a drone with photogrammetry capabilities that allows us to map images to a high fidelity 3-D twin of the area. This twin was then imported to the simulator for the purpose of reinforcement learning.

1) *RL architecture & training*

The presented Reinforcement Learning (RL) approach makes use of the DD-PPO (Decentralized Distributed Proximal Policy Optimization) architecture [22] (Figure 7).

The Reinforcement Learning (RL) approach is able to map high dimensional inputs to discrete actions. The DD-PPO model consists of a visual pipeline, for which in our case we use a ResNet18 [21].

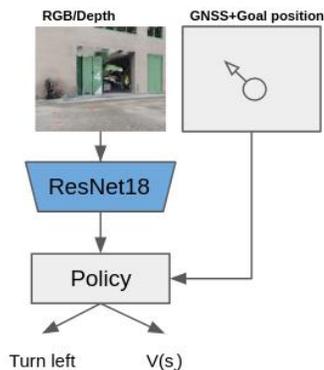


Figure 7. DD-PPO architecture overview

The resulting learned visual representation is concatenated together with a GNSS sensor. This output is then passed onto a recurring policy consisting of 2 Long short-term memory (LSTM) [23] layers. The final outputs of the model consists of a state value estimation, and an action distribution from which actions (move forward, turn left, turn right and stop) can be sampled. The stop-action should be executed by the agent when positioned less than 2 meters of the goal position. As inputs for the model we tested a single depth camera, a single RGB camera, or a combination of both RGB and depth. We use these sensors as they are cheap and widely available. The camera is positioned on the front of the AGV.

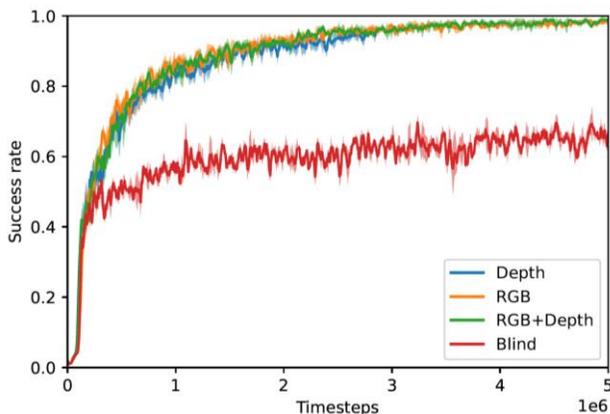


Figure 8. Training performance. The blind agent can perform basic navigation by relying on the GNSS sensor, however to further improve to near perfect results an additional RGB or depth sensor is required to detect and avoid collisions.

To train the agent we use the improvement in geodesic distance between the agent and the goal position as a dense reward signal. A slack penalty of -0.01 is subtracted on each step, and a termination bonus of 2.5 is awarded upon

successfully utilizing the done action. We train the agent entirely in the Habitat simulator [8] where a photorealistic scan of the environment is used. This allows the agent to interact with the terrain in a safe way. While in this case we trained the agent to specifically work on a single environment, DD-PPO also allows generalization to unseen environments, given enough different training environments and training samples. Figure 8 shows the required number of interactions with the environment. These results indicate that in this setting the agent relies mostly on the GNSS sensor, as the blind agent performs reasonably (60% success rate after 5M training interactions). However, by adding either a depth or RGB sensor the agent achieves near perfect navigation capabilities on the training set after 5M interactions with the simulated environment.

2) *RL validation*

Realizing Reinforcement learning on a large autonomous platform brings in multiple challenges to the board. For safety concerns, the approach to validate the system was to use a Hardware-in-loop setup along with the digital twin of the environment. The main input from the real world was the signal from the GNSS receiver (Septentrio AsteRx-U) on the AGV, which was then mapped to the digital twin coordinates system. The GNSS had a dual antenna setup which could then provide the heading of the platform as well. Using a cloud-based service updates were provided in real time to the simulator/digital twin environment to position the simulated tractor same as the one in real world. The output from the simulator was the suggested trajectory to the goal pose.

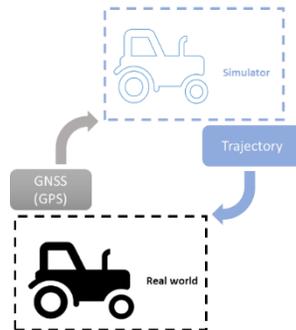


Figure 9. Hardware In Loop setup (overview).

To evaluate the navigation capabilities of the agent, we created a holdout dataset. This holdout dataset contains goal positions the agent did not see during training. Table 3 contains the results of 100 tested episodes. In Table 3, the success rate indicates the amount of episodes the agent could complete successfully. The Success weighted by Path Length (SPL) measurement also considers the length of the path taken.

TABLE 3. SUMMARY OF TESTED EPISODES

| Sensors | Success Rate | SPL    | Avg. Collisions |
|---------|--------------|--------|-----------------|
| RGB     | 100%         | 0.9454 | 0.4355          |
| Depth   | 100%         | 0.8882 | 0.1129          |

|              |        |        |        |
|--------------|--------|--------|--------|
| <b>RGBD</b>  | 100%   | 0.9272 | 0.5161 |
| <b>Blind</b> | 91.94% | 0.7294 | 4.3548 |

### VI. AGV MULTIMODAL AI DEMONSTRATION

To demonstrate the full methodology, we combined the methods respectively described in Sections III, IV and V in one demonstrator implemented in the AGV. We added all the information in a new docker environment to be able to run on the dedicated PC in the AGV. There is a similar user interface compared to the *Speech* (NLP) model (Section III) where you can record your voice and use the NLP model to predict the voice commands. These commands consist of the description of the object and the task the AGV should do. Then the fusion model uses this information to link an object description with a detection from the *Vision* model to predict the location of the describer object on the image. As a last step the lidar data is used to link the 2D location on the image to a 3D location of the object in the world coordinate space. This location can then be sent further as a goal to the control systems together with the described task from the NLP model. A significant improvement could be made in the parameters of the fusion model. There was a bias against using spatial information in the voice command. The material of the object is more difficult to extract on the image than its location, so using the location for finding the correct object is more reliable. Hence, we tuned some of the weights to have a bigger focus on this kind of information. Another small improvement could be made to the audio side. The person dedicated to controlling the AGV added some voice samples and gave feedback to the model through the user interface. This way the model was more confident in recognizing their accent and way of talking. With regards *Navigation*, although the approach is not fully implemented in the rea system, the approach can

already be demonstrated by Hardware-In-the-Loop. In this setting an instance of the simulator is constantly synchronized with the AGV. This is done by using the GNSS position from the real-world AGV to set the position of the agent in the simulator. We can use the digital twin to generate trajectory paths. These generated trajectories can then be used in the real-world by the AGV. A snapshot from the full demonstrator is depicted in Figure 10.

### VII. CONCLUSION

In this paper, we developed and demonstrated a Multimodal AI framework that allows to intuitively instruct production AGVs to perform multiple tasks. The interface with operators is allowed by speech interaction that is decoded through an AI NLP model to translate to interpretable instructions both by AGV controller and the other components of the AI Framework. Associations with Vision and Navigation data is done respectively through an AI detector ad classifier model that recognize different types of objects in a varying environment, as well as a Reinforcement Learning model that estimate the optimal trajectory between the AGV and the objects. The Multimodal AI framework proves to work in different varying conditions where the AGV (Autonomous tractor) is configured to perform different outdoor missions (handling different types of objects such as boxes, pallets, etc.) under different ambient conditions (sunny, rainy, day, night, etc.). The demonstrator remains however a research proof of concept (to demonstrate the approach) and requires different improvements before an effective industrial usage. This includes amongst others, training with larger datasets (speech, vision, navigation) and evaluation in extended number of scenarios. Our research will continue on demonstrating this Multimodal AI approach in other industrial applications where AGVs are typically used, such as in Logistics.

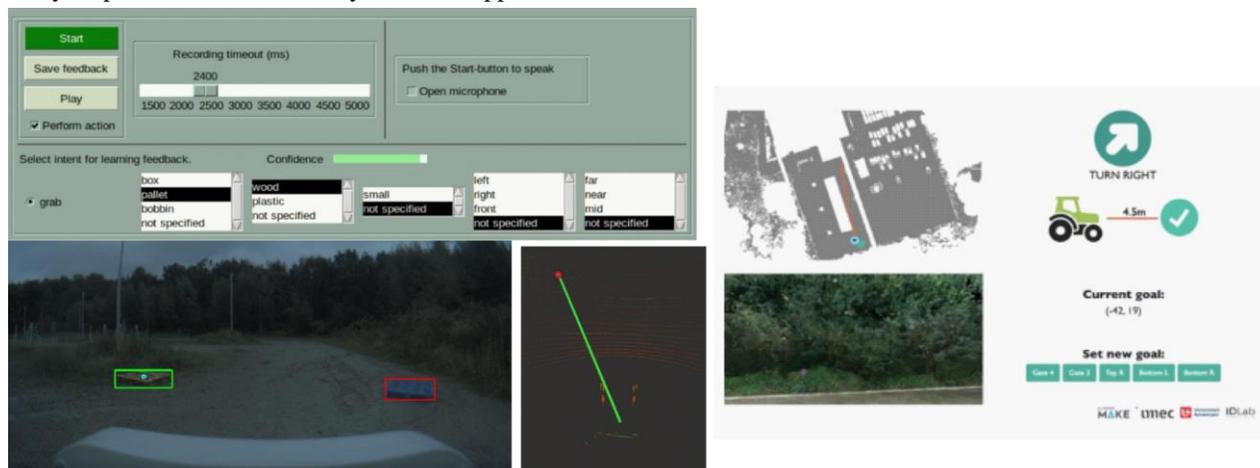


Figure 10. Snapshot of the demonstrator of the AGV Multimodal AI framework: (top left), the *Speech* model interface, (bottom left), the *Vision* model interface, (right), the *Navigation* digital twin interface, (bottom middle), the estimated trajectory between the AGV and the object that is dynamically calculated by fusing all models together

ACKNOWLEDGMENT

This research is supported by AI Flanders that is financed by EWI (Economie Wetenschap & Innovatie) (<https://www.flandersairesearch.be/en>) where all the authors of this publication are collaborating to perform AI research for different applications including industrial applications, and Flanders Make (<https://www.flandersmake.be/en>), the strategic research Centre for the Manufacturing Industry who owns the AGV infrastructure. The authors would like to thank everybody who contributed with any inputs that support to make this publication.

REFERENCES

1. Li, Dong, Bojun Ouyang, Duanpo Wu and Yaonan Wang. "Artificial intelligence empowered multi-AGVs in manufacturing systems." *ArXiv abs/1909.03373* (2019): n. pag.
2. Radder, Laetitia & Louw, Lynette. (1999). Mass customization and mass production. *The TQM Magazine*. 11. 35-40. 10.1108/09544789910246615.
3. M. De Ryck, M. Versteyhe, F. Debrouwere, Automated guided vehicle systems, state-of-the-art control algorithms and techniques, *Journal of Manufacturing Systems*, Volume 54, 2020, Pages 152-173.
4. Herrero-Perez, D. & Barberá, Humberto. (2008). Decentralized coordination of automated guided vehicles. 1195-1198.
5. Mousavi M, Yap HJ, Musa SN, Tahiri F, Md Dawal SZ (2017) Multi-objective AGV scheduling in an FMS using a hybrid of genetic algorithm and particle swarm optimization. *PLoS ONE* 12(3): e0169817.
6. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 6 (June 2017), 84–90. DOI:<https://doi.org/10.1145/3065386>.
7. Lin *et al.*, "Microsoft COCO: Common objects in Context", *European Conference on Computer Vision, ECCV 2014*, p. 740-755
8. Lin, TY. *et al.* (2014). Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) *Computer Vision – ECCV 2014*. *ECCV 2014. Lecture Notes in Computer Science*, vol 8693. Springer, Cham. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
9. Savva, Manolis & Kadian, Abhishek & Maksymets, Oleksandr & Zhao, Yili & Wijmans, Erik & Jain, Bhavana & Straub, Julian & Liu, Jia & Koltun, Vladlen & Malik, Jitendra & Parikh, Devi & Batra, Dhruv. (2019). Habitat: A Platform for Embodied AI Research.
10. Mishkin, Dmytro & Dosovitskiy, Alexey & Koltun, Vladlen. (2019). Benchmarking Classic and Learned Navigation in Complex 3D Environments.
11. Automated off-highway vehicle test platform. [Online]. Available form : <https://www.flandersmake.be/en/testing-validation/product-validation/automated-off-highway-vehicle-test-platform>
12. Real-time testing system (dSpace). [Online] Available from: <https://www.dspace.com/en/pub/home.cfm>
13. Open source robot simulator. [Online]. Available from: <https://cyberbotics.com/>
14. Bart Ons, Jort F. Gemmeke, Hugo Van hamme, Fast vocabulary acquisition in an NMF-based self-learning vocal user interface, *Computer Speech & Language*, Volume 28, Issue 4, 2014, Pages 997-1017,
15. Wang, Pu & Van hamme, Hugo. (2021). Pre-training for low resource speech-to-intent applications. arXiv:2103.16674
16. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.
17. Oostdijk, Nelleke. (2000). The Spoken Dutch Corpus: Overview and first evaluation. *Proceedings of LREC-2000*, Athens. 2.
18. Daniel D. Lee and H. Sebastian Seung. 2000. Algorithms for non-negative matrix factorization. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*. MIT Press, Cambridge, MA, USA, 535–541.
19. ALADIN: Adaptation and Learning for Assistive Domestic Vocal Interfaces. [Online]. Available from: <https://www.esat.kuleuven.be/psi/spraak/downloads/>
20. T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936-944, doi: 10.1109/CVPR.2017.106.
21. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell*. 2017 Jun;39(6):1137-1149. doi: 10.1109/TPAMI.2016.2577031. Epub 2016 Jun 6. PMID: 27295650.
22. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
23. Erik Wijmans, Abhishek Kadian, Ari Morcos, Stefan Lee, Irfan Essa, Devi Parikh, Manolis Savva, Dhruv Batra, (2019, September). DD-PPO: Learning Near-Perfect PointGoal Navigators from 2.5 Billion Frames. In *International Conference on Learning Representations*. arXiv:1911.00357
24. Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. *Neural Comput* 1997; 9 (8): 1735–1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>
25. Kang Tong, Yiquan Wu, Fei Zhou, Recent advances in small object detection based on deep learning: A review, *Image and Vision Computing*, Volume 97, 2020, 103910, ISSN 0262-8856.
26. M. Majdi, M. Deldar, R. Barzamini and J. Jouzdani, "AGV Path Planning in Unknown Environment Using Fuzzy Inference Systems," *2006 1ST IEEE International Conference on E-Learning in Industrial Electronics*, 2006, pp. 64-67, doi: 10.1109/ICELIE.2006.347213.
27. H T, Sreenivas & C, Arjun. (2017). Design of Voice Controlled Automated Guided Vehicle.
28. Aachen Impulse Response Database. [Online]. Available from: <https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>
29. Howard, Andrew G., Zhu, Menglong, Chen, Bo, Kalenichenko, Dmitry, Wang, Weijun, Weyand, Tobias, Andreetto, Marco and Adam, Hartwig MobileNets: Efficient Convolutional Neural Networks for Mobile Vision. Applications. (2017). , cite arxiv:1704.04861

# FPGA Frontend for Highly Efficient Automotive LIDAR Perception

Sanaz Asgarifar  
 Bosch Car Multimedia  
 Braga, Portugal  
 Sanaz.asgarifar@pt.bosch.com

Pedro Barbosa  
 Bosch Car Multimedia  
 Braga, Portugal  
 Pedro.barbosa@pt.bosch.com

Amir Farzamiyan  
 Bosch Car Multimedia  
 Braga, Portugal  
 Amir.farzamiyan@pt.bosch.com

Marcelo Alves  
 Bosch Car Multimedia  
 Braga, Portugal  
 Marcelo.alves@pt.bosch.com

Alexandre Correia  
 Bosch Car Multimedia  
 Braga, Portugal  
 Alexandre.correia@pt.bosch.com

João C. Ferreira  
 INESC TEC,  
 Faculty of Engineering, University of  
 Porto,  
 Porto, Portugal  
 jcf@fe.up.pt

**Abstract**— This paper briefly presents the recent progress on automotive perception and the corresponding hardware implementation for the emerging application of autonomous driving systems. The requirements for an automotive road perception algorithm are presented. An FPGA based design for creating custom implementation of road perception are discussed to form an efficient hardware platform for real-time purpose. A special attention is given to determining the efficiency of hardware implementation in terms of speed and power consumption. Finally, the technical challenges are presented to motivate future research and development in this field.

**Keywords**- Automotive driving; Field Programmable Gate Array (FPGA); Light Detection and Ranging (LIDAR); Road perception; Artificial Intelligence (AI)

## I. INTRODUCTION

The last decade has witnessed tremendous development of autonomous vehicles. These autonomous systems present a great potential for improving safety, increasing productivity, and minimizing their impact on the environment [1]-[3].

Light Detecting and Ranging (LIDAR) system plays an important role in autonomous vehicle systems and recognized as key enabling technology for Advanced Driver-Assisted Systems (ADAS) and autonomous driving, as they enable 3D mapping of objects. Regularly, more than one LIDAR sensors are installed on vehicle for the sense modality of perception, mapping, and localization. One major challenge for LIDAR system is real-time perception.

The core competencies of an autonomous vehicle system are classified into three categories: perception, planning, and control. Figure 1 presents the interactions between these competencies and the vehicles interactions with the environment. Perception refers to the ability of an autonomous system to collect information and extract relevant knowledge from the surrounding environment. The real world is a complex place for partial or even full autonomy, and the lack of

predictability and structure poses serious challenges to the deployment of self-driving vehicles. The autonomous system needs to sense the environment, to determine the exact position on the road, and to decide how it should behave in each situation. That is why self-driving cars are highly dependent on software to bridge the gap between sensor physics and the mechanical actuation of the vehicle, e.g., steering and brakes.

Due to the complexity of perception, there are highly challenging requirements and constraints in terms of real-time determines. Besides, there are limitations on hardware properties, such as size and power consumption. For autonomous vehicles, both real-time processing and low power consumption are desirable.

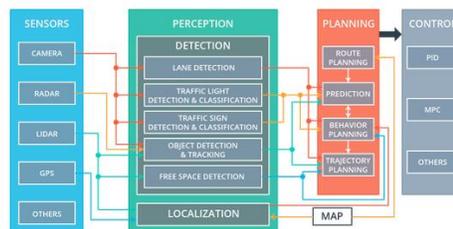


Fig. 1 A vehicle system overview, highlighting core competencies.

Software implementations of LIDAR perception for embedded applications, cannot satisfy the above constraints. Consequently, research has focused on the design of custom hardware architectures for object detection. [4]-[6] Reconfigurable hardware platforms, such as Field Programmable Gate Array (FPGA), have emerged as a very attractive platform for implementing architectures for LIDAR perception applications. FPGAs offer high flexibility with regards to area, power, and performance. They can meet application-specific constraints, which is difficult to achieve with other platforms such as CPUs and GPUs due to fixed interconnect and high-power demands of FPGAs.

## II. METHODOLOGY

Graphics-Processing Unit (GPU) systems have been popular for parallel processing method and implementation of road perception algorithms, but still struggle to comply with the space and power limitations of vehicles. Conversely, a mobile/embedded CPU/GPU system, simply lacks the computing power required for computational processing.

An FPGA can be developed as customized integrated circuit, which is able to perform massive parallel processing and data analysis on a chip. FPGAs emerged in image processing and Deep Learning (DL) due to their incredible benefits of faster mathematical computations, and processing operations. Additionally, capability of recent frameworks allowed to import the work more straightforwardly into an FPGA.

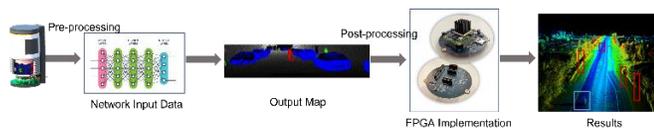


Fig. 2 FPGA implementation of LIDAR perception.

The methodology to implement the FPGA front-end for LIDAR perception consists of the LIDAR data as an input, an intelligent algorithm as a processing tool and top-view predictions as the main output to evaluate the road perception (Figure 2). The algorithm has four steps to process the LIDAR data including:

1. Pre-processing which arranged the input data and created the tensors that can flow through the layers for processing phase.
2. Neural-network processing which do the adaptation of Neural Network (NN) algorithm for perception to improve computational efficiency and suitability for hardware implementation.
3. Evaluation of multiple types of Neural Network in the context of hardware and computational energy efficiency.
4. Development, implementation and testing of reconfigurable hardware architectures exploiting parallelism to increase speed and reduce memory requirements.

Additionally, limited computational resource in an embedded system, raise the need of efficient compression method. To aim this goal, in this paper the new compression algorithm has been design and implemented to compress the point-cloud data and minimize the FPGA memory in post processing. Results obtained from neural network are projected back to targeted views for performance validation. Figure 3 presents the FPGA implementation of compression algorithm for LIDAR point-cloud.

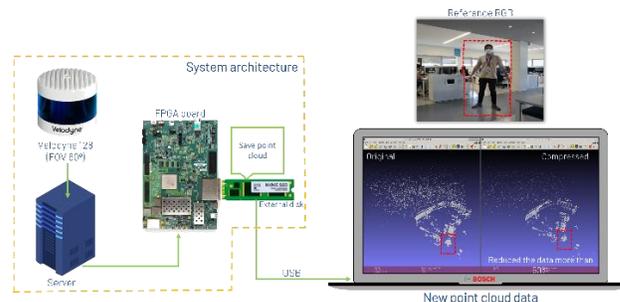


Fig. 3 Compression algorithm implementation for LIDAR point-cloud.

## III. RESULTS AND FUTURE WORKS

The experimental results indicated that the proposed hardware architecture implemented on FPGA, could process each LIDAR scan in 15-18 millisecond, which is significantly faster than convenient hardware implementation and pervious works.

All convolution layers have been taken several milliseconds to complete due to FPGA parallelization. Since LIDAR normally scans at 10HZ, this FPGA implementation fulfils the requirements of the real-time processing. The results of compression algorithm showed significant compress rate of point-cloud data, resulting in speed of entire post-processing stages.

Conclusively, this work introduces the framework for implementation of highly complex pipelines, such as deep learning approach that has the potential to speed-up LIDAR perception processing. Evaluations showed that the proposed LIDAR processing algorithm could achieve state-of-art performance in accuracy and real-time processing. Such a system will have far better real-time determinism than a software-based approach, while providing sufficient computational complexity for object detection and road perception.

Regardless of all progresses in hardware implementation of LiDAR perception, the FPGA implementation still consumes a large amount of on-chip memory. This is the reason that for future work, the Spike Neural Network (SNN) can be considered as a solution to reduce the on-chip memory usage.

### ACKNOWLEDGMENT

“This work is supported by European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project n° 047264; Funding Reference: POCI-01-0247-FEDER-047264].”

### REFERENCES

- [1] J. Levinson et al., *Towards fully autonomous driving: Systems and algorithms*, IEEE Intelligent Vehicles Symposium, Proceedings, 2011.
- [2] R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, *Models, algorithms, and evaluation for autonomous mobility-on-demand systems*, in Proceedings of the American Control Conference, 2015.
- [3] S. D. Pendleton et al., *Perception, planning, control, and coordination for autonomous vehicles*, Machines, 2017.
- [4] Y. Lyu, L. Bai, and X. Huang, *Real-Time Road Segmentation Using LiDAR Data Processing on an FPGA*, in Proceedings - IEEE International Symposium on Circuits and Systems, 2018.
- [5] J. P. Mitchell, C. D. Schuman, and T. E. Potok, *A small, low-cost event-driven architecture for spiking neural networks on FPGAs*, International Conference on Neuromorphic Systems 2020, 2020, pp. 1–4.

- [6] y. lyu, l. bai, and x. huang, *chipnet: real-time lidar processing for drivable region segmentation on an FPGA*, IEEE trans. circuits syst. i, vol. 66, no. 5, pp. 1769–1779, May 2019.

# Self-Aware Industrial Control Systems through Cloud Based Autonomic Computing

Christopher Rouff\*, Ali Tekeoglu\*, Joseph Mauro\*, Alexander Beall\*

\*Johns Hopkins University Applied Physics Laboratory, Critical Infrastructure Protection Group  
email: {christopher.rouff|ali.tekeoglu|joseph.mauro|alexander.beall}@jhuapl.edu

**Abstract**—Critical infrastructure (CI) is being attacked and needs the ability to identify, protect and recover from attacks automatically. Autonomic Computing can provide self-awareness to critical infrastructure so that it can identify and continue to operate through attacks. In this paper, we propose a cloud-based autonomic computing manager that will give Industrial Control Systems (ICS) self-awareness to detect anomalies in their operation, protect themselves and self-organize with other critical infrastructure to thwart attacks.

**Keywords**—Self-Aware Computing, Industrial Control Systems, Cloud Computing.

## I. INTRODUCTION

Historically, Operational Technology (OT) has run on air-gapped private networks; thus, security was achieved through lack of public network access. With the proliferation of cloud computing resources and the Industrial Internet of Things (IIoT) paradigm, OT networks are now connected to enterprise networks for remote access. The security of the ICS networks is now based on the security of the enterprise networks. This has left the ICS vulnerable to malicious actors who can compromise the enterprise networks, and laterally move to have full access to the ICSs components on the mission critical Supervisory Control and Data Acquisition (SCADA) networks.

Detecting malicious activities is now left to OT operators who must detect anomalies on the networks based on their experience and alarms, which malicious actors often circumvent. The autonomic managers will be able to automatically detect anomalies from learning normal OT network behavior and automatically defend the ICSs against attacks, collectively, at network speed.

There have been a number of high-profile attacks on critical infrastructure, including a water treatment plant and a pipeline [1]. If the water treatment attack had not been detected, many people could have been harmed. Though shutting down the pipeline did not cause direct injury, it did affect the economy by reducing the supply of gasoline and keeping people from getting to work.

If the actors wanted to make a major disruption to the economy, they could have attacked multiple facilities and caused major harm to people and businesses. By making critical infrastructure self-aware, these attacks could be thwarted, and other infrastructure could be informed of the attacks. This would cause OT to go into self-protection mode to prevent an attack on their systems and self-organize to recover.

Autonomic Computing (AC) [2] has as its vision the creation of self-managing systems to address today's concerns

of complexity and total cost of ownership while meeting tomorrow's needs for pervasive and ubiquitous computation and communication. Providing security self-awareness to AC-controlled ICSs that can communicate attacks to other AC-controlled ICSs, and collectively defending against these attacks, can provide a higher level of assurance to critical infrastructure.

## II. RELATED WORK

This section presents, to the best of our knowledge, closely related work in the recent literature. In [3], authors describe two different approaches that can provide security assurances to cyber-physical systems: (i) Through the use of micro-services that reconfigure the systems dynamically during attacks or failures, researchers embedded ICSs with autonomic properties to allow them to automatically detect and recover from cyber-attacks and other failures. (ii) Resiliency of autonomous unmanned aerial systems are tested through intelligent agents in a modeling and simulation framework. Researchers in [4] investigated the autonomy of individual cyber-physical systems within a larger cyber-physical system-of-systems (CPSoS), and they looked into potentially insecure and unsafe situations as a result of failures in autonomy.

Another study surveyed the methods used for embodied self-aware computing systems, in application of areas of systems-on-chip control systems, health monitoring and condition monitoring in industrial production systems [5]. Embodied self-aware computing systems are compared to traditional embedded systems. They are defined as being significantly more flexible, robust and autonomous such that they can adapt to a wide range of environmental variation and can cope with deterioration and shortcomings of their own performance.

In [6], authors presented a unified framework for integrating Cyber Physical Systems (CPS) in manufacturing. They utilized an adaptive clustering method for interconnected systems and investigated a case study of self-aware machines by CPS integration. Researchers in [7] surveyed potential challenges that are important in the near future to achieve self-aware smart city objectives. They claimed that cyber-physical systems can extract awareness information from the physical world, thus a holistic approach from the physical to cyber-world is necessary for a successful smart city outcome.

## III. APPROACH

In this paper, we propose a cloud-based autonomic

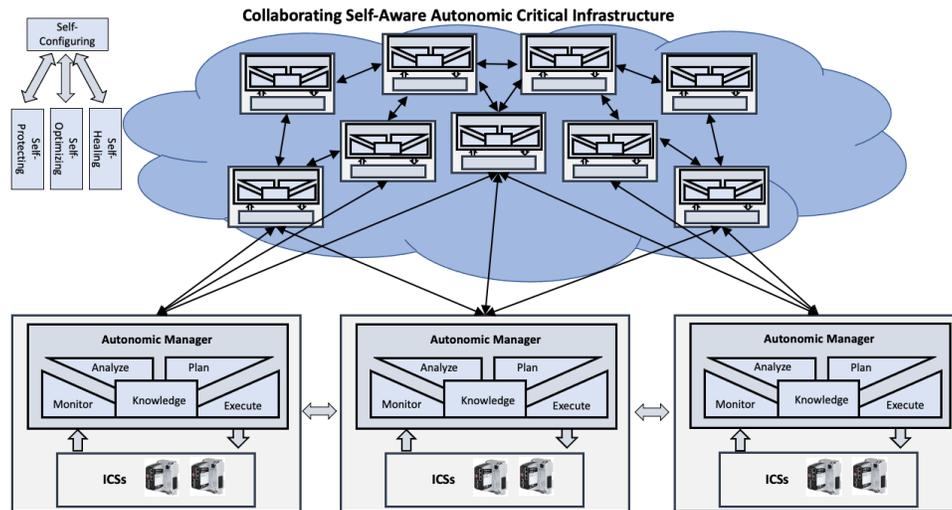


Fig. 1. Cloud Based Self-Aware Autonomic ICS

manager for ICS that will allow them to be self-aware [8] of their operating and computing environment and self-organize with other autonomic managers. Overall system architecture is depicted in Figure 1. Autonomic Computing provides self-awareness properties of self-configuring, self-healing, self-optimizing and self-protecting, among others (referred to as self-CHOP properties) [9].

The proposed design utilizes anomaly detection techniques for ICSs to identify deviations from normal operating conditions. Anomalies will be inputted to a model-base and reasoner in the autonomic manager to identify attacks and failures. Self-CHOP properties will also provide protection and self-healing through reconfiguration and re-optimization. An autonomic MAPE-K (Manage, Analyze, Plan, Execute, Knowledge) architecture will be implemented in the cloud. The cloud will provide the computing resources to store the models and perform the reasoning and computations to implement the self-CHOP properties. The cloud will also provide communications between ICS autonomic managers. This will allow for communication of attacks and provide for self-organization to protect and reconfigure CI components based on the type of attack and the effects on the ICS.

The novelty of the proposed work is the design of autonomic self-aware critical infrastructure that can identify anomalous activity in ICSs, act based on the threat and communicate that threat to other parts of the infrastructure to warn and collaborate with them. Infrastructure that would not be affected by a particular attack would only need to take minimal, or no action. The use of autonomic self-CHOP properties by ICS through cloud resources allows computationally constrained systems to still take advantage of the self-awareness that autonomic computing provides. Cloud-based autonomies prevents overloading an ICS, or needing to upgrade their memory and CPUs to handle the additional load of an autonomic manager.

The model-base and reasoning engine will implement the knowledge component of the autonomic MAPE-K architecture. It will provide the information to detect attacks, protect the ICS, heal, re-configure and re-optimize the ICS. The rest

of the MAPE-K components provide the communications to the ICSs, analyze alternatives when attacked, plan and execute changes to the ICSs, communicate and self-organize with other autonomic managers about attacks.

#### IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed a cloud-based autonomic computing manager that will give ICS self-awareness to detect anomalies in their operation, protect themselves and self-organize with other critical infrastructure to thwart attacks. By working together, autonomic computing enhanced industrial control systems can provide the means for critical infrastructure to have security self-awareness and provide the needed robustness against attacks.

#### REFERENCES

- [1] E. Montalbano, "Florida Water Plant Hack: Leaked Credentials Found in Breach Database," Feb 2021. [Online]. Available: <https://threatpost.com/florida-water-plant-hack-credentials-breach/163919/>
- [2] R. Sterritt, "Autonomic Computing," *Innovations in Systems and Software Engineering*, vol. 1, no. 1, pp. 79–88, 2005.
- [3] J. Maurio, P. C. Wood, S. A. Zanlongo, J. Silbermann, T. I. Sookoor, A. Lorenzo, R. Sleight, J. Rogers, D. Muller, N. Armiger, C. A. Rouff, and L. A. Watkins, "Agile Services and Analysis Framework for Autonomous and Autonomic Critical Infrastructure," *Innovations in Systems and Software Engineering*, pp. 1–10, 2021.
- [4] M. Gharib, L. Dias da Silva, and A. Ceccarelli, "A Model to Discipline Autonomy in Cyber-Physical Systems-of-Systems and its Application," *Journal of Software: Evolution and Process*, vol. 33, no. 9, p. e2328, 2021, e2328 smr.2328.
- [5] H. Hoffmann, A. Jantsch, and N. D. Dutt, "Embodied Self-Aware Computing Systems," *Proceedings of the IEEE*, vol. 108, no. 7, pp. 1–20, July 2020.
- [6] B. Bagheri, S. Yang, H.-A. Kao, and J. Lee, "Cyber-Physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 1622–1627, 2015, 15th IFAC Symposium on Information Control Problems in Manufacturing.
- [7] L. Gurgun, O. Gunalp, Y. Benazzouz, and M. Gallissot, "Self-Aware Cyber-Physical Systems and Applications in Smart Buildings and Cities," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2013, pp. 1149–1154.
- [8] J. Cámara, K. L. Bellman, J. O. Kephart, M. Autili, N. Bencomo, A. Diaconescu, H. Giese, S. Götz, P. Inverardi, S. Kounev, and M. Tivoli, *Self-aware Computing Systems: Related Concepts and Research Areas*. Cham: Springer International Publishing, 2017, pp. 17–49.
- [9] J. Kephart and D. Chess, "The Vision of Autonomic Computing," *IEEE Computer Society, Computer*, vol. 36, no. 1, pp. 41–50, Jan 2003.

# Agility and Semantic Structures to Scaffold Modern Academic Education

## Supporting the Digital Transformation in Higher Education Institutions

*Karsten Böhm*

WEBTA Institute

FH Kufstein Tirol University of Applied Sciences

Kufstein, Austria

email: karsten.boehm@fh-kufstein.ac.at

**Abstract**—Higher Education Institutions (HEI) are faced with a number of challenges in these days: The sector is becoming more competitive and the education profile need to become more dynamic: the amount of knowledge is ever increasing and – especially in the technical subjects – knowledge is out-dating fast and new knowledge is emerging. As a consequence, the education system in HEI needs to become more agile and more modular. The current situation is reflected upon the VUCA paradigm and some solutions will be proposed to address those challenges. As education programs are carefully crafted in a manual but unstructured way, this contribution suggests the support by a scaffolding of semantic structures that helps to create connected, modular education items that can be used as Pre-built information spaces for learning environments. The main contribution of this paper is a concept of a simple and universal model to structure learning outcomes in a more structured form in order to create a foundation for a better understanding for human users and machines alike.

**Keywords**—*Agile Methods; Semantic Web; HEI; VUCA; Digital Transformation.*

### I. INTRODUCTION

The digital transformation of our economies and societies is also a challenge for the sector of Higher Education Institutions (HEI). Education programs become more digital components; this process had been accelerated by the COVID pandemic but is probably also an aspect that is bound to remain a substantial part of the sector. Moreover, the sector is becoming more competitive – at least in the DACH region (Germany [D], Austria [A], Switzerland [CH]) – and HEI have to deal with new and global competitors, also coming from the private sector. At the same time, knowledge to be transferred to students becomes more in terms of volume and complexity and is outdated fast due to technology developments. Lifelong learning as a concept that deals with constant learning is also slowly becoming a reality with the direct consequence that part-time programs become more popular for secondary studies (e.g., in Masters programs). This seems to be especially important and demanding in all the engineering sciences, with the Computer Science and its related disciplines and cross-cutting subjects as probably one of the fastest evolving subjects.

As a result, the education programs offered by HEI have to address these challenges by becoming more modular and open in order to be combined with each other and to be innovated faster. The demanded agility is currently hard to

achieve as education programs are designed and accredited in a process that takes years instead of months and the execution phase usually starts yearly and has a duration of two to three years for one cycle. Thus, innovation in education programs is at the scale of 5-10 years.

The specification of an education process is a careful and manual process and is usually being carried out with extensive text-based documentation. Qualification profiles, competencies and curriculum content is being described in an unstructured way, which makes it hard to validate and aggregate the content from the program level to the lecture level and back. A better structural and semantically oriented support could be useful as a scaffolding support. This could be helpful in all phases of the process (during design, while executing and finally when assessing student results) and Semantic Web technologies are a promising way to support such a scaffolding in terms of providing additional information (meta-data) to trigger and to support directed self-learning services. With that respect the textual specification documents evolve into knowledge-based user interface for students and lecturers alike that can be used for knowledge intensive task in the learning phase (students deciding which module to study next) and the teaching phase (lecturers designing content in a more outcome-oriented way according to the competences that ought to be taught).

While a semantic structuring supports the consistency of programs with respect to learning outcomes and competences it also contributes towards the agility of the programs by exposing interfaces in terms of learning outcomes that could connect to other education programs within an HEI or outside of an HEI. And finally, agility might also be of interest for the student who is increasingly involved in Problem-Based Learning scenarios (PBL) in which knowledge is being created and used, based on the tasks given.

The rest of the paper is structured as follows: In Section II the current situation of HEI is analyzed with respect to the VUCA paradigm. VUCA is the abbreviation for a number of properties of certain environments: V – Volatility, U – Uncertainty, C – Complexity and A – Ambiguity [1][2]. A number of high-level suggestions for addressing those challenges is made, again employing the VUCA solution paradigm with an interpretation to the domain of HEI [3][4]. After that the concept of hierarchic competence matrices is introduced in section III with the focus on connected application throughout the different levels of an education program. The support for the creation and execution in terms of semantic scaffolding using Semantic specification

documents is introduced in Section IV. The paper concludes in the final Section V with an outlook at future research activities.

## II. HIGHER EDUCATION IN A VUCA-WORLD

Higher Education is process that is spanning several years in each iteration and prepares students for their professional life. Consequently, long innovation cycles in their curricula are the result which does not suit well in a dynamic and fast changing environment [5]. The concept of a VUCA-world is commonly used to describe environments with a large degree of change and the required frequent adaptation to a changing or new environment. The current changes in the digital transformation of society and economy are also having an impact on the HEI and thus creating a VUCA-environment for education programs that HEIs need to address.

In the following, the author will analyze how the different aspects of VUCA can be addressed in the execution level and the management level of education programs in HEI. Both levels are quite interconnected as the execution level is dependent on a certain flexibility of the lecturing in a certain while still maintaining the original learning goals of the curriculum. Furthermore, it is important that the novel and agile approaches still exhibit a wide range of scalability from small groups to large groups and from presence teaching to distant teaching as well. At management level, it becomes more important, that individual lectures are dynamically orchestrated into a curriculum, that might be more individualized that structures education programs in the past. At the same time, individualization and the interfacing to prior knowledge and post-graduate education becomes more important.

Now each of the components of a VUCA world in the context of HEI are dealt with in more detail:

(V) – Volatility in the context of HEI can be interpreted in terms of changing topics that are concerned relevant and/or interesting by the external stakeholders (students, companies), but also in terms of volatile group sizes due to those trends. As programs are designed and funded in the long run, adaptation to those volatile aspects is becoming a challenge.

(U) – Uncertainty in HEI can be interpreted as the fact that the education system is currently changing by external drivers like the digital transformation, the lasting effects of the COVID pandemic and changing expectations of future generation of students. As topics and education profiles are changing and new job profiles are emerging or did not even exist yet, education planning for the long run is difficult – development and financing phased is not designed for such an uncertain environment and thus has constantly to adapt (which requires resources to employ the change). Uncertainty in the economy also leads to an increased amount of part time profiles of working students. This leads to a demand of lifelong learning and lifelong education, that both sides (students and HEI) are not yet ready for.

(C) – Complexity is a common pattern in fast evolving subjects as most engineering programs and can be interpreted in two ways: an increasing complexity of the fields in terms of subjects becoming broader as well as subjects having a deeper level of knowledge that is needed to master it. This is

leading to the observation, that educating students becomes a challenge, as the time for education remains the same (2-3 yrs. in BA/MA) and could not be extended. Part-time students add to the difficulty of the problem as their time budgets are even more limited. For lecturers it becomes more relevant to select the right content and to moderate the learning process – in that sense they are becoming guides and curators for the knowledge that is transferred to the students. In order to keep a close contact to the practical application they need to work with real-world examples but also need to convey more general pattern that are relevant to their field.

(A) – Ambiguity in the context of HEI can be interpreted in the fast-evolving knowledge domain in many subjects. Concepts that attain a lot of general interest like Digital Transformation, Artificial Intelligence and the Cloud technologies, for example have multiple meanings and require different levels of knowledge to become actionable. Understanding those concepts and applying them in real-life scenarios if often the requirement for engineering students but often to conflict with the expectation from the general public what technology could do or could not do.

Different stakeholder groups (e.g., business, technology, society) have a quite different understanding and expectations. For the education this leads to the challenge of educating students in an informed and adequate way without the danger of getting too much in the complexity trap.

The solution space in a VUCA-world used the same letters with a different meaning and in the following an interpretation in the context of HEI could is proposed:

(V) – vision is adhering to the fact that an operational guidance is needed to navigate through changing topics and still retain a USP or core competence for the HEI. Here, the situation is similar to those of companies, but might be interpreted in a different way, e.g., it could be value driven in the sense that it is important to develop applicable knowledge, or to employ a guiding attitude to the education of the students. Such a vision needs to be employed in practice and is therefore more of a cultural value that is developed by employing a vision.

(U) – understanding could be seen as an active and ongoing reflection process on the requirements of the application domain (business & society) but also the expectations and requirements of the current and next generations of students to act accordingly. More than in the past this is also an interplay and a communication of values between different generations: lecturers (“older generations”) and students (“newer generations”). That always had been the case, but in a VUCA world this process is being accelerated and in the sense of a dialogue more important.

(C) – clarity in a VUCA world is an important role that HEI can play: by building on existing knowledge and by employing scientific methods and an objective view of the world HEI can help students to provide orientation in a complex and changing world and convey them important tool-sets to navigate in that world at topics that they are faced later in their life.

(A) – agility becomes more important on the strategic level (for the development of programs) and the operational level

(the execution of programs). While the former is required to develop and adapt programs faster the second aspect addresses the aspect on how the execution of a program is tailored to the specified group of students in front of the lecturer and even towards the individual. This leads to the fact that learning analytics becomes more important and that digitization can be used to provide additional or alternative learning paths.

### III. HIERARCHICAL SKILL MATRICES FOR SCAFFOLDING COMPETENCES IN EDUCATION PROGRAMS

In order to address the challenges for education programs derived by the VUCA analysis a more structured approach on designing and executing education programs is needed. Currently those programs are specified mostly using informal descriptions a non-standardized structure for the curricula. Especially for the core concepts of competencies that are the result of the learning outcomes of the individual lectures of a program no specific support structure is being provided. This approach has two important drawbacks: (1) The problem of *consistency* among the different abstraction layers of a program and the different interpretation at design time and execution time. (2) a limited support for *modularity*, as there are no clearly defined interfaces among the different components. Modules in current programs are designed around its content but do not clearly expose the competences in a way that could act as an ‘interface’ for other modules. Making competences more visible and structurally comparable might help to create more consistent programs that provides modules, which can be reused more easily.

In order to address those issues, the concept of hierarchical Competence Matrices is being created. The concept roots in the observation that in modern Curriculum Vitae (CV) competences are often presented in a more visual form of scales, as illustrated in Figure 1 below. Although being not exactly precise the representation delivers an immediate profile of the competencies of a person, albeit on a very general scale.

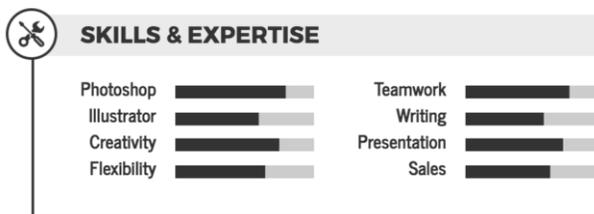


Figure 1. Template of a skills & expertise in CVs [6]

A second observation is taken from the domain of foreign language competences. Here, the Common European Framework of Reference for Languages (CEFR) established an accepted framework for language proficiency that encodes a rather complex subject in just six reference levels (A1, A2, B1, B2, C1, C2).

Both models are simple to understand, and despite their generality widely used. This contribution builds on those ideas and extends it towards a hierarchical relation among different

competence or skill matrices. A *Competence Matrix (CM)* is a set of competences that are identified with a brief description and are assigned an attribution according to a skill or competence level that could use Bloom’s taxonomy [7] or a similar model [8][9]. Such a taxonomy defines a number of levels at which a competence could be classified, such as remember, understand, apply, analyze evaluate and create.

The central idea of *hierarchical concept matrices (HSM)* is the connection of different but related CMs to create a relation along different levels of an education program (general to specific) and from different parts of the life cycle of such a program (from design to execution)

The following Figure 2 illustrates a number of HSM along three levels

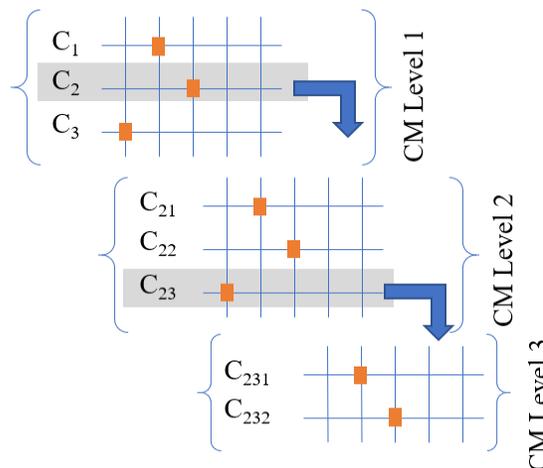


Figure 2. Illustration of HSM along three levels that are interconnected, e.g., on program level (1), at syllabus level (2) and at the level of a lecture being held (3). Scale and structure of the HSM remain the same and a direct connection of  $C_1 \rightarrow \{C_{21}, C_{22}, C_{23} \equiv \{C_{231} C_{232}\}$  can be derived.

The HSM could be connected in a strict hierarchic fashion, leading to a taxonomy of competences that are being constructed from more general (at program level) to more specific (at the level of instruction). In practice it might be the case that a competence is being provided or needed by different CMs at a deeper level, which leads to a graph like structure as cycles might be a part of the hierarchical structure. Another way of connecting the CMs would be the interpretation as a function that maps a set of competences (at a lower level) to a single competence (at a higher level), changing its attribution value  $attr: f: \{C_1, C_2, C_3, \dots\} \rightarrow attr(C)$ .

The most important feature is the connectivity of the CMs that resembles the HSM, because it constitutes a connected competence model for all phases of an education program of an HEI that can be used for students and lecturers alike an explanation model to provide guidance and structuring. In that sense it really provides a scaffolding for designing and executing education at HEI. Moreover, it lays the foundation for a validation of a competence model by enabling to check for missing links or competences that are taught but never used, for example. This could help to improve the education modules in an iterative way and to ensure quality with respect to didactic design and content.

#### IV. A PREBUILT INFORMATION SPACE FOR LEARNING

The concept of HCM contribute to the notion of Pre-built Learning Information Spaces for Learning Environments [10] that are aiming at supporting students in a VUCA world to address the needs of the future in Higher Education efficiently. It provides the component to connect different specifications documents together that are otherwise unconnected, such as syllabi of different lectures or the syllabus and the learning content in a learning management system (LMS).

HCM could be embedded in specification documents that contain semantic encoding, so called *Semantic Specification Documents (SSD)* – for a detailed description, see [10]. In essence, an SSD is a document that contains metadata in a semantic encoding that models textual descriptions in a machine-readable way; see Figure 3 below.

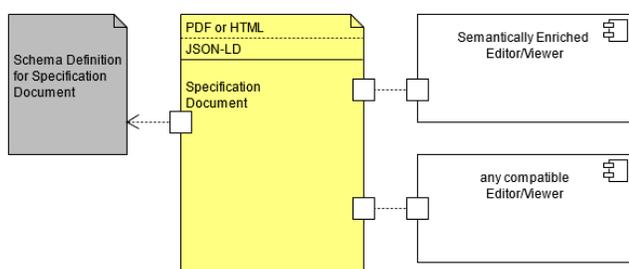


Figure 3. Semantically enriched documents as containers for HCM [10]

The important aspect of such an SSD is the fact that the semantic annotation is directly attached to the document itself and is available to creators and viewers of the document without the need for an external IT-system. It is self-contained. In order to retrieve the information that is dispersed over different SSDs, an aggregation step is needed. Such an aggregation would collect and analyze all the documents of an education program and collect the extracted semantic information in a specific database (called a triple store). Queries would be run against this store in order to satisfy an information demand; see Figure 4 below.

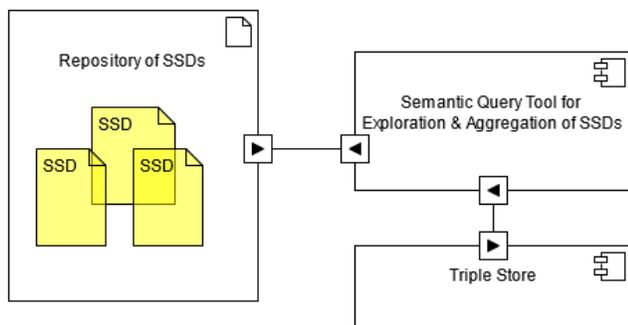


Figure 4. Aggregating of SSDs in a repository of documents to provide querying against a triple store [10]

The combination of HCM as a conceptual model with SSDs as a distributed and document centric representation model helps to work towards a concept of semantically defined learning specifications that can be used to describe the benefits of explicit formal descriptions to structure the curricular content in HEI to build semantically and agile learning environments (“SALE”) in order to support the teaching and learning process at the level of meaningful learning objectives. It also provides interfaces to prior knowledge and post-graduate education, which will become more important in a world of true lifelong learning as it is required in a VUCA world.

#### V. CONCLUSION

This contribution analyzed the current situation of the HEI with the help of the VUCA paradigm and presented how a more structured, modular and agile approach is needed to address the changes that are ahead of the sector in the process of digital transformation. In order to provide a more structured approach, it is important to involve all stakeholders in the process with an easy to understand and easy to use model to express competences at different levels. Hierarchical Competence Matrices are introduced for that respect and the concept of semantic specification documents can be used to implement this tool set. The main contribution of HCM to VUCA is the increased focus on connected and explicit learning outcome that are specified in a uniform way. This improves clarity (“C”) and helps to students and lecturers to understand (“U”) the intention (or vision (“V”) of the education program better. Due to the exposed structure it also helps to make missing aspects or changing topics more visible and provide intervention points for improving programs in an agile (“A”) way, helping designers of course programs. For the student, however, the structure can be used to provide agile teaching support by considering his or her progress using learning analytics, derived from the interaction with the LMS.

Based on this conceptual work the next phases of this research are going to provide an implementation and codification of the hierarchical competence matrices from the bottom up starting with modeling of a single lecture to a whole program and then potentially to the programs of a whole faculty. Evaluation of those implementation both in terms of understanding and scaffolding support will be the next steps in the research that is being carried out of the author who is both in the lecturer and the researcher role, thus being able to interpret findings from multiple perspective together with colleagues and students.

#### ACKNOWLEDGMENT

The author would like to thank the Tyrolean Science Fund (“Tiroler Wissenschaftsförderung”), which supported this research under grant number F.33280/6-2021 and the remarks of the reviewers of this contribution that helps to improve it.

REFERENCES

- [1] A. Codreanu, "A VUCA Action Framework for a VUCA Environment. Leadership Challenges and Solutions," *J. Def. Resour. Manag. JoDRM*, vol. 7, no. 2, pp. 31–38, 2016.
- [2] C. Pangaribuan, F. Wijaya, A. Djamil, D. Hidayat, and O. Putra, "An analysis on the importance of motivation to transfer learning in VUCA environments," *Manag. Sci. Lett.*, vol. 10, no. 2, pp. 271–278, 2020.
- [3] S. Green, A. F. Page, P. De'ath, E. Pei, and B. Lam, "VUCA Challenges on the Design Engineering Student Spectrum," presented at the 21st International Conference on Engineering & Product Design Education (E&PDE 2019), 2019. doi: 10.35199/epde2019.100.
- [4] G. Baskoro, "Designing a Master Program to Cope with the New and Next Normal (VUCA World, Industry 4.0, and Covid 19): a case study," *IPTEK J. Proc. Ser.*, no. 3, Art. no. 3, Oct. 2021, doi: 10.12962/j23546026.y2020i3.11078.
- [5] V. C. Tassone, C. O'Mahony, E. McKenna, H. J. Eppink, and A. E. J. Wals, "(Re-)designing higher education curricula in times of systemic dysfunction: a responsible research and innovation perspective," *High. Educ.*, vol. 76, no. 2, pp. 337–352, Aug. 2018, doi: 10.1007/s10734-017-0211-4.
- [6] Affde Marketing, "20+ Infographics CV-templates and degisn suggestions to get the job." <https://www.affde.com/de/infographic-resume-template.html> (accessed Apr. 11, 2022).
- [7] B. S. Bloom, *Taxonomy of Educational Objectives: The Classification of Educational Goals*. D. McKay, 1956.
- [8] L. Anderson, D. Krathwohl, and B. Bloom, "A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives," undefined, 2000. Accessed: May 05, 2021. [Online]. Available: /paper/A-Taxonomy-for-Learning%2C-Teaching%2C-and-Assessing%3A-A-Anderson-Krathwohl/23eb5e20e7985fca5625548d2ee6d781a2861d41
- [9] V. Prikshat, S. Kumar, and A. Nankervis, "Work-readiness integrated competence model: Conceptualisation and scale development," *Educ. Train.*, vol. 61, no. 5, pp. 568–589, Jan. 2019, doi: 10.1108/ET-05-2018-0114.
- [10] K. Böhm, "Towards Semantically Enriched Curricula as pre-built Information Spaces in Higher Education Institutions," in *ECKM 2021 22nd European Conference on Knowledge Management*, 2021, p. 71.