ADAPTIVE 2024

The Sixteenth International Conference on Adaptive and Self-Adaptive Systems and Applications

ISBN: 978-1-68558-153-4

April 14 - 18, 2024

Venice, Italy

ADAPTIVE 2024 Editors

Steve Chan, VTIRL, VT/I-PAC, USA

ADAPTIVE 2024

Forward

The Sixteenth International Conference on Adaptive and Self-Adaptive Systems and Applications (ADAPTIVE 2024), held on April 14 – 18, 2024, continued a series of events targeting advanced system and application design paradigms driven by adaptiveness and self-adaptiveness. With the current tendencies in developing and deploying complex systems, and under the continuous changes of system and application requirements, adaptation is a key feature. Speed and scalability of changes require self-adaptation for special cases. How to build systems to be easily adaptive and self-adaptive, what constraints and what mechanisms must be used, and how to evaluate a stable state in such systems are challenging duties. Context-aware and user-aware are major situations where environment and user feedback is considered for further adaptation.

The conference had the following tracks:

- Self-adaptation
- Adaptive applications
- Adaptivity in robot systems
- Fundamentals and design of adaptive systems
- Computational Trust for Self-Adaptive Systems
- Assurances and metrics for adaptive and self-adaptive systems

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the ADAPTIVE 2024 technical program committee, as well as the numerous reviewers. The creation of a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to ADAPTIVE 2024. 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 also gratefully thank the members of the ADAPTIVE 2024 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope ADAPTIVE 2024 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of adaptive and self-adaptive systems and applications. We also hope that Venice provided a pleasant environment during the conference and everyone saved some time to enjoy this beautiful city

ADAPTIVE 2024 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

ADAPTIVE 2024 Steering Committee

Constantin Paleologu, University Politehnica of Bucharest, Romania
Claudia Raibulet, University of Milano-Bicocca, Italy
Sebastian Herold, Karlstad University, Department for Mathematics & Computer Science, Sweden
Andreas Rausch, TU Clausthal, Clausthal-Zellerfeld, Germany
Marc Kurz, University of Applied Sciences Upper Austria, Faculty for Informatics, Communications and Media, Austria
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France
Yukio Hayashi, Japan Advanced Institute of Science and Technology, Japan

ADAPTIVE 2024 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

ADAPTIVE 2024

Committee

ADAPTIVE 2024 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

ADAPTIVE 2024 Steering Committee

Constantin Paleologu, University Politehnica of Bucharest, Romania
Claudia Raibulet, University of Milano-Bicocca, Italy
Sebastian Herold, Karlstad University, Department for Mathematics & Computer Science, Sweden
Andreas Rausch, TU Clausthal, Clausthal-Zellerfeld, Germany
Marc Kurz, University of Applied Sciences Upper Austria, Faculty for Informatics, Communications and Media, Austria
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France
Yukio Hayashi, Japan Advanced Institute of Science and Technology, Japan

ADAPTIVE 2024 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

ADAPTIVE 2024 Technical Program Committee

Habtamu Abie, Norwegian Computing Center/Norsk Regnesentral-Blindern, Norway
Harvey Alférez, Universidad de Montemorelos, Mexico
Raid Al-Nima, Northern Technical University, Iraq
Felipe Arevalo, Rice University, USA
Mark Balas, Texas A&M University, USA
John S. Baras, University of Maryland, College Park, USA
Nik Bessis, Edge Hill University, UK
Glen Bright, University of KwaZulu Natal, Durban, South Africa
Antonio Brogi, University of Pisa, Italy
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France
Constantin F. Caruntu, “Gheorghe Asachi” Technical University of Iasi, Romania
Angel P. del Pobil, Jaume I University, Spain
Laura-Maria Dogariu, University Politehnica of Bucharest, Romania
Ibrahim Abdallah Abbas Atwa Elgendy, School of Computer Science and Technology | Harbin Institute of Technology, China
Holger Eichelberger, University of Hildesheim, Germany
Fairouz Fakhfakh, University of Sfax, Tunisia

Francisco José García-Peñalvo, University of Salamanca, Spain
Zhiwei Gao, Northumbria University, UK
Yukio Hayashi, Japan Advanced Institute of Science and Technology, Japan
Sebastian Herold, Karlstad University, Department for Mathematics & Computer Science, Sweden
Koen Hindriks, Vrije University Amsterdam, Netherlands
Christopher-Eyk Hrabia, Technische Universität Berlin | DAI-Labor, Germany
Marc-Philippe Huget, Polytech Annecy-Chambery-LISTIC | University of Savoie, France
Imène Jraidi, McGill University - ATLAS Lab, Canada
Yasushi Kambayashi, Nippon Institute of Technology, Japan
Michael Katchabaw, Western University, London - Ontario, Canada
Christoph Knieke, Technische Universität Clausthal, Germany
Marc Kurz, University of Applied Sciences Upper Austria - Faculty for Informatics, Communications and Media, Austria
Mikel Larrea, University of the Basque Country UPV/EHU, Spain
Yichuan Li, University of Nevada Reno, USA
Rajini M., PES University, India
Mieke Massink, CNR-ISTI, Italy
James E. McCarthy, Instructional Systems - Sonalysts Inc., USA
René Meier, Lucerne University of Applied Sciences and Arts, Switzerland
Philippe Merle, University of Lille, France
Andreas Metzger, University of Duisburg-Essen, Germany
Mehrdad Moharrami, University of Illinois at Urbana-Champaign, USA
Filippo Neri, University of Napoli "Federico II", Italy
Karol Niewiadomski, University of Wuppertal, Germany
Ashley Oiknine, DCS Corporation / Army Research Laboratory / University of California, Santa Barbara, USA
Krzysztof Okarma, West Pomeranian University of Technology in Szczecin, Poland
Joanna Isabelle Olszewska, University of West Scotland, UK
Ali Pakniyat, University of Alabama, USA
Constantin Paleologu, University Politehnica of Bucharest, Romania
Marcello Pellicciari, University of Modena and Reggio Emilia, Italy
Marcin Pietron, University of Science and Technology in Cracow, Poland
Agostino Poggi, Università degli Studi di Parma, Italy
Claudia Raibulet, University of Milano-Bicocca, Italy
Andreas Rausch, TU Clausthal, Clausthal-Zellerfeld, Germany
Juha Röning, University of Oulu, Finland
Joerg Roth, Nuremberg Institute of Technology, Germany
José Santos Reyes, University of A Coruña, Spain
Jagannathan (Jag) Sarangapani, Missouri University of Science and Technology, USA
Ichiro Satoh, National Institute of Informatics, Japan
Melanie Schranz, Lakeside Labs GmbH, Austria
Vasco N. G. J. Soares, Instituto de Telecomunicações / Instituto Politécnico de Castelo Branco, Portugal
Erik Sonnleitner, University of Applied Sciences Upper Austria - Faculty for Informatics, Communications and Media, Austria, Germany
Mohammad Divband Soorati, University of Southampton, UK
Cristian-Lucian Stanciu, University Politehnica of Bucharest, Romania
Roy Sterritt, Ulster University, UK
Justyna Swidrak, August Pi & Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain / Institute

of Psychology - Polish Academy of Sciences, Warsaw, Poland
Nikolaos Tsiogkas, KU Leuven, Belgium

Copyright Information

Table of Contents

Agent-based Modeling in the Edge Continuum using Swarm Intelligence

Melanie Schranz, Kseniia Harshina, Peter Forgacs
Lakeside Labs
Klagenfurt, Austria
{schranz, harshina, forgacs}@lakeside-labs.com

Fred Buining
HIRO MicroDataCenters
Voorburg, Netherlands
fred.buining@hiro-microdatacenters.nl

***Abstract*—The edge continuum presents a dynamic and evolving paradigm in the future's world of computing, offering a versatile and efficient solution for a wide range of applications and industries. The edge infrastructure is more challenged in its stability and performance because of more stringent latency and autonomy requirements, distribution across multiple sites, their local limited size, multi-tenancy and multi-operators, local management, with components being concurrent and asynchronous. This paper introduces an innovative framework that combines agent-based modeling and swarm intelligence to address complex challenges such as resource allocation, workload scheduling, and data management in the edge continuum. This framework, at the core of the architecture, enhances edge autonomy, reduces latency, improves energy efficiency, and optimizes cloud connectivity by applying agent-based modeling. By integrating autopoietic characteristics like self-organization, regeneration, and regulation, the system dynamically adapts to changing conditions. Two candidate algorithms, the hormone algorithm and ant algorithm, emulate decentralized decision-making processes observed in nature. The paper reviews related work in swarm intelligence for network optimization and emphasizes the need for distributed, agent-based solutions. This research paves the way for robust, adaptive, and scalable systems in the complex edge environment, promising emergent behaviors and enhanced efficiency. In this position paper, we propose the edge continuum with its characteristics and limitations as a novel field of application for swarm intelligence by conceptually proposing agent-based modeling and simulation.**

***Index Terms*—Swarm Intelligence, Bio-inspired Algorithm, Edge Continuum, Agent-Based Modeling**

I. Introduction

The emergence of local processing capacity at the edge is driven by numerous advantages essential for upcoming processing tasks. These benefits encompass heightened security and reliability, alongside reduced latency and energy consumption. The management of the edge infrastructure, the so-called edge continuum, presents a dynamic computing landscape. Within the edge continuum, for which we consider a mesh of Edge Micro Data Centers (EMDCs) in this paper (see Figure 1), intelligence is spread across the edges forming a distributed environment. This will make the edge more autonomous and fine-grained in local decision making within a regional context and make it more independent from a central coordination point. This is especially necessary, if we talk about real-time applications such as autonomous driving or monitoring and control of smart grids. The edge infrastructure is more challenged in its stability and performance because of more stringent latency and autonomy requirements, distribution across multiple sites, its local limited size, multi-tenancy and multi-operators, local management, with components being concurrent and asynchronous. This challenge to edge infrastructures is growing rapidly due to the increasing i) number of connected devices and their data-producing and data-consuming capabilities, ii) intelligence embedded in edge devices, iii) atomization of monolithic applications, iv) scale, speed, and complexity of edge device interactivity in a zero-trust environment. Resource allocation, workload scheduling, and data management are challenges that increase in the complexity of the edge orchestration and edge-cloud interaction (see Figure 1 for a schematic architecture provided by the ACES project).

This position paper introduces a conceptual, but novel framework that combines agent-based modeling and swarm intelligence as an emergent orchestrating mechanism to address these complexities. Agent-based modeling and swarm intelligence are known for providing advantages in simulating complex systems with autonomous entities including adaptability, scalability and robustness. They utilize collective decision-making processes as observed in nature by swarms of insects, fish or birds [1]. Central to our approach is the integration of these autopoietic characteristics that include the emergent intelligence of self-organization, regeneration, and regulation. These characteristics enable the system to dynamically adapt and optimize in response to changing conditions. AI-driven optimization methods (including swarm intelligence) in cloud infrastructure are successfully being researched (see Section VI for more details). Among recent notable examples of utilization of swarm intelligence to optimize complex systems, is the work of Schranz et al. [2], where authors successfully utilize bottom-up job shop scheduling applying swarm intelligence algorithms for optimizing a large production plant. Thus, we propose the edge continuum with its characteristics and limitations as a novel field of application for swarm intelligence.

This framework is at the core of the architecture, required to manage the edge infrastructure, EMDCs capable of processing big data and AI at the edge-to-edge environment independent from a distant cloud. Key to our conceptual approach is the use of swarm agents, representing demand and supply entities.

 ISBN: 978-1-68558-153-4

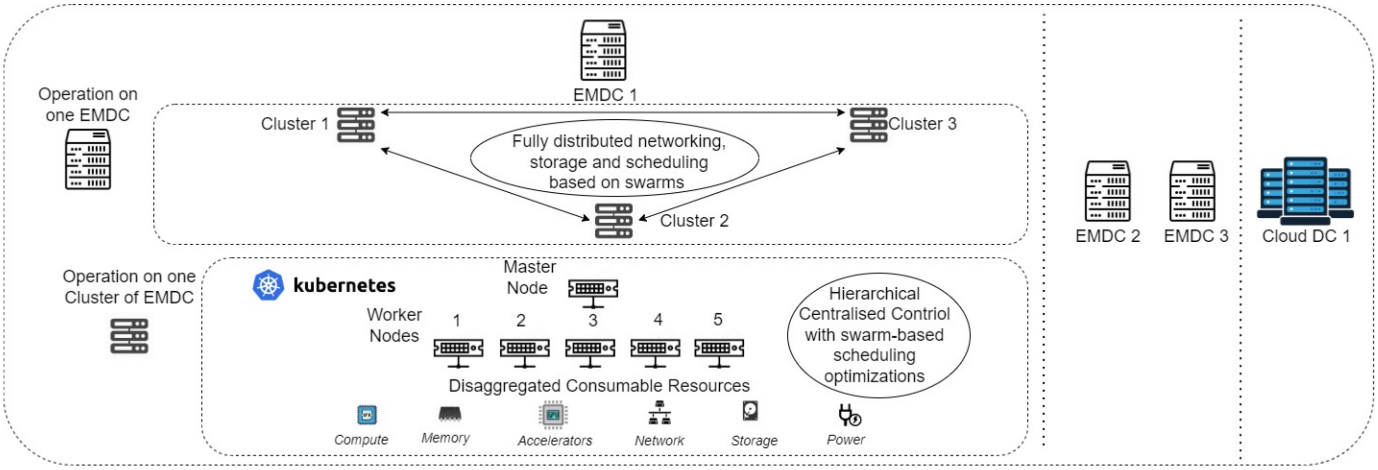

Fig. 1. Schematic architecture demonstrating the inter-edge resource allocation in clusters (nodes, pool of resources - see Section III - and the overall edge-cloud interaction.

Demand swarm agents represent workload behaviors at the pod level, ensuring pod-level optimization. On the other hand, supply swarm agents represent node dynamics. These agents collaborate within an EMDC environment, orchestrating processes such as workload placement, storage management, and caching optimization. Exemplary swarm algorithms, in this paper the hormone and ant algorithms are utilized in order to accomplish the desired functionality of the system. For example, demand swarm agents deploy synthetic hormones to communicate their requirements and priorities. Supply swarm agents, detect these hormones to make informed allocation decisions. The ant algorithm dynamically optimizes workload-node-assignments by simulating the foraging behavior of ants, depositing pheromones to guide subsequent decisions.

The paper follows the following structure: Section 1 introduces the problem and outlines specific challenges of the research process. In Section 2, we explore the edge continuum setting. In Section 3, we discuss the agent-based model, and address challenges specific to the infrastructure and interactions in the EMDCs. In Section 4, we evaluate candidate swarm algorithms, the hormone and ant algorithms on a conceptual basis. We conclude the paper with Section 5 by reviewing the related work and giving an outline on future work in Section 6.

II. The Edge Continuum

Industry surveys show that edge infrastructure is a driver for new initiatives and business operations. According to recent studies, the percentage of respondents that have or will implement edge locations within the next three years raised from 55% last year to 87% this year [3]. According to the EU, a decentralized intelligently orchestrated edge infrastructure (hardware and software) is needed to support platforms, data spaces and applications, e.g., an Industrial Metaverse utilizing a combination of cloud, edge and IoT to enable a wide range of new solutions to transform processes, automate operations, and launch new products and services. At the edge where traffic patterns are becoming sudden and unpredictable, more traffic will be handled in far edge or metro edge data centers, micro edge data centers embedded in metro networks [4].

The edge is created from a mesh of EMDCs that are completely composable, customizable, heterogeneous, and therefore different in capacity and capabilities. An EMDC will supply resources; nodes with servers, accelerators, storage, and networking capabilities. While the application owners and edge devices will demand certain resources, capabilities, and performances, the orchestration of matching demand and supply are made dynamically and decentralized in a bottom-up approach. This will allow the EMDC to be configured and sized to satisfy 1) local autonomy: the demands of the local edge clients, and 2) regional elasticity: the demands upon groups of EMDCs caused by fluctuating local demand and mobile demand traveling along several EMDCs in a region. This means that each EMDC needs to be:

- aware of its hardware configuration, hardware capabilities, software services (supply) and the local and regional requests for services (demand);
- self-intelligent and autonomous (autopoietic AI) in matching local and regional supply and demand;
- efficient in the execution of its services and self-intelligence.

Currently, hardware-based composability of an EMDC means that the nodes can be of any type, such as Central Processing Unit (CPU), Field Programmable Gate Array (FPGA), Graphical Processing Unit (GPU), or Non-Volatile Memory express (NVMe).

III. Creating an Agent-Based Model for Optimizing the Edge Continuum

When shaping the edge continuum to an agent-based system, we analyze a group of possible swarm agents and their attributes. In this context, we need to determine the eligibility of an entity to serve as a member of the EMDC swarm [5]. The swarm can exhibit homogeneity (with all agents being of the same kind, like numerous pods) or heterogeneity (comprising agents of various types, such as pods and resources). For an entity to qualify as a swarm member, it should possess the capacity to effectively function within a swarm. This entails the presence of a significant number of other swarm members (for instance, a single instance of an FPGA, existing in isolation, would not make a suitable swarm member). Additionally, the entity should exhibit an appropriate degree of abstraction to facilitate modeling, possess the capability to sense dynamic information from the immediate environment, respond to information originating from the local vicinity (such as making decisions), and be logically coherent and comprehensible, fostering trust in the proposed solution [1].

A. Modeling Agents in the Edge Continuum

Our agent-based approach introduces two distinct types of swarm agents: demand swarm agents and supply swarm agents. These agents collaborate within an EMDC environment, orchestrating processes such as pod placement, storage management, and caching optimization. The model for the problem consists of an edge continuum with resources, queues, pods, and processes.

1) Demand Swarm Agents: An application is split into a set of services $\mathfrak{S}$ that are represented as a set of related pods $P^s = \{p_1^s, p_2^s, \dots\}$ with s as the specific service. Each service s is defined by a compilation of resources R^s which prescribes the processing steps necessary to compute the individual pods. The pod p_j^s can choose which of the suitable nodes N_i^n to use for each necessary process step P^r.

2) Supply Swarm Agents: The EMDC $\mathfrak{E}$ contains several sets of nodes or nodes, consisting of different types of resources $N^r = \{N_1^r, N_2^r, \dots\}$, where r is/are the resource type(s). A node with different resources presents a typical EMDC node, whereas a node with a single resource presents, e.g., a CPU that is part of a pool of resources. In the course of this work, we consider the following resources along with their respective capacities: CPU, FPGA, RAM, and NVMe. Each resource N_i^r has a queue Q_i^r.

3) Agent Collaboration and Self-Organization: The interaction between demand swarm agents and supply swarm agents is orchestrated through swarm intelligence algorithms. Demand swarm agents autonomously seek out the most suitable node for workload placement, while supply swarm agents determine the optimal workload to process based on available resources and capacity. This collaborative decision-making process enables the system to efficiently allocate workloads to nodes, optimizing processing, latency, and resource utilization.

Our agent-based model is designed to exhibit autopoietic characteristics, fostering self-organization, regeneration, and regulation within the edge continuum. As demand and supply agents interact and adapt to changing workloads and resource availability, the system as a whole displays emergent behaviors that contribute to its resilience and efficiency.

IV. Challenges in Modeling Agents for the EMDC

In the agent-based modeling of an EMDC, we face a set of challenges that need to be considered in the modeling process.

A. Pool of Resources

Additionally to the nodes in an EMDC, we consider a pool of resources that presents an innovation to the current definitions of the edge continuum. This means that besides the processing capabilities in a node (that is a constitution of multiple resources), single resources can be requested for pod processing. This pool of resources is part of the EMDC and can be consulted by the edge(-cloud) management as requested. Such a pool mainly prevents resource limits, increased latencies, and stability of the performance of other pods, as their assigned resources are not tapped. Currently, the Compute Express Link (CXL) is being implemented in CPUs (Intel, AMD), in memories and storage (Samsung) and the PCIe switches are expected in 2025. Besides the hardware development, the biggest challenge currently is how these pools of resources can be orchestrated.

B. Application Types

For the different services, we can differ between the three application types that come with diverse requirements in their response time.

- The **Long-Running Applications (LRAs)** instantiate long-standing pods to enable iterative computations in memory or unceasing request-response. LRAs include processing frameworks (e.g., Storm [6], Flink [7], Kafka streams [8]), latency-sensitive database applications (e.g., HBase [9] and MongoDB [10]), and data-intensive in-memory computing frameworks (e.g., TensorFlow [11]).
- **Batch processing** is typically used when you have a large amount of data that needs to be processed all at once, and when the results of that processing can be stored and used later. Data is typically processed on a schedule or at regular intervals. There are two types of batch processing: Regular returning requests, and opportunistic requests with little to no SLA (Service Level Agreement).
- **Stream processing** also deals with large volumes of data, but the data needs to be processed in real-time.

Future workloads will become even more complex with LRAs, batches, and stream processes being interconnected. Therefore, it will be challenging to categorize an application and tune its agents accordingly.

C. Relationships among Pods

The demand swarm agents are related pods P^s split from a specific service s. These pods can have several relations with each other. There can be different needs, e.g., that they need to be processed in parallel or that they depend on each other. Additionally, if one pod is too slow, the current system creates more pods to reach the given response times of the specific service s. Currently, these relationships are not used in the scheduler and orchestration optimization. For example, placing interacting services closer together can significantly enhance their performance, e.g., i) if there are multiple services with microservices that frequently interact, it is advisable to locate the microservices of one service within the same region to improve performance, ii) for pods that are heavily dependent

 ISBN: 978-1-68558-153-4

on a database, it is best to place them near the database to reduce latency and improve overall performance.

V. Candidate Algorithms

In this section, we introduce two candidate algorithms for the edge continuum. These algorithms adopt a bottom-up approach, by modeling real-world entities as agents, including the attributes that enable them to interact with each other and their environment. By applying swarm intelligence principles to these agents, we enhance their capabilities to manage the complexity in the edge continuum. This allows them to make context-aware decisions by drawing from both local and global information. This approach embraces decentralized decision-making, promising effective resource management in the complex edge(-cloud) continuum.

A. Hormone Algorithm

Artificial hormone systems draw inspiration from the biological endocrine system, which regulates various metabolic processes within our bodies [12], [13]. This creates a self-organizing system characterized by scalability, adaptability, and robustness. In our simulation, supply swarm agents correspond to nodes within the continuum, and demand swarm agents represent pods seeking optimal node placement.

Demand swarm agents release synthetic hormones into the environment based on their resource requirements and preferences. These hormones carry information about the demands and priorities of the pods. Supply swarm agents, representing nodes, detect these hormones and adjust their behavior accordingly. Nodes release their own hormones indicating resource availability and capacity.

The concentration of hormones guides demand swarm agents toward nodes that match their requirements, fostering autonomous and informed decision-making. The communication of synthetic hormones replaces traditional centralized control mechanisms with decentralized coordination, allowing the system to adapt to pod variations and resource fluctuations.

The underlying principle is inspired by the use of artificial hormones for reorganizing agents in self-organizing systems for technical applications [14], [15], which can be extended to the dynamic edge environment. In our framework, we will implement the artificial hormone system as a software layer distributed across the processing nodes within the edge continuum as inspired by the applications in production plants (see Elmenreich et al. [16] for more details). The hormone algorithm used for optimization in the edge continuum can be dissected into six key mechanisms:

Production: Supply swarm agents, representing nodes, produce hormones in response to the number of demand swarm agents, pods, in the EMDC. Nodes that are currently processing fewer pods produce more hormone. Each node as well as a pool of resources (e.g., CPU, storage) may produce a distinct type of hormone with

$$H^r = \frac{1}{|Q_i^r| + \beta} \tag{1}$$

where H_i^r is the hormone corresponding to the the node N_i^r, β is a smoothing factor, and $|Q_i^r|$ is the number of waiting workloads in the EMDC for the node N_i^r.

Evaporation: The hormone levels at each node gradually decrease over time through a process of evaporation, controlled by a parameter α given with

$$H_{i,t+1}^r = H_{i,t}^r \cdot (1 - \alpha) \tag{2}$$

where $H_{i,t+1}^r$ and $H_{i,t}^r$ represent the state of hormone at the node N_i^r before and after a discrete evaluation step.

Diffusion: Hormones diffuse from one node to another based on the compilation of resources per pod that also connects the resources similar to hormone propagation in biological systems. Hormones move upstream, following the reverse of this resource graph by calculating

$$\Delta H = H_i^r \cdot \gamma \tag{3}$$

$$H_i^r - = \Delta H \tag{4}$$

where ΔH is the amount of hormone moving upstream from the node N_i^r, and γ is a parameter setting the motility of hormone.

The link strength $l^{r,p}$ between two nodes N_i^r and N_j^p is equivalent to the number of compilations of resources R_t containing processes P^r and P^p in direct succession. Each node connected upstream receives a proportional part of the upstream hormone with

$$H_j^p + = \Delta H \frac{l^{r,p}}{\sum_c l^{r,c}} \tag{5}$$

where $\sum_c l^{r,c}$ represents the sum of all upstream links from P^r.

Diffusion through pod movement: When pods move between nodes within the EMDC, e.g., due to a lack or loss of resources in a node, they carry hormones with them, influencing the hormone levels at both the initial and destination nodes.

$$\Delta H = H_i^r \cdot \delta \tag{6}$$

$$H_i^r - = \delta H \tag{7}$$

$$H_j^p + = \delta H \tag{8}$$

where ΔH defines the amount of hormone that moves with the pod, calculated from the amount of available hormone H_i^r at the node N_i^r.

Attraction: Pods are attracted by the nodes whose processing capabilities match the pods' requirements from the corresponding compilation of resources. The amount of attraction decreases exponentially based on the order of the node. The attraction force is applied to pods as soon as they enter the processing queue Q_i^r and it can make the pod move towards a distinct node.

$$attraction = \sum_{i,r} H_i^r \cdot \varepsilon^n \tag{9}$$

where H_i^r is the hormone amount at a node that is n edges away, and ε is a factor < 1 defining the degradation of the hormone attraction over edge distance in the graph G.

Each mechanism comes with a parameter indicating the strength of each part, that is evaporation rate (α), hormone production factor (β), upstream diffusion factor (γ), hormone distribution factor (δ), and attraction factor (ε). A possible configuration of these parameters is stated in [16]. Due to the interaction between each of the mechanisms forming feedback control loops, the algorithm can operate with a broad set of possible parameter settings.

B. Ant Algorithm

Ant algorithms draw inspiration from the decentralized foraging behavior of ants, a natural phenomenon where ants can efficiently find near-optimal paths to food sources without relying on global knowledge. They achieve this by leaving pheromone trails to communicate with other ants. In our simulation within the edge continuum, this concept will be applied to optimize the allocation and processing of pods by supply swarm agents, analogous to ants, representing nodes within the continuum.

Here's how the ant algorithm is adapted to the edge continuum:

Trail Following: In our context, we frame the allocation of pods as a routing problem in the edge continuum. Pods probabilistically select the next suitable node N_i^n from the set of potential nodes N^n based on both local pheromone values associated with that node and a local heuristic considering the node's current pod, which can be assessed by metrics like queue length or resource utilization. The probability $P_{i,j}$ of selecting node N_i^n is computed as shown in Equation 10.

$$P_{i,j} = \frac{\tau_{i,j,d} + \alpha\eta_{i,j}}{1 + \alpha(N_i - 1)} \tag{10}$$

with

$$\eta = 1 - \frac{q_{i,j}}{\sum q}. \tag{11}$$

In this equation, η represents the relative queue length of node with N_i as the number of possible nodes. The parameter α allows for fine-tuning the influence of pheromone τ (see update rules below) versus the local pod heuristic. In our adaptation, the destination d corresponds to the next step in the pod's compilation of resources within the EMDC, rather than a specific destination node.

Trail Laying: Pheromone values are updated after a pod has been processed on a node within the EMDC. However, unlike traditional ant algorithms where backward ants are used to update pheromone values, we utilize communication and coordination among nodes within the continuum. Each pod maintains a memory of the processing, effectively measuring the time it waited for resources. When a pod moves from one node to another, this information is used to update the pheromone values.

For a chosen node N_x^n, the pheromone value is updated as follows:

$$\tau_{x,d} \leftarrow \tau_{x,d} + r(1 - \tau_{x,d}) \tag{12}$$

For all potential nodes N_n^n that were not chosen, the pheromone values are updated according to

$$\tau_{n,d} \leftarrow \tau_{n,d} - r\tau_{n,d}. \tag{13}$$

The reinforcement r depends on the processing time of the pod, which reflects the waiting time and resource utilization. This approach ensures that nodes with shorter pod processing times and lower resource utilization become more attractive for incoming workloads.

Evaporation: Periodically, pheromone values are subject to evaporation with a rate p. This process simulates the natural fading of pheromone trails and helps remove paths that may have become less optimal due to changes in resource availability or demand (Equation 14).

$$\tau(t+1) = \tau(t)(1-p) \tag{14}$$

This adaptation effectively models and optimizes the allocation and processing of pods within the EMDC, drawing inspiration from the decentralized behavior of ants.

VI. Related Work

Next-Generation Networks (NGN) are growing fast, and this rapid growth is becoming more and more demanding to optimize resource management in cloud computing, edge computing, and edge-cloud computing. As big data analytics is gaining size, optimization is becoming problematic, because those optimizers, which seek an exact global optimum, can have an exponentially growing complexity. Some examples of optimization problems in big data analytics that can exhibit expensive computational complexity:

Combinatorial Feature Selection: When dealing with a large number of features (variables) in a dataset, selecting the optimal subset of features for a machine learning model can be computationally intensive. The number of possible feature combinations grows exponentially with the number of features, leading to exponential complexity [17], [18].

Clustering in High-Dimensional Spaces: In high-dimensional spaces, clustering algorithms like k-means can become computationally expensive. As the number of dimensions increases, the data points tend to become more distant from each other, making it challenging for clustering algorithms to identify meaningful clusters. This phenomenon is often referred to as the curse of dimensionality [19], [20].

Optimizing Distributed Systems: Optimizing the allocation of computing resources in distributed machine learning systems for big data analytics can be computationally expensive. These systems often involve multiple nodes and parallel processing of large data sets. Ensuring efficient resource allocation to reduce training time and resource waste is a challenging optimization problem [21], [22].

Pham et al. [23] do an in-depth review of the implementation of swarm intelligence for NGN, and state the advantage of swarm intelligence in guaranteed convergence, robustness, near-optimal solution, and computationally-traceability. There are several research works addressing the major issues in NGN using swarm intelligence. The way they are approaching is by creating a random set of solutions. This set of candidate solutions is improved iterations by iterations optimizing the objective function, which quantifies the goodness of a solution. The review also mentions possible swarm intelligence implementations for spectrum management and resource allocation, wireless caching and edge computing, network security, and miscellaneous issues. The following lines of this section present some even more recent research done in the optimization of task offloading in edge computation than the ones presented in the review.

In smart homes, to minimize the energy consumption of a residential consumer-centric load-scheduling, Lin & Hu [24] proposed a constrained Particle Swarm Optimization (PSO) algorithm, where the possible solutions are modeled as agents. Feng et al. [25] also describe a task offloading strategy, which is able to reduce the energy consumption, the time latency and the service price in mobile edge computing, however, the strategy is to use a Grey Wolf Optimizer (GWO), Whale Optimization Algorithm (WOA) and GWO – WOA where the agents are the percentages of how much of a mobile device's task is computed locally on the mobile device (because a task can also be partially offloaded to the edge server). This means that whenever there is a change in the tasks or the mobile devices, the algorithm needs to compute the optimal solution with a new set of input. Lee et al. [26] provides a swarm intelligence algorithm, an Artificial Bee Colony (ABC) algorithm for the allocation of a given task set to a given edge server set and a cloud, where the solutions from the solution population are interacting with each other. A relatively recent work, Mahenge & Sanga [27] presents a strategy to offload resource-intensive tasks in mobile edge computing energy-efficiently using a hybrid approach (PSO and GWO), where the algorithm gathers the information about the tasks and servers and then calculates the optimal offloading strategy. Bacanin et al. [28] perform energy optimization in 5G-enabled edge nodes using PSO which first has to obtain as input all the data about the tasks and edges. Attiya et al. [29] aims to tackle the problem of IoT application task scheduling. It uses the Manta Ray Foraging Optimization (MRFO) combined with Salp Swarm Algorithm (SSA) which is also initialized with a set of N solutions. In Singh et al. [30], the authors write all available resources into an availability list. On this list, a swarm algorithm (Ant Colony Optimization, ACO) is executed for searching an optimized (centralized) solution for resource allocation and scheduling. Another approach is presented in de Melo et al. [31]. Here, the focus is on a review of several methodologies to parallelize swarm algorithms on parallel hardware to increase execution performance. The aim is to accelerate finding an optimal solution to a problem which is then mostly applied in a centralized manner. No decentralized agent-based approach is revealed in this work.

Although the proposed solutions in the literature apply different swarm intelligence algorithms, they are executed centrally. Typical problems that arise from this approach are single point of failure, higher computational effort, and lack of dynamicity to occurring changes in the environment or incoming demands. Therefore we propose an optimization from the bottom-up, using swarm intelligence literally with interacting embodied agents that make decisions based on local information. These are then the algorithms that are robust, adaptive, and scale due to their distributed characteristic leading to a real emergent behavior of a complex system.

To the best of our knowledge, no one has ever tried an agent-based approach in the edge-cloud domain where resources and requests are regarded as agents, and scheduling along with relevant objectives (utilization, low latency, energy efficiency, etc.) are considered emergent properties of the agent's local decision making and interaction (autopoiesis). This will direct the NGN of EMDCs on the edge to a powerful, self-organized network, where we can generate a main contribution towards the scheduling of a resource pool and the dynamics of pod arrivals.

VII. Conclusion and Future Work

The management of the edge continuum presents a multi-faceted computing landscape that continues to grow in complexity. Our paper introduces a novel conceptual framework that leverages agent-based modeling and swarm intelligence to address these complexities, focusing on enhancing edge autonomy, reducing latency, improving energy efficiency, and optimizing cloud connectivity.

We propose utilizing swarm agents to represent demand and supply entities within an Edge Micro Data Center (EMDC) environment. Demand swarm agents optimize at the pod level, while supply swarm agents manage node dynamics. The application of swarm algorithms, including hormone and ant algorithms, facilitates intelligent workload placement, storage management, and caching optimization.

As a next step, we will elaborate on the efficiency of the proposed candidate algorithms using a simulation approach based on an abstracted version of the edge continuum using SwarmFabSim [32], a NetLogo implementation, as inspiration. Additionally, real-world implementation and validation of the framework will be essential to demonstrate its practical effectiveness in managing the dynamic edge(-cloud) landscape. This is a step that will need some workarounds first, as the hardware to realize resource pools on the edge is still in the development phase.

Acknowledgement

This work was performed in the course of the EU-project ACES[1] supported by European Union's Horizon Europe re-

[1] https://www.aces-edge.eu/

search and innovation programme under the grant agreement No. 101093126 (HORIZON-CL4-2022-DATA-01-02).

References

[1] M. Schranz, G. A. Di Caro, T. Schmickl, W. Elmenreich, F. Arvin, A. Şekercioğlu, and M. Sende, "Swarm intelligence and cyber-physical systems: concepts, challenges and future trends," *Swarm and Evolutionary Computation*, vol. 60, p. 100762, 2021.

[2] M. Schranz., M. Umlauft., and W. Elmenreich., "Bottom-up job shop scheduling with swarm intelligence in large production plants," in *Proceedings of the 11th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, pp. 327–334, INSTICC, SciTePress, 2021.

[3] B. Kleyman, B. Gillooly, and K. Letourneau, "The 2021 state of the data center report," *AFCOM*, p. 20, 2021. https://datacenterworld.com/sites/default/files/AFCOM_State%20of%20the%20Data%20Center_FINAL_2021_5-10-21.pdf [Online; accessed 8-March-2024].

[4] P. Fetterolf, "Next-generation metro network and edge computing architectures," *ACG Research*, p. 10, 2021. https://www.acgcc.com/media/reports/files/ACG_Research_Next-Generation_Metro_Network_and_Edge_Computing_Architectures_2021_1.pdf [Online; accessed 8-March-2024].

[5] M. Schranz, M. Umlauft, M. Sende, and W. Elmenreich, "Swarm robotic behaviors and current applications," *Frontiers in Robotics and AI*, vol. 7, p. 36, 2020.

[6] A. Storm. https://storm.apache.org/. [Online; accessed 8-March-2024].

[7] A. Flink. https://flink.apache.org/. [Online; accessed 8-March-2024].

[8] A. Kafka. https://kafka.apache.org. [Online; accessed 8-March-2024].

[9] A. HBase. https://hbase.apache.org/. [Online; accessed 8-March-2024].

[10] MongoDB. https://www.mongodb.com/. [Online; accessed 8-March-2024].

[11] TensorFlow. https://www.tensorflow.org/. [Online; accessed 8-March-2024].

[12] A. Sobe, W. Elmenreich, T. Szkaliczki, and L. Böszörmenyi, "Seahorse: Generalizing an artificial hormone system algorithm to a middleware for search and delivery of information units," *Computer Networks*, 2015.

[13] A. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, vol. 237, no. 641, pp. 37–72, 1952.

[14] W. Elmenreich, R. D'Souza, C. Bettstetter, and H. de Meer, "A survey of models and design methods for self-organizing networked systems," in *International Workshop on Self-Organizing Systems*, pp. 37–49, Springer, 2009.

[15] W. Elmenreich and H. de Meer, "Self-organizing networked systems for technical applications: A discussion on open issues," in *International Workshop on Self-Organizing Systems*, pp. 1–9, Springer, 2008.

[16] W. Elmenreich, A. Schnabl, and M. Schranz, "An artificial hormone-based algorithm for production scheduling from the bottom-up.," in *International Conference on Agents and Artificial Intelligence*, pp. 296–303, 2021.

[17] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proceedings of the 20th international conference on machine learning*, pp. 856–863, 2003.

[18] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1060–1073, 2022.

[19] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Proceedings of the 8th International Conference on Database Theory*, pp. 420–434, Springer, 2001.

[20] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia computer science*, vol. 148, pp. 291–302, 2019.

[21] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *2nd USENIX Workshop on Hot Topics in Cloud Computing*, 2010.

[22] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, "A survey on distributed machine learning," *ACM Computing Surveys*, vol. 53, no. 2, pp. 1–33, 2020.

[23] Q. Pham, D. Nguyen, S. Mirjalili, D. Hoang, D. Nguyen, P. Pathirana, and W.-J. Hwang, "Swarm intelligence for next-generation networks: Recent advances and applications," *Journal of Network and Computer Applications*, vol. 191, p. 103141, 2021.

[24] Y.-H. Lin and Y.-C. Hu, "Residential consumer-centric demand-side management based on energy disaggregation-piloting constrained swarm intelligence: Towards edge computing," *Sensors*, vol. 18, no. 5, p. 1365, 2018.

[25] S. Feng, Y. Chen, Q. Zhai, M. Huang, and F. Shu, "Optimizing computation offloading strategy in mobile edge computing based on swarm intelligence algorithms," *EURASIP Journal on Advances in Signal Processing*, vol. 7, no. 36, pp. 1–24, 2021.

[26] C. Lee, Y. Huo, S. Zhang, and K. Ng, "Design of a smart manufacturing system with the application of multi-access edge computing and blockchain technology," *IEEE Access*, vol. 8, pp. 28659–28667, 2020.

[27] M. Mahenge, C. Li, and C. Sanga, "Energy-efficient task offloading strategy in mobile edge computing for resource-intensive mobile applications," *Digital Communications and Networks*, vol. 8, no. 6, pp. 1048–1058, 2022.

[28] N. Bacanin, M. Antonijevic, T. Bezdan, M. Zivkovic, K. Venkatachalam, and S. Malebary, "Energy efficient offloading mechanism using particle swarm optimization in 5g enabled edge nodes," *Cluster Computing*, vol. 26, pp. 587–598, 2023.

[29] I. Attiya, M. Elaziz, L. Abualigah, T. Nguyen, and A. El-Latif, "An improved hybrid swarm intelligence for scheduling iot application tasks in the cloud," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6264–6272, 2022.

[30] H. Singh, A. Bhasin, and P. R. Kaveri, "Qras: Efficient resource allocation for task scheduling in cloud computing," *SN Applied Sciences*, vol. 3, pp. 1–7, 2021.

[31] B. A. de Melo Menezes, H. Kuchen, and F. Buarque de Lima Neto, "Parallelization of swarm intelligence algorithms: Literature review," *International Journal of Parallel Programming*, vol. 50, pp. 1–29, 2022.

[32] M. Umlauft, M. Schranz, and W. Elmenreich, "Swarmfabsim: A simulation framework for bottom-up optimization in flexible job-shop scheduling using netlogo," in *Proceedings of the 12th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, pp. 271–279, 2022.

To Refurbish or not to Refurbish? Towards an AI-based Evaluation System for Power Tool Batteries

Dominique Briechle, Marit Briechle-Mathiszig, Tobias Geger, Nelly Nyeck, Robert Werner
Institute for Software and Systems Engineering
University of Technology Clausthal
Clausthal-Zellerfeld, Germany
dominique.fabio.briechle@tu-clausthal.de; marit.elke.anke.mathiszig@tu-clausthal.de;
thomas.tobias.marcello.geger@tu-clausthal.de; nelly.nicaise.nyeck.mbialeu@tu-clausthal.de;
robert.werner@tu-clausthal.de

***Abstract*—The earth's resources are limited. Nevertheless, humans use more natural resources every year than the earth can provide. For that reason, sustainable usage of products is needed. Refurbishing processes offer an opportunity to extend the lifecycle of products like accumulators. For the refurbishing process, it is important for the operator to not only know the condition of the product but as well the possible expenditures it will cost to restore its functioning condition. The question whether it is possible to determine this kind of information about an accumulator from its external image has not yet been answered. Investigating this question can support the velocity of decision processes of whether a battery should be refurbished or given directly to recycling. This work describes the development of a refurbish and data collecting service and the design of a concept for adjunct data evaluation to investigate if Artificial Intelligence can draw a connection between the external features of an accumulator and the internal condition of the same. The preliminary results include the conception of the service as well as the derivation of assumptions based on the so far collected images of the batteries.**

***Index Terms*—Circular Economy, Recommendation, Digital Service Design, Product Lifecycle, Refurbishing, Artificial Neural Networks**

I. Introduction

The depletion of global resources is at an all-time high. Although the environmental impact of the currently conducted linear economy is drastically decreasing the quality of life in many countries, the broad expansion is a long time coming [1]. This can, however, be traced back to a few governing factors that are preventing the transition from the before-mentioned economic model toward a Circular Economy. Therefore, it is important to focus on those barriers, find solutions to remove them and support the transformation to a sustainable society. The main reason for the scarce repairing, refurbishing, and remanufacturing (3Rs) landscape in many countries is the monetary and time expenditure owners have to raise, which makes the continuation of product lifecycles difficult in comparison to the alternative of dumping the product and, in the best case, recycling its materials [2]. However, recycling quotas are still limited, and the resources which can be retracted are limited by the recycling plant installations and technical boundaries [3]. In terms of environmental sustainability, the 3Rs have an advantage in comparison with the production of a new good of the same make. We address the aforementioned obstacles, making 3R more attractive to both end users and service operators (e.g., companies that have the knowledge and capabilities to repair, refurbish, and remanufacture products) [4]. The conducted experiment therefore focused on two different thematic priorities: the first one is the design of the different aspects of a refurbishing process system, while the second one is the setup of a data recording pipeline that can catch the significant details of the specific power tool batteries to enable the development of Artificial Intelligence (AI)-based support tools. The present short paper aims to provide an overview of the refurbishing structure used for the restoring of the functioning conditions of the tool batteries, as well as the data foundation used for the further conduction of the battery assessment. The paper is therefore structured in section 2 in the Background and State of the Art section, followed in section 3 by the description of the initial Project Setup and the conduct of the project. This section is followed in section 4 by a review of the Preliminary Results. The paper closes in section 5 with a Conclusion regarding the usage of the recorded data in an AI-based support system to enhance end-user and operator knowledge.

II. Background and State of the Art

The development of the AI requires a understanding in both the domain-wise background and the technology-wise. Therefore this chapter is separated into two sub-chapters.

A. Domain Background

Fundamentally, repair entails restoring functionality to damaged or malfunctioning products, while refurbishment revitalizes products to nearly match their original state, thereby prolonging their lifespan [5]. In the context of the circular economy, repair and refurbishment are essential as they tackle the root causes of malfunctions, encourage the preservation of

 ISBN: 978-1-68558-153-4

resources, and reduce the carbon footprint of producing new products. Although repair procedures are consistent with the concepts of sustainability, they encounter a significant obstacle due to the societal mindset of prioritizing convenience over sustainability, resulting in a preference for new products over refurbished ones [6]. To promote repair- and refurbishment-centric services, technological improvements, such as applying AI-based technologies in streamlining refurbishment systems, greatly contribute to improving their effectiveness and accessibility, as consumers' perceptions will be more inclined to repairing and acquiring refurbished products than buying brand-new ones. Expanding the market for repair services has a beneficial effect on the environment [7] in addition to having favorable economic effects for repair companies [8]. This is backed by current studies regarding the shortcomings of the current state of repair services and the suggestions to improve those with different measures, including up-scaling, better networking, usage of innovative technologies and governmental support [7]. However, additional thresholds are caused by user behavior. Laitala et al. are stating, that the willingness of the repair is highly dependent on the cost of a product, environmental concerns, initial cost and age of the product to name a few [18] Traditional repairs frequently cause uncertainty for consumers, raising concerns about reliability and efficiency. The proposed solution ensures repairability, reduces resource waste and saves time by providing a refurbished product upfront.

B. *Technical State of the Art*

Repair research is growing in popularity [5], thereby playing an important part in limiting environmental effects on the planet [9] via waste resource reduction. According to the findings of Sonego et al., [10], the literature review underscores that consumers encounter various barriers refraining from engaging in repair activities and the need to reduce them to achieve a circular economy. Highlighting these consumer-related obstacles goes in the same direction as emphasizing product lifespan extension through repair instead of discarding. Moreover, McLaren et al. [11] critically review circular economy literature and introduce four categorizations of repair: reconstruction/restoration, remediation, reconciliation, and reconfiguration [5]. This ongoing experiment could be categorized as reconstruction/restoration because the aim is to restore the product from its damaged state to its original function and purpose using standard materials to achieve an authentic refurbished product. Moreover, repair here might also be viewed as remediation, as long as the original purpose of the product is maintained and functionality is prioritized.

In order to derive a continuous function of the state of the battery based on image data or the expected expenditure for repair, we are dealing with a regression problem similar to that of human age or pose estimation [15]. For this task's general purpose, Convolutional Neural Networks (CNNs) like ResNet or VGG have proven to be applicable [15]. This area of research has made major progress on the topic of image classification over the last couple of years. For example, EfficientNetV2M (2021) [16] achieved an increase of 10 percentage points in accuracy compared to ResNet-50 (2015) with other recent models performing even better [17].

III. Project Setup

As mentioned in the Introduction, the project has two main goals: on the one hand, the first phase of the experiment contains the development of a structure that enables 3R operations and product-centered data recording and sets the foundation for the second phase of the experiment, the development of the AI-based expenditure assessment. This section includes a description of the first part of the experimental setup regarding the refurbishment system and the data recording and storage. The product chosen for the experimental conductivity of the system is a power tool battery, which is a widely used battery power system for a variety of different power tools. Therefore, clients can trade a refurbished accumulator, and in exchange, they send in their defective batteries of the same kind in order to stock them after restoring the functional condition of the product.

The described system, as shown in Figure 1, has two main **actors**, which interact with the system: On the demand side, we have our **client**, whose desire is to receive a functioning product. On the other side, we have our **operator**, who is responsible for the conduction of the recording and restoring operations in our system. The **first step** in our process flow contains the acceptance by the client of the previously generated offer from the operator. With the acceptance of the offer, the client agrees to pay a certain amount of money for the refurbishment service, while at the same time, the operator agrees to conduct the concluding operations. This leads to the **second step**, in which the operator sends the refurbished item together with a *Product Condition Survey (PCS)* to the client before he receives the defective product. The client therefore does not receive his battery in a restored condition back. He rather gets a product of the same kind, which is restored to a functional condition. In the **third step**, the client sends his defective battery with the filled-out *PCS* to the operator. The *PCS* contains questions regarding the type of defect and additional product life information and gives, on one side, the operator a short pre-assessment of the product's condition, on the other side it already gives a good insight into the kind of information a client is able to give the operator based on his own assessment abilities. Followed by the **fourth step**, the product is externally recorded by an *Automated Image Capturing Toolchain (AICT)*, which provides the foundation for the later aspired AI image assessment system. The *AICT* takes images from different angles of the product as well as an image of the product label to support the later-designed system with Optical Character Recognition (OCR). This enables the system, designed in the second phase, to receive additional information for certain products with an attached product label. In the case of our before-mentioned battery, this is for example the manufacturing location and date along with product line specifications. The **fifth step** contains the actual refurbishing operation, the conduction of additional product measurements,

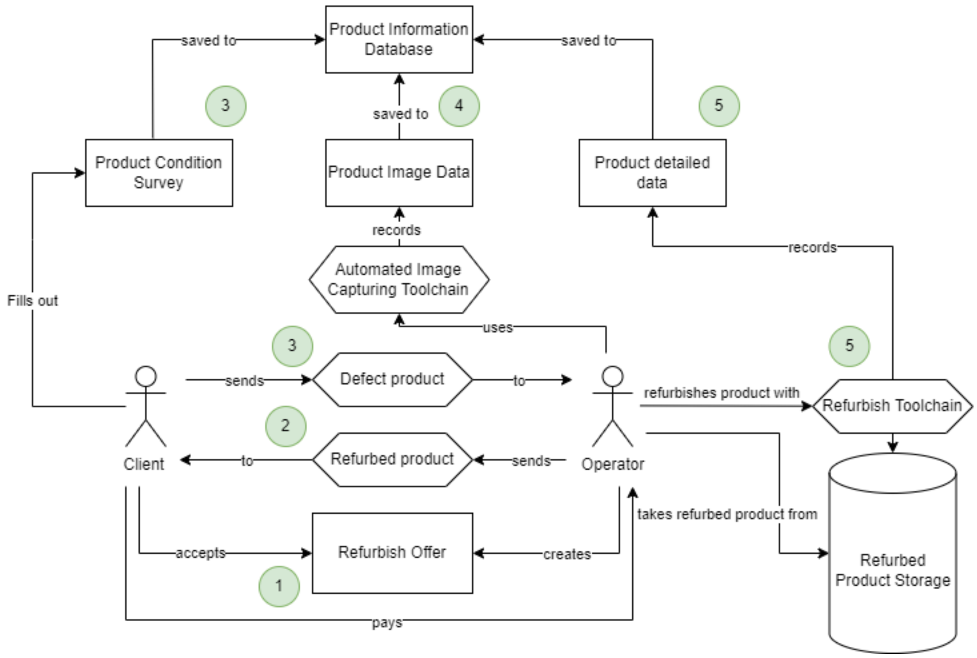

Fig. 1. Refurbishing structure for process conduction and data recording

and the storage of the refurbished product in the *Refurbished Product Storage*. The additional information about the product is in the case of the accumulator electrical data, for example, cell capacity, voltage, and current, as well as other repair data such as repair time and spare parts required. The experimental setup is highly modular and can therefore easily be retrofitted to suit other product type needs. Apart from that, the systems' modules can be uncoupled from their place in the system to be reassigned into another frame.

IV. Preliminary results

The preliminary results so far give a good overview of the applicability of both the overall concept as well as the results, which can be expected from the image collection unit. However, in order to determine the impact of such a system, further systemics regarding the logistic system as well as the financial backup regarding repair operations have to be evaluated and taken into account to refine the concept. On a technological level, the evaluation of suitable models and the training of the AI will be the next step of the project conduction. For this reason, images of the batteries were taken in a semi-automated way with the help of cooperative robots. In addition, the *PCS* sent to the user surveyed the estimated condition and usage scenario, among other information. In order to train for the cost of refurbishing operations, defects, and condition, the operator recorded the electrical information of the defective batteries as well as the resources required for restoration.

A. Image recording

To record the images for the future planned second phase of the experiment, the *AICT*, consisting of two co-bots of the same kind, is used to place the battery in a specific orientation. In the beginning, the battery is placed on a conveyor belt so that it can be transported automatically to the robot. The image-capturing device consists of two stereo cameras that enable the system to take pictures from a top-down and a 45° slanted perspective from the side of the belt. The first robot turns the battery 90 degrees around while the conveyor belt transports it back to the camera location where the cameras take an image of the battery. The process repeats after the battery is back in the *original* position. The other one grasps the battery from the side and places it facing the camera *above* the conveyor belt. This camera is used because the information on the product label can be used for OCR-based information extraction. At the end, the robot returns to the *handover* position and turns the battery 180° around on the top side. In total, 17 different views of each of the batteries are taken via both cameras: eight from each side where the battery is on top, eight where it is turned around, and one with a close up view of the battery label. To make sure that pictures are taken at the same position and angle, a script is written that takes the images automatically and switches between the two camera ports. Besides that, the use of co-bots gives the advantage that the images for all batteries are taken in the exact same position.

B. Battery Status

Preliminary results show that there are two main reasons for battery failure: Deep discharge with batteries in good condition or degraded cells with low remaining capacity, sometimes combined with a deep discharge. The latter is more often the case with 90% versus 10%. Interestingly, it is quite easy to detect deeply discharged cells as a user, however, it remains

 ISBN: 978-1-68558-153-4

to be seen, whether the condition can be gauged as well. This would be highly beneficial, since refurbishing an accumulator with deeply discharged cells in good condition requires no repairing material and little time. We noticed that the quality of the battery depends on the type and the manufacturer of the cells. This is not apparent from the outside, but could be derived from the production year and number as well as design changes like the choice of colors and logos over different periods of time, like displayed in Figure 2.

Fig. 2. Captured batteries of two different production batches

C. Future Research Agenda

Adjunct to the first phase of the experiment, which is currently running, the future phase will be concentrated on the usage of the acquired product data, specifically training a machine learning model to estimate the expenditure for refurbishing a specific product based on the product type and its condition. To conduct this part of the experiment, the batteries are labeled and indexed to connect the specific records (images, data readings, refurbish time, and monetary cost) to a specific battery. Furthermore, data from user *PCS* can be used to derive the state of the battery and the refurbishment cost. For this, a regression is applied to make predictions. We believe Deep Neural Networks (DNNs) regression to be better suited for this kind of prediction over regression analysis since we are dealing with noisy, unreliable data provided by the user. Further, we have many dimensions with no clear correlation apparent and lastly, we want to have multiple outputs to predict both the state of the battery as well as refurbishment cost [13]. Finally, we want to use mixed data for the neural net to combine both the image data as well as data from the *PCS*. The models' input consists both of numeric and categorical values from the *PCS*, as well as image data of the battery to be assessed. In addition, data from the product's label is captured via OCR and used as additional input parameters. This has previously been done in the context of housing price prediction [14].

It is therefore the goal of the research to find the correlation between the external product features, like geometric or color-based anomalies, and the internal state of the product and investigate if these can be coherently assigned to one another. The outcome will show if causality can be drawn between the different parameters, which would also open up the research for other product groups. Based on those findings, guidelines and regulations can be formed, which in turn could fuel product lifecycle prolonging even further. The regulations could in turn enable those services on a wider scale and increase the economic feasibility of such operations. Further results might be able to outline specific parameters of the products in order to derive specifications for lifecycle prolonging operations.

V. Conclusion

The paper showed how the experiment can pave the way to increase service generation in the area of the 3Rs and how it can help with the design of digitized support systems. The current stage of the experiment allowed a first review of a potential instance of the service system for power tool accumulators and the review of the so far acquired product recordings to plan and design the adjunct AI-based systems for the second project phase. However, the full potential and efficiency of such a service systems relies heavily on the kind of product, the information density of the product generated by the stakeholders, and the functionalities of the designed solution. The second phase of the experiment will further show how AI-based systems could be implemented in the service system and how such systems can help to assess and prepare the clients' products for the concluding steps. Future research in this area could be continued by applying the service system to other product branches while investigating the business models that can fuel a broad application to accelerate sustainable business fields.

Acknowledgment

This work was funded by the Federal State of Lower Saxony, Germany.

References

[1] J. Grafström and S. Aasma, "Breaking circular economy barriers, Journal of Cleaner Production", vol. 292, 2021.
[2] J. Kirchherr, L. Piscicelli, R. Bour, E. Kostense-Smit, J. Muller, A. Huibrechtse-Truijens, and M. Hekkert, "Barriers to the Circular Economy: Evidence From the European Union (EU)", Ecological Economics, vol. 150, p. 264-272, 2018.
[3] C. Hagelüken, M. Schmidt, L. Schebek, and C. Liedtke, "Opportunities and limits of recycling in the context of the circular economy: Framework conditions, requirements, and recommendations for action, position of the resource commison at the german environment agency (kru)", June 2023.
[4] G. D. A. Galvão, J. de Nadae, D. H. Clemente, G. Chinen, and M. M. de Carvalho, "Circular Economy: Overview of Barriers", Procedia CIRP, vol. 73, p. 79-85, 2018.
[5] M.Van Der Velden, E. Maitre-Ekern, and D. K. Wanja, The Role of Independent Repair in a Circular and Regenerative Economy. Circular Economy and Sustainability, pp.1-26, 2023.
[6] S. Svensson-Hoglund, J. D. Russell, and J. L. Richter, "A process approach to product repair from the perspective of the individual", Circular Economy and Sustainability, 3(3), pp.1327-1359, 2023.

[7] I. Güsser-Fachbach, G. Lechner, T. B. Ramos, M. Reimann, ”Repair service convenience in a circular economy: The perspective of customers and repair companies”, Journal of Cleaner Production, vol. 415, 2023.
[8] M. D. Bovea, V. Ibáñez-Forés, and V. Pérez-Belis, ”Repair vs. replacement: Selection of the best end-of-life scenario for small household electric and electronic equipment based on life cycle assessment”, Journal of environmental management, vol. 254, p.109679, 2020.
[9] W. Steffen, K. Richardson, J. Rockström, S.E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. De Vries, C. A. De Wit, and C. Folke, ”Planetary boundaries: Guiding human development on a changing planet”, Science, 347(6223), p.1259855, 2015.
[10] M. Sonego, M. E. S. Echeveste, H. G. and Debarba, ”Repair of electronic products: Consumer practices and institutional initiatives”, Sustainable Production and Consumption, 30, pp.556-565, 2022.
[11] D. McLaren, J. Niskanen, J. and Anshelm, ”Reconfiguring repair: Contested politics and values of repair challenge instrumental discourses found in circular economies literature”, Resources, Conservation ‘I&’ Recycling: X, vol. 8, p.100046, 2020.
[12] M. Young, ”The Technical Writer’s Handbook”, Mill Valley, CA: University Science, 1989.
[13] M. Imaizumi, and K. Fukumizu, ”Advantage of Deep Neural Networks for Estimating Functions with Singularity on Hypersurfaces”, Journal of Machine Learning Research, vol. 23, p. 1–53, 2020.
[14] E. H. Ahmed, and M. N. Moustafa, ”House price estimation from visual and textual features”, Computer Vision and Pattern Recognition, 2016.
[15] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, ”A Comprehensive Analysis of Deep Regression”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, issue 9, p. 2065–2081, 2018.
[16] M. Tan, Q. v. and Le, ”EfficientNetV2: Smaller Models and Faster Training”, Proceedings of Machine Learning Research, vol. 139, p. 10096–10106, 2021.
[17] X. Yu, Y. Xue, L. Zhang, L. Wang, T. Liu, and D. Zhu, ”NoisyNN: Exploring the Influence of Information Entropy Change in Learning Systems”, 2023.
[18] K. Laitala, I. G. Klepp, V. Haugrønning, H. Throne-Holst, P. Strandbakken, ”Increasing repair of household appliances, mobile phones and clothing: Experiences from consumers and the repair industry”, Journal of Cleaner Production, vol. 282, 2021.

CollectByCycle: Towards an Automatized Condition Assessment for Bicycles

Tobias Geger, Dominique Briechle, Marit Briechle-Mathiszig, Nelly Nyeck, Robert Werner
Institute for Software and Systems Engineering
University of Technology Clausthal
Clausthal-Zellerfeld, Germany
email: thomas.tobias.marcello.geger@tu-clausthal.de; dominique.fabio.briechle@tu-clausthal.de;
marit.elke.anke.mathiszig@tu-clausthal.de; nelly.nicaise.nyeck.mbialeu@tu-clausthal.de;
robert.werner@tu-clausthal.de

***Abstract*—The detection of defective parts of complex products is a great challenge for the untrained eye and impedes the assessment of different conditions. Therefore, the necessity for smart solutions that bridge the gap between user awareness of a product state and the restoration service of an operator is at an all-time high. The goal of this paper is to outline the preliminary results regarding the generation of a data set in order to improve bicycle repair and reduce the verification time of the component condition assessment. To enable such a service, a total of 115 bicycles were collected and classified.**

***Index Terms*—Repairing, Data set, Labeling, Product breakdown Structure, Circular Economy**

I. Introduction

The earths' resources are scarce, and therefore the overall depletion capacity is restricted. This is demonstrated by the Earth Overshoot Day. This day represents the date from which on humanity has exceeded the resource depletion in comparison to the natural budget of the year. This date was in 2023 the 2$^{\text{nd}}$ of August [1]. For the reduction of resource consumption, it is important to avoid the production of new products. This would help to save resources. Repairing offers an opportunity to avoid the production of new products and extend the life span of products that are already in use. However, at the beginning of the repair process, the question arises of which components are exactly in need of repair in order to restore the products' functioning condition. Therefore, the aim of this work is to investigate, if it is possible to identify certain conditions of product components and enhance the process of defect detection by reducing the amount of time necessary. For this purpose, condition data and images of the retrieved bicycles were taken and labeled and an evaluation concept, for the later used Artificial Intelligence (AI)-based solution was developed. The first section of this work contains the Introduction. Section 2 includes the Domain Background and the Technological State of the Art. The following section describes the initial Project Setup and section 4 deals with Preliminary Results. In section 5, the overall Conclusion and Discussion is drawn with an outlook of the subsequent steps of the project.

II. Background

The development of support tools in the sustainability domain for various tasks is a highly anticipated field in nowadays computer science and the IT-Industry [2]. Sustainability demands and necessities accelerate the development of these tools and lead to scientific challenges to enhance the information significance of the designed solution. In order to design solutions for domains like repair, not only the technological foundation must be assessed and evaluated but the domain-wise governing factors as well. Therefore, the *Background* section is divided into the *Domain Background* and the *Technological State of the Art*.

A. Domain Background

Product lifecycle-prolonging actions are one of the core mechanisms enabling the transformation of our linear economy. Short product usage cycles lead to higher amounts of waste that need to be processed, as well as a higher amount of energy consumption due to the production of new goods to replace the old or defective ones [3]. Therefore, the need for tools to support lifecycle prolonging measures becomes more and more important. Especially in the field of repairing, the hurdles limiting repairers are often, besides the economic barriers, the information gap regarding the specific defect, which causes the malfunction of the product [8]. Digitized tools can hereby help overcome these hurdles by enhancing the information density of a specific product, which reduces repairing time and supply chain planning for the repairer [9].

B. Technological State of the Art

The usage of Artificial Intelligence for support in analysing data is nowadays widely state of the art [4]. In the field of optical defect detection the usage of machine learning, especially Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs), is already used and showed promising results in many applications. Papageorgiou et al. line up in a short overview of the application of Computer Vision for defect detection in manufacturing, which technologies are the most common, and discuss the impact of the different machine learning technologies for certain applications. Hereby, they describe 2D as well as 3D based detection technologies

 ISBN: 978-1-68558-153-4

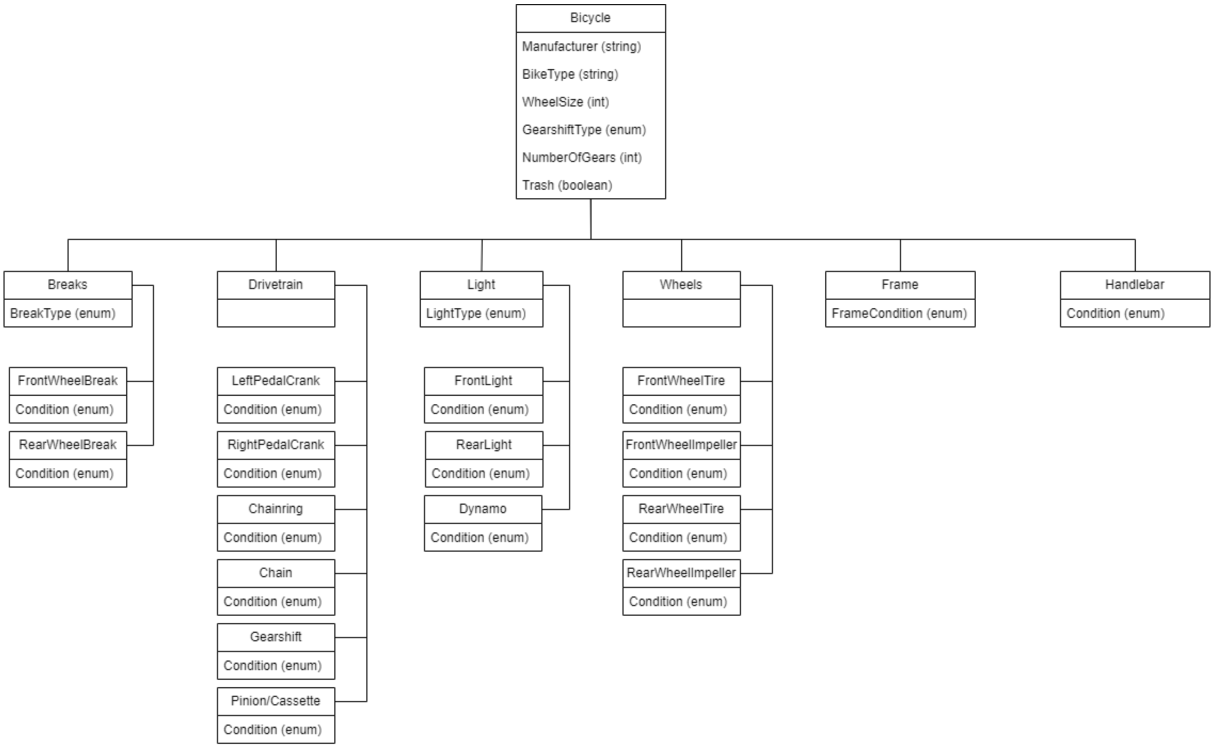

Fig. 1. PBS for the Bike data acquisition

[6]. Especially in the area of detecting product production issues on the external surface, the usage of CNNs is already an accepted technology. Hereby, the usage of pattern recognition showed how surface failures can be detected to enhance quality control [5]. A broader application-driven view is shown by Saberironaghi et al., where the scope includes also different types of defects products can show based on their product lifecycle. Therefore, the application of deep learning for different kinds of data was shown in order to classify defects in products based on their specific fault pattern [7]. The mentioned references showed hereby three things: deep learning is a possible solution for defect identification, the necessity of a strategy to tackle small sample data sets, and the need to label those data in a structured way in order to gain accurate data out of the machine learning model. As a result, our experiment is dissected into different phases, which are described in the following section.

III. Project Setup

The conducted experiment aims to provide the data foundation for later targeted an AI-based defect identification demonstrator that is not only able to classify bicycles but further to detect missing or damaged parts, which affect the overall functionality of the bicycle in mechanical as well as legal terms. Therefore, the technical relevant parts of the bicycle have been classified in the form of a *Product Breakdown Structure (PBS)*, which sets the scope for the different product phases and is used for structuring our product labeling Figure 1. [10]. The PBS therefore defines the layout for the collection phase as well as the adjunct phases of the data recording described in the next subsection.

A. Project phases

The project is divided into four different phases. The first phase consisted of the collection of the bicycles. This was maintained with the help of a designed collection app, where users could register their old and/or superfluous bicycles. The bicycles were subsequently collected and stored, which led to the next phase. The bicycles were then inspected with the help of the designed questionnaire, which also set the foundation for our PBS. Afterward, images were taken from different perspectives of the bicycles and their components to generate a sufficient data set for each of the bicycles. This data set will be used as foundation in the following phase to train an AI-demonstrator, which detects missing components of bicycles.

B. Appraisal

The appraisal phase of the project contained the recording of the images to capture the features of the bicycle. Additionally, the bicycles were classified with the checklist containing the PBS shown in (Fig. 1.). Therefore, the image-capturing process consists of a total of six different images taken for each of the bicycles. The bicycles were photographed from both sides, as well as the front and rear, in order to capture the relevant components. The different views capture the components of the bicycle as well as the bicycle itself to later provide a sufficient sample data set to train the aspired model. In addition, two images were taken which show a more detailed view of the front tire as well as the back tire. This enables a more detailed view of critical components, like the brake system and the drive train.

For further information about the bicycles, which may not be visible, we developed the before-mentioned checklist, which will additionally be used to label the images taken above. In the first step, the checklist was developed with the goal of getting all the necessary information about the bicycle. This development was based on literature research and feedback from repairers. In the end, a useful checklist for scientific research purposes and for repair was created. This checklist yields to the already mentioned PBS, which is shown in (Fig. 1.). At first, general information about the bicycle was stored in different data types in accordance with the possible states of the components. The components were in advance classified. For example, the *condition* state can have three different states {**functional, defective, missing**} to determine the necessary repairs to reinstate the functional condition.

With the images and the results of the checklist, we can build the data set by linking specific conditions and bicycle types to the images in order to train the AI. We have therefore planned to use the PBS as the foundation for the labeling of the images.

IV. Preliminary results

The current state of the experiment already allows to draw of a few conclusions. The preliminary results can therefore be divided into a section concerning the overall derivation of certain parameters already visible in the collected data, and a section concerning the contextualized view regarding the later-developed machine learning system and the potential significance of the former. In the case of the images and their classification, we were able to link different states directly to the proposed PBS in order to determine the conditions of the bicycles' different components. Because of this, it was possible to separate the collected bicycles already in their responsive categories, which enables the operators to further process the bicycles.

A. Data set Overview

In order to design efficient assessment support for the operator, a sufficient data set is needed that fulfills the requirements regarding the categorization of the repairer. For this reason, a total of 115 bicycles were recorded and, as mentioned before, classified according to their type. Additionally, the data set was further enriched with the detailed data of the PBS, namely the condition of the different components the bicycle is composed of. For the recording process, the use of a twelve megapixel camera proved to be sufficient in terms of the overall resolution. The images provide ample resolution to identify the different components and potential damages. However, it needs to be seen if this is sufficient in order to receive an accurate assessment of the product state in terms of part damages and missing components. An example of an image taken from one side is visible in Figure 2. where we can see the labeling of both the tires of one of the bicycles. In this case, the front and rear tires are labeled and marked as a colored polygon. Also, the front and light, as well as the left pedal crank are edged and can be seen in the image. The bicycles that were collected are already mapping a great variance of the potential conditions and are therefore suitable to provide the variety needed for the proper training of the Artificial Intelligence. Additionally, a few special types were collected as well, for example, e-bikes, which require particular treatment based on their more complex product composition.

As already mentioned, the decision if a bicycle can be reused depends on different influencing factors. There are, however, some critical components regarding the mechanical functionality and critical components linked to legal requirements, like the German road traffic regulations (StVGO), that highly affects this decision. One example of such a part is the frame because it is not exchangeable. The frame of some bicycles is in very bad condition and could possibly be detected and classified with the according label with the help of a trained AI.

If we say ”bad condition”, we mean the frame may have corrosive spots, bulges, or superficial issues like numerous stickers that do not affect the functionality of the bicycle. However, this will be further divided since one affects the overall functionality deeply whereas the other solely is a cosmetics issue.

As in comparison, there are other components where the condition may or may not be visually apparent. As an example, it is fairly easy to tell if a tire is flat and therefore defective. However, a defect light caused by a damaged portion of the component is not always visible.

Fig. 2. Recorded bicycle with highlighted (human conducted) labeling for both tires.

 ISBN: 978-1-68558-153-4

B. Domain context

The collected bicycles showed a huge variety in terms of the collected models and types, as well as their condition state. The total of 115 bicycles we received over the course of the collection period over one month was more than the authors initially anticipated over the given time frame. The project therefore already showed, that there is a certain willingness in the public to support sustainable causes. However, the authors did not conduct any economical or social science related research to investigate this further. For the broader application of such systems, additional influencing factors, especially economical ones like spare part prices, labor cost and transportation costs must be evaluated in depth in order to provide a substantial assessment of the disability of such models. This project therefore serves as a pilot for small sample product collection to examine the potential of this kind of experiment, where citizens can give their old and superfluous products away. After that, the products will be repaired and then returned to the usage cycle. Therefore, resources can be spared and production capacities can be optimized further. Still, the question remains, if such processes are feasible and can provide a suitable business model for product restoration companies.

V. CONCLUSION AND DISCUSSION

The paper described the collection of an image data set of bicycles in order to use them later on to train deep learning models. Therefore, an emphasis was put on the recording of data and the structure of the component labels. With the help of the PBS, the data can be labeled according to the suggested data model, which allows a component-wise classification of the parts which will therefore allow to assign the labels in a harmonized way. The project data set can therefore be used as well for a variety of classification or identification tasks and can bolster computer vision systems in the domain of repairing, refurbishing, and remanufacturing. The task the authors of the paper will try to investigate in the next step of the paper is the detection of missing components of a bicycle based on the bicycle's type, as well as a condition assessment of the quality of the different parts. If this step proves successful, the authors aim to design a system that already generates concrete repair instructions based on the current status of the bicycle, which could help the repairer further by already managing the supply chain necessary for repair. However, the overall mechanisms of the aspired system must be investigated further to clarify the applicability of the trained model, as well as the overlaying business model. For the Artificial Intelligence, the question remains, how additional data input could be used to bolster the accuracy of the model. A possible approach could be the use of a multimodal deep learning model, which benefits from multiple data inputs, for example, pre-assessments generated by the bicycle owners. Therefore, potential non visual factors could be accounted in a more detailed manner, which could led to other results in the overall assessment of the condition of the products. The success of the overall system is however depending as well on the accuracy of the AI component as on the proper conduction of the repairing task. To tackle both personal and economic shortcomings for the actual repair operation, a possible solution would be the attachment of a expert consulting system, which generates repair instructions based on the condition of the product. This could enable on the one hand enable repairing for laypeople, which do not possess the knowledge for bike repairing. On the other, it could benefit the repairers with additional knowledge and information to enable a faster repair conduction by providing services like the pre-selection spare parts. The data set therefore provides an interesting opportunity to investigate, how different types of data can improve services for Circular Economy applications.

ACKNOWLEDGMENT

This work was funded by the Federal State of Lower Saxony, Germany.

REFERENCES

[1] "Earth Overshoot Day 2023 fell on august 2", https://overshoot.footprintnetwork.org/ , status: 01.02.2024.
[2] A. T. Rosário, J. C. Dias, "Sustainability and the Digital Transition: A Literature Review" Sustainability, vol. 14, no. 7, p. 4072, 2022
[3] K. Laitala, I. G. Klepp, V. Haugrønning, H. Throne-Holst, P. Strandbakken, "Increasing repair of household appliances, mobile phones and clothing: Experiences from consumers and the repair industry", Journal of Cleaner Production, vol. 282, 2021.
[4] S. Athmaja, M. Hanumanthappa and V. Kavitha, "A survey of machine learning algorithms for big data analytics", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, p. 1-4, 2017.
[5] P. C. Lien and Q. Zhao, "Product Surface Defect Detection Based on Deep Learning", 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, p. 250-255, 2018.
[6] E. I. Papageorgiou et al., "Short Survey of Artificial Intelligent Technologies for Defect Detection in Manufacturing", 12th International Conference on Information, Intelligence, Systems 'I&' Applications (IISA), Chania Crete, Greece, p. 1-7, 2021.
[7] A. Saberironaghi, J. Ren , M. El-Gindy, "Defect Detection Methods for Industrial Products Using Deep Learning Techniques: A Review", Algorithms. vol. 16, no. 2, p. 95, 2023.
[8] M. Jaeger-Erben, V. Frick, T. Hipp, "Why do users (not) repair their devices? A study of the predictors of repair practices", Journal of Cleaner Production, vol. 286, 2021.
[9] N. Roskladka, G. Bressanelli, G. Miragliotta, N. Saccani, "A Review on Design for Repair Practices and Product Information Management", In: E. Alfnes, A. Romsdal, J. O. Strandhagen, G. von Cieminski, D. Romero,(eds) "Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures", APMS 2023, IFIP Advances in Information and Communication Technology, vol. 692, Springer, Cham.
[10] E. Brusa, "Digital Twin: Toward the Integration Between System Design and RAMS Assessment Through the Model-Based Systems Engineering," in IEEE Systems Journal, vol. 15, no. 3, p. 3549-3560, Sept. 2021.
.

A Method for the Runtime Validation of AI-based Environment Perception in Automated Driving Systems*

Iqra Aslam*, Abhishek Buragohain†, Daniel Bamal‡, Adina Aniculaesei§, Meng Zhang¶, and Andreas Rausch‖
Institute for Software and Systems Engineering, Technische Universität Clausthal
Clausthal-Zellerfeld, Germany
Email: { iqra.aslam*, abhishek.buragohain†, daniel.bamal‡, adina.aniculaesei§, meng.zhang¶, andreas.rausch‖ } @tu-clausthal.de

***Abstract*—Environment perception is a fundamental part of the dynamic driving task executed by Autonomous Driving Systems (ADS). Artificial Intelligence (AI)-based approaches have prevailed over classical techniques for realizing the environment perception. Current safety-relevant standards for automotive systems, International Organization for Standardization (ISO) 26262 and ISO 21448, assume the existence of comprehensive requirements specifications. These specifications serve as the basis on which the functionality of an automotive system can be rigorously tested and checked for compliance with safety regulations. However, AI-based perception systems do not have complete requirements specification. Instead, large datasets are used to train AI-based perception systems. This paper presents a function monitor for the functional runtime monitoring of a two-folded AI-based environment perception for ADS, based respectively on camera and LiDAR sensors. To evaluate the applicability of the function monitor, we conduct a qualitative scenario-based evaluation in a controlled laboratory environment using a model car. The evaluation results then are discussed to provide insights into the monitor's performance and its suitability for real-world applications.**

Keywords-runtime monitoring; function monitor; dependable safety-critical system; automated driving system; perception system.

I. Introduction

In principle, fully autonomous vehicles are technically feasible. However, after the initial proof-of-concept testing under ideal conditions, e.g., in lab environments [1], or on restricted test fields [2], further innovative verification and validation techniques are needed during the approval and release processes. These additional verification and validation phases are necessary to gather the required evidence for the safety and reliability of the autonomous vehicle in real-world scenarios. For the commercial approval of autonomous vehicles by certification bodies, the current state-of-the-art practices require verifying specific maneuvers using predefined test scenarios and statistically analyzing real-time data covering millions of kilometers of driving. For autonomous vehicles at Society of Automotive Engineers (SAE) Level 3/4 and above, the key challenge lies in ensuring the safe commercial release and the safe vehicle operation in all possible situations, not just only in those situations encountered during the system development, e.g., through random tests.

Today's automated driving systems are primarily designed to be fail-safe systems, capable of switching the ego-vehicle to a safe state, e.g., by activating an emergency brake. However, future ADSs must be designed as fail-operational systems, especially as there are many situations in which an immediate fail-state might not be readily accessible, e.g., when driving at high speed on the highway. Moreover, in case issues occur during the vehicle operation, the control over the dynamic driving task can no longer be simply handed back to the human driver, since human intervention is not mandated anymore at SAE L4+ [3]. Without a human fallback system, the ADS must be able to take over control and establish a safe state for the vehicle, if a problematic situation arises.

In recent years, a high-level functional architecture has been established for ADSs comprising three main subsystems: (1) environment and self-perception, (2) situation comprehension and action planning, and (3) trajectory planning and motion control [4], [5]. The environment and self-perception is particularly significant as it strongly impacts the performance of the entire ADS and the safety of the autonomous vehicle, as shown by a series of accidents involving (partially) automated driving functions, e.g., Tesla's autopilot. In the first notable accident in 2016, the Tesla's autopilot has failed to detect an articulated lorry driving in the opposite direction which was engaged in a turn maneuver, despite having been successfully tested over 200 million kilometers. In the context of a brightly lit sky, both the driver and the autopilot were unable to recognize the lorry, which had white sides [6]. In response to this accident, Tesla announced the introduction of Shadow Mode to enhance the safety of its autopilot [7]. An approach for continuous monitoring of autonomous driving functions for development, validation and series operation has been proposed in in [8] and demonstrated for the lane changing functionality of the highway pilot in [9]. This approach essentially extends the concept of shadow mode, by addressing two questions: (1) does the autonomous driving function operate correctly (qualitative oracle) and (2) is the autonomous driving function currently operating in a known environment (quantitative oracle) [9].

There is a noticeable gap in research work regarding the validation of environment and self-perception methods applied in automated driving applications. AI-based approaches have prevailed over classical approaches for realizing environment perception, as the former are used both in image processing and in signal processing of other raw sensor data, such as radar. Current safety-relevant standards for automotive systems, e.g., ISO 26262 [10] and ISO 21448 [11], assume the existence of

 ISBN: 978-1-68558-153-4

complete requirements specifications. The system development process is usually organized using a structured process model, e.g., the V-model. However, challenges arise with AI-based perception systems since they do not have complete requirements specifications. Instead, the development usually begins with incomplete artifacts, e.g., system requirements formulated for the entire ADS structured test cases derived from other artifacts, e.g., in the format of OpenSCENARIO [12] or OpenDRIVE [13] formats. Additionally, the development of AI-based systems requires extensive training. For instance, for the development of an AI-based system for pedestrian detection an substantial training data set consisting of diverese pedestrian images is required.

The approach outlined in [9] by Mauritz and his co-authors can be understood as developing a dependability cage for monitoring the lane changing functionality of the highway pilot. While their focus is monitoring the entire autonomous driving function without particularly considering the environment perception, in this work we focus on monitoring the environment and self-perception subsystem of an ADS. We propose a dependability cage for validating an environment and self-perception system including redundant perception components that operate with multiple sensor data sources. In this dependability cage, an essential component is a function monitor which evaluates the consistency of the outputs produced by the redundant perception components during the ADS operation. This function monitor is derived from a high-level safety requirement established for the environment and self-perception subsystem. In this work, we evaluate the dependability cage approach for AI-based environment perception qualitatively in a lab environment using predefined test scenarios.

The rest of the paper is structured as follows. Section II gives an overview of related work in the area of verification and validation of AI-based system, with a specific focus on environment perception systems in both robotic and automotive applications. Section III presents the dependability cage approach for the monitoring and validating AI-based environment perception during the ADS operation. In Section IV, we conduct a a qualitative scenario-based evaluation and discuss the obtained results. Section V concludes the paper by summarinzing its contributions and outlining potential future work directions.

II. RELATED WORK

Environment- and self-perception is an integral part of the dynamic driving task that is carried out by the ADS. Primarily based on AI models, it provides essential inputs for further safety-critical functions of the ADS, such as Situation Comprehension and Trajectory Planing. Through the interpretation of various raw sensor data into detailed semantic information about the surrounding driving environment, the AI-based environment perception subsystem enables the ADS to complete its decision making, motion planning and control command execution, by relying on its respective planning and decision components.

For this reason, monitoring and evaluating the functional behavior and the performance of AI-based environment perception systems at runtime is extremely important for the safety of the autonomous vehicle. Recently, diverse research approaches have addressed the safety issues of AI-based environment perception. Czarnecki [14] identifies a set of influencing factors for AI-based environment perception: (1) conceptual uncertainty, (2) development situation and scenario coverage, (3) situation or scenario uncertainty, (4) sensor properties, (5) labeling uncertainty, (6) model uncertainty, and (7) operational domain uncertainty. Identifying these factors is understood as the first step, which should be followed by a systematic analysis of their impact on the perceptual uncertainty. In addition, methods to eliminate or reduce their negative effects on the perceptual uncertainty. Subsequently, mitigation measures are proposed to be applied for the cases when the control of the negative effects is not possible [14]. The concept in [14] revolves around using these methods to gather the essential evidence to substantiate claims regarding the environment perception uncertainty in safety cases that contribute to demonstrating the safety of the overall autonomous vehicle [14].

Another survey presented in [15] identifies various research directions concerning the runtime performance monitoring of AI-based perception in autonomous robots. One direction encompasses approaches using past examples of failures and successes or similarity of operational context to previous experiences to predict the quality of perception output [15]. The second direction involves by methods that detect inconsistencies in perception output, using the input data from a single sensor or from multiple sensors, or by comparing outputs from different models [15]. The third direction focuses on methods for uncertainty estimation by indicating low confidence in prediction output, calculating confidence scores as a measure of the target model's output quality, and detecting anomalies [15]. Several other studies also aim to validate the accuracy or robustness of the perception system outputs concerning specific inputs of the neural network. However, the approaches surveyed in [16] and [17] are primarily limited to offline verification.

In addition to estimating uncertainty, some recent studies have shifted their primary focus to object detection as the main task of the perception system under analysis. Thus, Feng et al. [18] evaluate LiDAR-based object detectors using the Jaccard Intersection over Union (IoU) metric and KITTI and Waymo datasets, incorporating label uncertainty specifically for the bounding boxes of a particular object class, e.g., the object class *car*. In [19], the authors propose a framework to predict performance monitoring of object detection at runtime without relying on any ground-truth data. Meanwhile, in [20], they monitor the performance of the object detector deployed on mobile robots by predicting the quality of its mean average precision using a sliding window technique over the input frames. However, the approach presented in [21] for monitoring of neural networks that estimate 3D human shapes and poses from images is limited to human pose estimation.

 ISBN: 978-1-68558-153-4

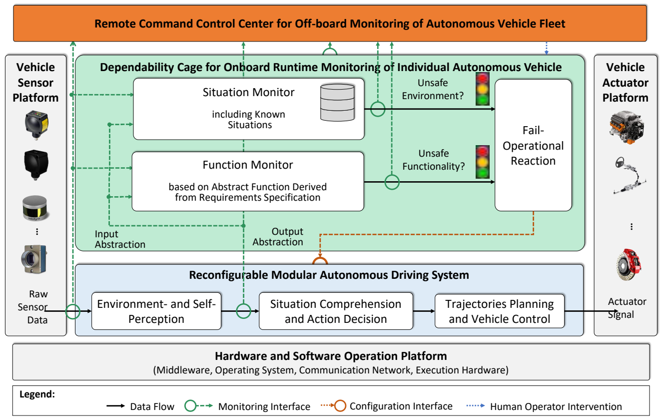

Figure 1: High-level Architecture of the Dependability Cage for Runtime Validation AI-based Environment Perception Systems.

In previous research, the majority of studies have made significant contributions towards the evaluation methods, applications and understanding of uncertainty in perception systems. Nevertheless, most of these studies focus on a single object detection network. They do not address the comprehensive validation of perception system outputs concerning reference sensors or the comparative analysis of outputs from redundant perception systems. Additionally, there is a research gap in detecting perception failures or incorrect behaviors of environment perception systems at runtime to ensure the safety of the perception-equipped system. In this work, we propose an approach for the runtime validation of environment perception in autonomous vehicles (AVs), by analysing and comparing the outputs of two redundant perception systems utilizing camera and respectively LiDAR sensors data. We evaluate this approach qualitatively using predefined scenarios and a model car in a lab environment.

III. Integrated Safety Architecture with Dependability Cage for the Runtime Validation of AI-based Environment Perception in Automated Driving Systems

This section introduces the Dependability Cage approach for the runtime monitoring and validation of AI-based environment perception in automated driving systems. Figure 1 illustrates the high-level architecture of this dependability cage, that can be understood as an instantiation of the overarching concept of the Dependability Cage approach, presented first in [22]. The subsequent sections offer a detailed introduction to the architecture of the function monitor, derived on the basis of the Dependability Cage approach.

A. Overall Concept of the Dependability Cage Approach

In the initial position paper, the Dependability Cage approach was introduced to address three main challenges in the development of autonomous systems: (1) guaranteeing the safe behavior of an autonomous system in an unknown and uncertain environment, (2) ensuring the safe behavior for all safety-critical system components, including machine-learning based components, even when deviations are detected in their behavior during system operation time, and (3) guaranteeing and improving the relevance and completeness of test cases for the validation of the system under test [22].

To tackle these challenges, the paper in [22] proposes an iterative development process consisting of three primary stages. The first stage is *Dependability Cages Engineering and Training in System Development*, in which the dependability cages are engineered in parallel to the autonomous system and tested using simulation tests or tests in a restricted and controlled lab environment [22]. In the second stage, *Runtime System Observation and Resilience System Stabilization*, the dependability cages are used to monitor the system behavior during its operation and record any deviation in the system behavior compared to the test results obtained during system development as well as any novel situations that occur in the environment [22]. In the third stage, *Monitored Data Analysis and Goal-Oriented System Evolution for Dependability Improvement*, the observations logged during system operation are leveraged to improve the development artifacts during the subsequent system development cycle [22]. The deployment of the dependability cages on the actual system during its operation is facilitated by a modular platform architecture used for the seamless development and operation of the system, monitor and system environment [22]. The end goal of this iterative development process is to continuously improve the system's quality in terms of its dependability requirements. For further details on the development process and its phases, the reader is referred to [22]. The safety architecture proposed for autonomous systems in [22] draws its inspiration from the second phase of the iterative development process, *Runtime System Observation and Resilience System Stabilization*. This phase is carried out through a continuous monitoring framework that observes the behavior of the overall system, as shown in [22], [23], and [2]. In this paper, we refine the focus of this monitoring framework and tailoring it to analyze the environment perception subsystem of an ADS, rather then the ADS as a whole. The monitoring framework consists of two types of monitors, a *function monitor* and a *situation monitor*. Additionally, it involves a component responsible for defining the fail-operational reaction of the ADS based on the results of these two runtime monitors, as depicted in Figure 1.

The responsibility of the *situation monitor* - denoted as the quantitative monitor in the original position paper [22] - is to evaluate the input abstract situations used in the environment and self-perception subsystem of the ADS. During system operation, the situation monitor assesses if the input situations encountered by the ADS align with those considered during the development phase of the environment perception. As the research in this paper does not focus on the situation monitor, the reader is referred to [24] and [25] for a more detailed description of its concept and functionality.

On the other hand, the *function monitor* - denoted as the qualitative monitor in the original position paper [22] - evaluates if there is a critical deviation in the behavior of the environment and self-perception subsystem during the operation of the ADS. The function monitor consists of an abstract

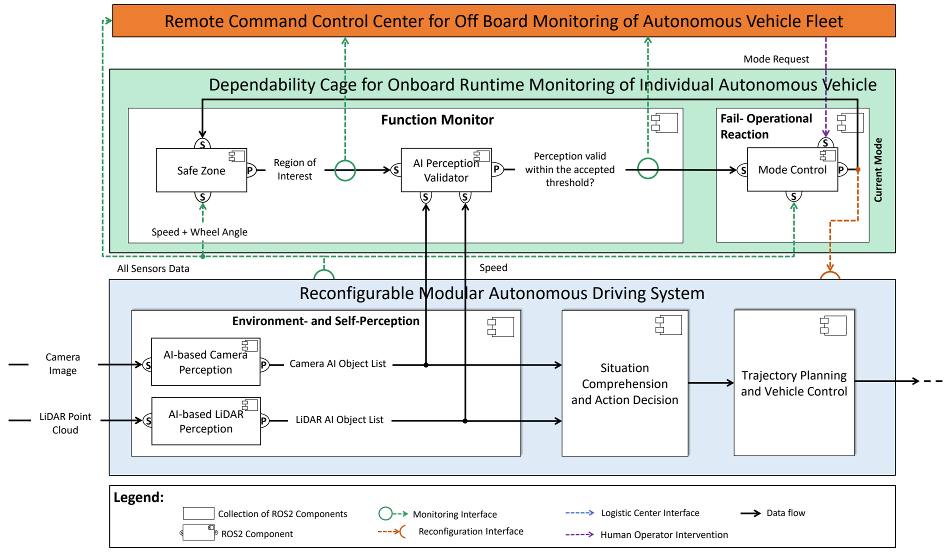

Figure 2: Architecture of the Function Monitor for AI-based Perception Systems.

behaviour's boundary function and a conformity oracle. The abstract behaviour's boundary function dynamically computes a set of safety boundaries, conceptualized as a Region Of Interest (ROI) for the ADS. The conformity oracle assesses the abstraction of the environment and self-perception's outputs in order to check if these outputs are consistent with each other within the safety boundary. Consistency is evaluated based on certain threshold values that are established by empirical tests during the development phase of the environment and self-perception of the ADS. The architecture of the function monitor is explained in detail in the Section III-B.

B. Dependability Cage for AI-based Environment Perception in the ADS's Integrated Safety Architecture

The dependability cage for AI-based environment perception is integrated in the three layered safety architecture developed for ADSs. Beginning from the top, the first layer is represented by the *Remote Command Control Center (CCC)*, where the sensors data stream is visualized along with the results provided by the components in the layers below [23], [1]. The remote CCC has been previously showcased in a separate work [1] and is therefore not the primary focus of the current paper.

The research focus of this paper lies in the middle layer and the bottom layer of the integrated safety architecture. The middle layer encompasses the dependability cage for the AI-based environment perception of the ADS. The dependability cage consists of two main components: (1) the function monitor, responsible for observing and analyzing the environment perception system during the ADS's operation and (2) the fail-operational reaction component, which triggers a fail-operational reaction of the ADS based on the results of the function monitor. The third layer represents the reconfigurable modular autonomous driving system [26], which draws inspiration from previous work [4], [5], and [27], and consists of three main subsystems: (1) Environment and Self-Perception, (2) Situation Comprehension and Action Decision, and (3) Trajectory Planning and Vehicle Control.

The environment and self-perception subsystem consists of two components, AI-based Camera Perception and AI-based LiDAR Perception. These components use camera and respectively LiDAR sensor data to detect objects in the AV's environment. They implement safety-critical machine-learned functions for the ADS's operation. Each component produces an object list, denoted as Camera AI-based (CAI) object list and LiDAR AI-based (LAI) object list. The object lists are used by the other components in the architectural pipeline of ADS. Several pieces of information are provided for each object in the two object lists: (1) object's class, (2) the object's dimensions, height and width, (3) object's distance from the ego-vehicle, (4) sensing timestamp, and (5) confidence level of detection. Figure 2 provides an overview of the integrated safety architecture with a focus on the function monitor and the environment and self-perception system.

The function monitor observes the behavior of the environment and self-perception subsystem against to a specified safety requirement. The safety requirement is informally formulated in a controlled natural language as follows:

Safety Requirement (Informal Specification). The environment and self-perception system must always consistently detect the objects located inside the autonomous vehicle's region of interest using at least two different sensor data sources.

This safety requirement mandates two aspects: first, that a ROI is computed around the AV, and second, that the objects detected by the perception components in the vehicle's region of interest are consistent with each other. The function monitor consists of two components, which allow it to check this safety requirement during the ADS's operation, denoted as *Safe Zone* and *AI Perception Validator*.

1) Safe Zone: This component utilizes various parameters of the AV, e.g., current speed, steering angle, physical dimen-

 ISBN: 978-1-68558-153-4

sions of the AV, acceleration and deceleration, to dynamically calculate ROI. The ROI expands around the AV in the vehicle's direction of travel, and is divided in two areas: a *focus zone* marked in orange and a *clear zone* marked in green around the ego-vehicle. Figure 5 gives a visual intuition of the ROI, depicted on the Graphical User Interface (GUI) of the remote CCC. The ROI around the ego-vehicle is understood as a safety-critical area in which the outputs of the two perception components align consistently. Further details regarding the algorithm used for the ROI computation can be found in [23].

2) AI Perception Validator: This component takes as inputs the ROI computed by the Safe Zone component and the object lists produced by the LiDAR-based and camera-based perception components and computes a boolean flag *valid*, which indicates if the object lists from both perception components are consistent within certain threshold limits (as indicated in lines 1 - 2 in Figure 3). The AI Perception Validator leverages the computed ROI to prune the set of objects detected by the camera and LiDAR sensors in the respective fields of view (as indicated in lines 3 - 4 in Figure 3). It prioritizes those objects detected in close proximity to the ego-vehicle, i.e. inside the ROI of the vehicle. Such a prioritisation differentiates clearly between the objects situated inside the ROI that are safety-relevant for the AV, from the objects situated outside the ROI, which do not pose an immediate safety concern. The threshold limits for determining the consistency of the two object lists are established through a Hazard Analysis and Risk Assessment (HARA). Given that the camera and LiDAR sensors operate at different time rates due to their inherent configuration [28], the CAI object list and the LAI object list will be generated at different frequencies. This time synchronization problem is addressed by using a timeout limit in the AI Perception Validator for the timestamp of the detected objects. It means that an object with a timestamp older than the timeout limit is filtered out from the respective object list (as indicated in lines 9 - 22 in Figure 3). Subsequently, the AI Perception Validator compares the attributes of each object in the CAI object list with the corresponding attributes of each object in the LAI object list, e.g. object class, object distance from the ego vehicle, width and height of the object. If the difference between the respective attributes does not exceed the respective threshold value determined through the HARA analysis, then it is considered a match between both object lists and the AI Perception Validator returns $true$ (as indicated in lines 23 - 26 in Figure 3). It means that the outputs of the two perception components are consistent with each other and all objects have passed the validation and matching criteria. Another case for returning $true$ is that both CAI object list and LAI object list are empty. In this case, there is no need to compare the two object lists. If at least one object in any of the two lists does not meet the comparison criteria, the AI Perception Validator returns $false$, meaning that the outputs of the two perception components are inconsistent (as indicated in lines 27 - 28 in Algorithm 3).

Algorithm: AI Perception Validator Algorithm

```
1  Input: caiObjList - CAI object list; laiObjList - LAI
     object list, roi - vehicle's ROI
2  Output: valid - a boolean flag
3  cameraObjList ← FilterObjectList(caiObjList, roi);
4  lidarObjList ← FilterObjectList(laiObjList, roi);
5  valid ← true;
6  removed ← false;
7  if cameraObjList and lidarObjList are empty then
8  |  return valid
9  foreach lidarObj in lidarObjList do
10 |  if timestamp of lidarObj is older than the
11 |  timeout limit then
12 |  |  if lidarObj is not matched then
13 |  |  |     valid ← false;
14 |  |  remove lidarObj from lidarObjList;
15 |  |  removed ← true;
16 foreach cameraObj in cameraObjList do
17 |  if timestamp of cameraObj is older than the
18 |  timeout limit then
19 |  |  if cameraObj is not matched then
20 |  |  |  valid ← false;
21 |  |  remove cameraObj from cameraObjList;
22 |  |  removed ← true;
23 foreach lidarObj in lidarObjList do
24 |  foreach cameraObj in cameraObjList do
25 |  |  if difference between lidarObj attributes and
   |  |  cameraObj attributes are under the threshold
   |  |  values then
26 |  |  |  mark lidarObj and cameraObj as matched;
27 |  if removed is true then
28 |  |  return valid
```

Figure 3: AI Perception Validator Algorithm.

The result of the AI Perception Validator as well as the two object lists are forwarded to the remote CCC for visualization (see Section IV). Additionally, the result computed by the AI Perception Validator serves as an input in the component *Mode Control*. In case of inconsistency between the two object lists, this component is responsible for triggering a fail-operational reaction of the AV, by gracefully degrading the ADS functionality. This process is similar to approach outlined in [23]. However, since the primary research focus of this paper is the function monitor, the definition, implementation and evaluation of the corresponding fail-operational reaction will be addressed in future work.

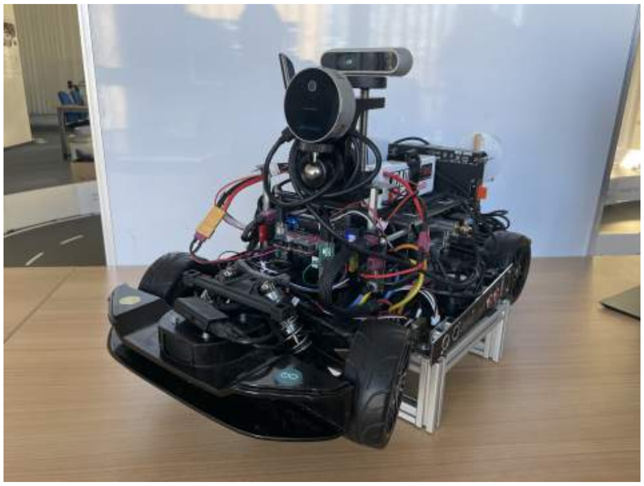

Figure 4: Model Car

IV. Evaluation and Discussion of Results

This section presents the evaluation of the function monitor for the runtime validation of AI-based environment perception in ADS. The first part of this section introduces a detailed overview of the setup, including both hardware and software details (see Section IV-A). Subsequently, in Section IV-B, various test scenarios and several working hypotheses are defined. A qualitative scenario-based evaluation is conducted to assess the defined hypotheses and the obtained results are discussed.

A. Evaluation Setup

1) Physical Hardware Platform and Test Track: For the evaluation of the function moniotr concept, a model vehicle on the scale of 1:8, developed by Digitalwerk [29], is chosen as the physical hardware platform. The model vehicle is equipped with a wide range of sensors, including a mono camera, a LiDAR sensor, wheel speed sensors, ultrasonic sensors, a Global Positioning System (GPS) sensor, and an Inertial Measurement Unit (IMU).

To enhance its environmental perception capabilities for the validation of the function monitor, further sensors have been installed on the model vehicle, e.g., an Intel RealSense LiDAR camera (L515) [30] and a stereo vision camera (D435f) [31]. The LiDAR camera provides sensor input data for the LiDAR AI-based perception component. Although a high-resolution 3D LiDAR sensor would have been ideal, the model vehicle's limited power supply led to the deployment of a LiDAR camera with lower power requirements as a good compromise solution, which still provides adequate data output. Both sensors have been calibrated based on the vehicle's rear axle, to ensure that generated object lists are in the same coordinate system, specifically in the vehicle coordinate system. This alignment is crucial for an accurate and coherent comparison of redundant perception systems. Figure 4 illustrates the model car with the LiDAR camera and stereo vision camera mounted on it.

The test track utilized for the evaluation was constructed in the lab environment using modular martial arts mats. Each black mat measures 1m $\times$ 1m and is adorned with street markings and track walls [32]. Figure 5 depicts the model vehicle placed on the test track along with other objects that emulate other traffic participants, such as a pedestrian represented by a wooden human dummy, and elements of the road infrastructure, e.g., traffic light.

2) Implementation Details: The function monitor conducts a consistency comparison between two object lists, a CAI object list and a LAI object list. These two lists are generated by respective AI-based perception components, one based on camera input and the other on LiDAR input. Both perception components apply YOLO Nano 2D object detectors [33], which yield 2D bounding boxes with object class names and and their respective confidence scores, but lack the distance information between the ego-vehicle and the corresponding objects. By leveraging the LiDAR point cloud provided by the LiDAR camera, we computed the distance between the model vehicle and the objects, referred to as depth, thereby producing 2.5D bounding boxes. The 2.5D bounding boxes differ from the 3D bounding boxes in that the latter include all three dimensions, i.e., height, width and length of the bounding box.

The outputs of the environment and self-perception subsystem along with the result of the function monitor are visualized in the GUI of the remote CCC (see Section III). Figure 7 depicts the visualization of the function monitor result in the Car Selection panel of the remote CCC, utilizing a flag called *AI perception Validator*. The flag's color indicates different results of the function monitor: (1) **green** - denotes consistency between the two object lists, (2) **red** - signifies inconsistency, and (3) **black** - denotes data not being received by the AI Perception Validator component in the function monitor.

In the center of Figure 7, two panels display the view of the LiDAR camera (upper panel) and the stereo vision camera (lower panel) with the bounding boxes corresponding to each object list highlighted in green on their respective sensor view. Additionally, the Sensor Visualization panel depicts the ROI computed around the model vehicle and comprising a focus zone, marked in orange, and a clear zone, marked in green, as introduced in Section III-B. The implementation of each component is based on the decentralized middleware ROS2, facilitating the communication between the ROS2 components through the publish-subscribe pattern. This solution provides advantageous features such as self-adaptation and component reconfiguration at runtime, aligning with the distributed nature of the AV, where components are distributed across different electronic control units (ECUs) [34]. Furthermore, the real-time capabilities of ROS2 make it appropriate for ensuring the safety and reliability of the ADS.

B. Definition of Test Scenarios and Research Hypotheses

To evaluate the function monitor concept, we conducted a qualitative scenario-based assessment. We defined several test scenarios along with working hypotheses to guide the evaluation. The test scenarios range from simple to complex, starting with a single static object and gradually increasing the

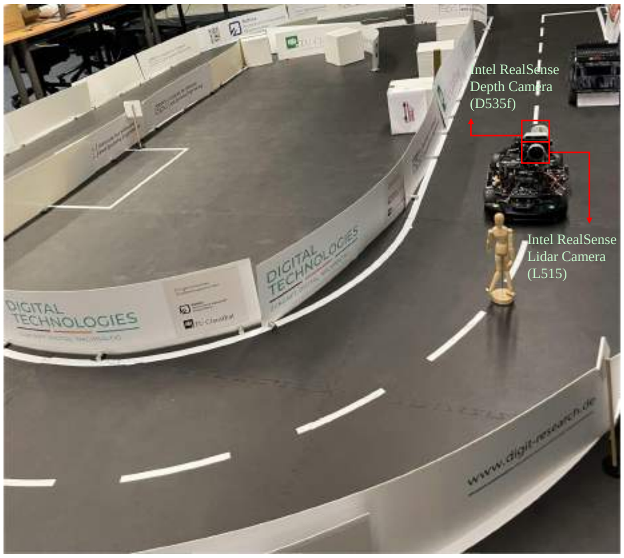

Figure 5: Model Car: bird's-eye view in a Lab Environment.

scenario complexity, by incorporating multiple static objects in a static environment. Each test scenario includes a description of the physical actions of the model vehicle and its environment. Following three test scenarios were defined evaluating the function monitor:

Test Scenario 1 (TS 1). The model vehicle is stationary on the test track and a pedestrian (represented by a wooden human dummy) is placed in front of the model vehicle, outside of its ROI. The pedestrian is placed in such a way that it is detected by the LiDAR camera, but not by the stereo vision camera.

Test Scenario 2 (TS 2). The model vehicle is stationary on the test track and a pedestrian (represented by a wooden human dummy) is placed in front of the model vehicle, inside of its ROI. The pedestrian is positioned in such a way that it is detected by the LiDAR camera, but not by the stereo vision camera.

Test Scenario 3 (TS 3). The model vehicle is stationary on the test track and a traffic light is placed in front of the model vehicle, inside of its ROI. The traffic light is positioned so that both the LiDAR camera and the stereo vision camera are able to detect it.

In addition to the test scenarion, several research hypotheses are formulated in this paper to assess the expected performance of the function monitor. The function monitor has been evaluated in the defined test scenarios with respect to the following two hypotheses:

Hypothesis 1 (H1). The function monitor accurately identifies that the CAI object list and the LAI object list are consistent with each other.

Hypothesis 2 (H2). The function monitor accurately identifies that the CAI object list and the LAI object list are not consistent with each other.

C. Discussion of Results

The evaluation results of the function monitor on hypotheses H1 and H2 in all the test scenarios defined for the evaluation

Figure 6: Wooden Human Dummy outside Model Vehicle's ROI.

are presented in Table I. In TS 1, in which the wooden dummy is placed in front of the vehicle and outside of its ROI, the LiDAR camera can detect it but the stereo vision camera cannot. In this scenario, the AI Perception Validator gives the result *consistent* since the wooden dummy is outside of the model vehicle's ROI, and thus, both processed object lists are empty. Therefore, in TS 1, hypothesis H1 is *true* and hypothesis H2 is *false*. The CAI object list and the LAI object list along with the flag of the AI Perception Validator are visually depicted in Figure 6, in which the AI Perception Validator flag shows a green status, indicating "consistent object lists".

TABLE I: EVALUATION RESULTS OF THE FUNCTION MONITOR.

Hypotheses \ Test Scenarios	TS 1	TS 2	TS3
H1	True	False	True
H2	False	True	False

In TS 2, the wooden human dummy is placed again in front of the vehicle, but this time inside its ROI. The human dummy is placed so that it is detected by the LiDAR camera but not by the stereo vision camera. In this scenario, the AI Perception Validator returns *inconsistent*, since there is at least an inconsistent object in either the CAI object list or the LAI object list, which in this case is the human dummy. Thus, in TS 2, hypothesis H1 is *false* and hypothesis H2 is *true*. The CAI object list and the LAI object list along with the flag of the AI Perception Validator are shown in Figure 7. The AI Perception Validator flag shows a red status, indicating "inconsistent object lists" since the objects are inside the vehicle's ROI but not aligned with each other. In TS 3, a traffic light is positioned in front of the model vehicle inside its ROI, so that both the LiDAR camera and the stereo vision camera can detect it. In this scenario, the AI Perception Validator returns *consistent*, since the objects are inside the ROI and were detected by both sensors. Thus, in TS 3, hypothesis H1 is *true* and hypothesis H2 is *false*. The CAI object list and the LAI object list along with the flag of the AI Perception Validator are shown in Figure 8, in which the AI Perception

 ISBN: 978-1-68558-153-4

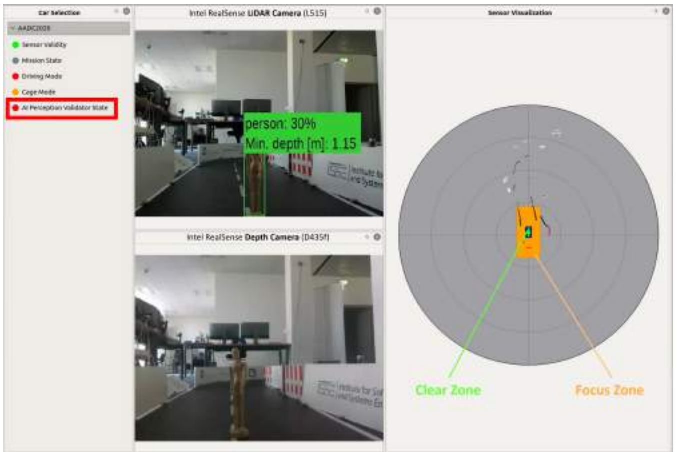

Figure 7: Wooden Human Dummy inside Model Vehicle's ROI.

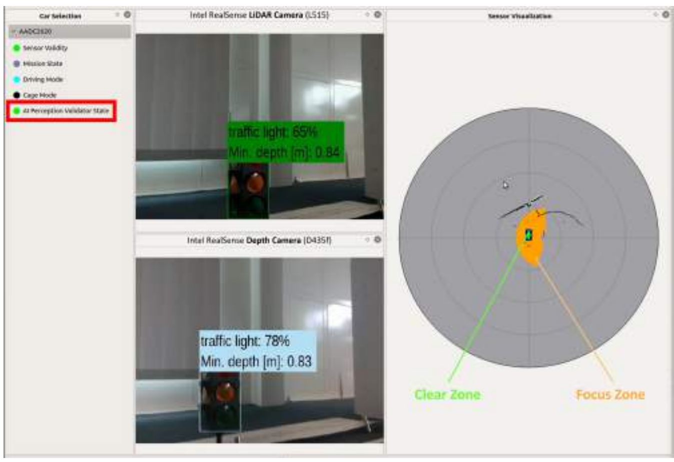

Figure 8: Traffic Light situated Inside Model Vehicle's ROI.

Validator flag indicates a green status, indicating "consistent object lists".

The results obtained in the three test scenarios have confirmed the formulated hypotheses in the defined test scenarios, showing that the function monitor is operating as intended.

V. CONCLUSION

This paper outlines a method for the runtime monitoring and validation of AI-based environment perception systems employed in autonomous driving contexts. It builds upon the Dependability Cage approach, initially proposed in [22], with a specific focus on the environment and self-perception subsystem in an ADS. The environment perception consists of two redundant perception components tasked with object detection in the ego-vehicle surrounding environment, which leverage multiple sensor data sources, e.g., LiDAR and camera. The dependability cage for the environment perception comprises a function monitor and a fail-operational reaction component. The function monitor checks at runtime whether the outputs of the two perception components remain consistent. Meanwhile, the fail-operational reaction component dictates the fail-safe or the fail-operational reaction of the ADS based on the feedback of the function monitor. This study was primarily focused on the function monitor, which was evaluated qualitatively using predefined test scenarios and a model car in a lab environment. The results of the evaluation demonstrated that the function monitor works as expected.

The test scenarios employed in the evaluation focused on relatively simple driving situations, with stationary objects and a stationary model car. However, in future work, we plan to enhance and extend the functionality of the function monitor so that it covers more complex scenarios, with both dynamic and static obstacles. Moreover, we intend to define a method for defining appropriate fail-operational reactions to gracefully degrade the ADS's functionality [35] in response to warning signals given out by the function monitor. Lastly, we plan to integrate the function monitor with the concept of situation monitor, similar to the one presented in [25]. Such integration enables the ADS to be aware of new object classes detected in its environment, thus enhancing its capability to handle novel environment situations. Ultimately, this integration will contribute to the safety and reliability of autonomous driving systems in diverse and challenging real-world scenarios.

ACKNOWLEDGMENT

The authors express their gratitude for the collaboration of all cooperative partners. This work was carried out as part of the project *safeWahr - Safe release and reliable series operation through continuous real-time monitoring of the environmental perception of autonomous vehicles* [36] under the grant number 19A21026E and was funded by the Federal Ministry for Economic Affairs and Climate Protection [37] in Germany.

REFERENCES

[1] A. Aniculaesei et al., "Connected dependability cage approach for safe automated driving," in *International Stuttgart Symposium*, pp. 3–21, Springer, 2023.

[2] I. Aslam, A. Aniculaesei, A. Buragohain, D. Bamal, and P. Rausch, "Runtime safety assurance of autonomous vehicles used for last-mile delivery in urban environments," *arXiv preprint arXiv:2307.04454*, 2023.

[3] SAE, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles." SAE International, Warrendale, PA, USA, June 2018. Norm.

[4] S. Behere and M. Törngren, "A functional architecture for autonomous driving," in *Proceedings of the First International Workshop on Automotive Software Architecture* (P. Kruchten, Y. Dajsuren, H. Altinger, and M. Staron, eds.), (New York, NY, USA), pp. 3–10, ACM, 2015.

[5] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomes Fahren - Technische, rechtliche und gesellschaftliche Aspekte*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.

[6] S. Zeitung, "Tödlicher unfall im selbstfahrenden tesla." online, July 2016. Newspaper.

[7] J. Golson, "Tesla's new autopilot will run in 'shadow mode' to prove that it's safer than human driving." online, Oct. 2016. The Verge.

[8] M. Mauritz, A. Rausch, and I. Schaefer, "Dependable adas by combining design time testing and runtime monitoring," in *FORMS/FORMAT 2014 - 10th Symp. Form. Methods Autom. Saf. Railw. Automot. Syst.*, pp. 28–37, 09 2014.

[9] M. Mauritz, F. Howar, and A. Rausch, "Assuring the safety of advanced driver assistance systems through a combination of simulation and runtime monitoring," in *7th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation (ISoLA)*, vol. 9953, pp. 672–687, 10 2016.

[10] ISO, "Road vehicles – functional safety." ISO, Geneva, Switzerland, Dec. 2018. Norm.

[11] ISO, "Road vehicles — safety of the intended functionality." ISO, Geneva, Switzerland, June 2022. Norm.
[12] "Asam openscenario." https://www.asam.net/standards/detail/openscenario-xml/, 2022. Norm.
[13] "Asam opendrive." "https://www.asam.net/standards/detail/opendrive/, 2023. Norm.
[14] K. Czarnecki and R. Salay, "Towards a framework to manage perceptual uncertainty for safe automated driving," in *Computer Safety, Reliability, and Security: SAFECOMP 2018 Workshops, ASSURE, DECSoS, SASSUR, STRIVE, and WAISE, Västerås, Sweden, September 18, 2018, Proceedings 37*, pp. 439–445, Springer, 2018.
[15] Q. M. Rahman, P. Corke, and F. Dayoub, "Run-time monitoring of machine learning for robotic perception: A survey of emerging trends," *IEEE Access*, vol. 9, pp. 20067–20075, 2021.
[16] W. Xiang et al., "Verification for machine learning, autonomy, and neural networks survey," *arXiv preprint arXiv:1810.01989*, 2018.
[17] F. Leofante, N. Narodytska, L. Pulina, and A. Tacchella, "Automated verification of neural networks: Advances, challenges and perspectives," *arXiv preprint arXiv:1805.09938*, 2018.
[18] D. Feng, Z. Wang, Y. Zhou, L. Rosenbaum, F. Timm, K. Dietmayer, M. Tomizuka, and W. Zhan, "Labels are not perfect: Inferring spatial uncertainty in object detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 9981–9994, 2021.
[19] Q. M. Rahman, N. Sunderhauf, and F. Dayoub, "Per-frame map prediction for continuous performance monitoring of object detection during deployment," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 152–160, 2021.
[20] Q. M. Rahman, N. Sünderhauf, and F. Dayoub, "Online monitoring of object detection performance during deployment," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4839–4845, IEEE, 2021.
[21] A. Gupta and L. Carlone, "Online monitoring for neural network based monocular pedestrian pose estimation," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8, IEEE, 2020.
[22] A. Aniculaesei, J. Grieser, A. Rausch, K. Rehfeldt, and T. Warnecke, "Towards a holistic software systems engineering approach for dependable autonomous systems," in *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems* (R. Stolle, S. Scholz, and M. Broy, eds.), (New York, NY, USA), pp. 23–30, ACM, 2018.
[23] F. Helsch, I. Aslam, A. Buragohain, and A. Rausch, "Qualitative monitors based on the connected dependability cage approach," vol. 11, no. 21, p. 46 to 55, 2021.
[24] A. Rausch, A. M. Sedeh, and M. Zhang, "Autoencoder-based semantic novelty detection: Towards dependable ai-based systems," *Applied Sciences*, vol. 11, no. 21, p. 9881, 2021.
[25] N. Habib, Y. Cho, A. Buragohain, and A. Rausch, "Towards exploring adversarial learning for anomaly detection in complex driving scenes," in *Deep Learning Theory and Applications* (D. Conte, A. Fred, O. Gusikhin, and C. Sansone, eds.), (Cham), pp. 35–55, Springer Nature Switzerland, 2023.
[26] L. Everding et al., "Dynamically configurable autonomous vehicles for urban cargo transportation," in *Towards the New Normal in Mobility: Technische und betriebswirtschaftliche Aspekte*, pp. 851–869, Springer, 2023.
[27] S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," *Information and Software Technology*, vol. 73, pp. 136–150, 2016.
[28] Z. Rozsa and T. Sziranyi, "Virtually increasing the measurement frequency of lidar sensor utilizing a single rgb camera," *arXiv preprint arXiv:2302.05192*, 2023.
[29] "Welcome to digitalwerk - regensburg." https://www.digitalwerk.net/. (Accessed on 01/31/2024).
[30] "Lidar camera l515 – intel® realsense™ depth and tracking cameras." https://www.intelrealsense.com/lidar-camera-l515/. (Accessed on 01/31/2024).
[31] "Depth camera d435f – intel® realsense™ depth and tracking cameras." https://www.intelrealsense.com/depth-camera-d435f/. (Accessed on 01/31/2024).
[32] T. Warnecke, J. Grieser, M. Zhang, A. Vorwald, and A. Rausch, "Teaching novices supervised learning with autonomous model vehicles," in *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T)*, pp. 1–10, IEEE, 2020.
[33] A. Wong, M. Famuori, M. J. Shafiee, F. Li, B. Chwyl, and J. Chung, "Yolo nano: A highly compact you only look once convolutional neural network for object detection," in *2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS)*, pp. 22–25, IEEE, 2019.
[34] C. Raulf et al., "Dynamically configurable autonomous vehicle concepts for passenger transport," in *Transforming Mobility–What Next? Technical and business aspects*, pp. 265–288, Springer, 2022.
[35] A. Aniculaesei, J. Grieser, A. Rausch, K. Rehfeldt, and T. Warnecke, "Graceful degradation of decision and control responsibility for autonomous systems based on dependability cages," in *5th International Symposium on Future Active Safety Technology toward Zero, Blacksburg, Virginia, USA*, 09 2019.
[36] "Safewahr:sichere freigabe und zuverlässiger serienbetrieb durch kontinuierlisches echtzeitfähiges monitoring der umgebungswahrnehmung autonomer fahrzeuge." https://zdin.de/digitales-niedersachsen/projektubersicht/safewahr, 2021. (Accessed on 11/03/2024).
[37] "Bmkw:bundesministerium für wirtschaft und klimaschutz." https://www.bmwk.de/Navigation/DE/Home/home.html. (Accessed on 11/03/2024).

On the Regularization of a Low-Complexity Recursive Least-Squares Adaptive Algorithm

Cristian-Lucian Stanciu, Cristian Anghel, Camelia Elisei-Iliescu, Laura-Maria Dogariu, Ionuţ-Dorinel Fîciu, and Constantin Paleologu

Department of Telecommunications, Politehnica Bucharest, Romania

Emails: cristian@comm.pub.ro, canghel@comm.pub.ro, camelia.elisei@romatsa.ro, ldogariu@comm.pub.ro, ionut.ficiu22@gmail.com, pale@comm.pub.ro

***Abstract*—The Recursive Least-Squares (RLS) family of adaptive algorithms can be an attractive choice for the identification of unknown acoustic systems, which have hundreds, or even thousands, of coefficients. The RLS have also been combined with Line Search Methods (LSMs) in order to obtain versions without numerical stability issues, and to decrease the corresponding arithmetic complexity. Despite the superior tracking speeds associated with the RLS-LSM methods (with respect to more consecrated algorithms), they remain vulnerable to Double-Talk (DT) situations, when the corresponding update process becomes inaccurate. This paper describes a variable regularization technique for the RLS-LSM general algorithm, which is designed to mitigate DT scenarios by adjusting the contents of the RLS correlation matrix. Simulation results demonstrate the proposed theoretical model in the stereophonic acoustic echo cancellation configuration.**

***Index Terms*—*Recursive Least-Squares (RLS)*; *Line Search Methods (LSMs)*; *Double-Talk (DT)*; *Variable Regularization (VR)*.**

I. Introduction

The identification of unknown acoustic echo paths using adaptive algorithms implies the estimation of impulse responses equivalent to hundreds of millisenconds. Most of the signal processing solutions rely on the Least-Mean-Square (LMS) adaptive methods [1], which have acceptable arithmetic complexities and poor performances when working with highly correlated signals, such as speech.

In this context, the Recursive Least-Squares (RLS) adaptive algorithms are possible alternatives, having superior tracking speeds [2]. However, most RLS versions require prohibitive amounts of resources on most modern chips and manifest numerical stability issues. The complexity problem is even worse when considering the Stereophonic Acoustic Echo Cancellation (SAEC) scenarios, where user terminals employ two microphones, respectively two loudspeakers, to create the impression of audio directionality. The associated setup needs to estimate four acoustic paths, corresponding to each loudspeaker-to-microphone pair.

In [3], Liu et al. combined the exponentially weighted RLS with Line Search Methods (LSMs) in order to approach the corresponding set of normal equations by solving an auxiliary system. The solution avoids the numerical stability issues, and allows the use of less complex LSM variants, which exploit the statistical properties of the input signal. Moreover, for the SAEC setup, the Widely Linear (WL) model was employed in [4] in order to simplify the handling and to allow easier developments of extra features, like mechanisms employed to mitigate the effects of Double-Talk (DT) situations.

This paper is organized as follows. Section II describes the WL-RLS-LSM adaptive algorithm with Variable Regularization (VR) of the associated correlation matrix. In Section III, simulation results are discussed for the proposed algorithm in the SAEC setup. The paper draws the main conclusions in Section IV.

TABLE I
WL-RLS-LSM ALGORITHM

Step	Action
Init.	Set: $\widetilde{\mathbf{g}}(0) = \mathbf{0}_{2L\times 1};\ \mathbf{r}(0) = \mathbf{0}_{2L\times 1}$ $\mathbf{R}(0) = \Phi\mathbf{I}_{2L},\ \Phi > 0; 0 < \lambda \leq 1$
	For $n = 1, 2, \ldots,$ number of iterations :
1	Update $2L \times 1$ input vector $\widetilde{\mathbf{x}}(n)$
2	Update correlation matrix $\mathbf{R}(n)$ (time shift) $\mathbf{R}^{(:,1)}(n) = \lambda\mathbf{R}^{(:,1)}(n-1) + x^*(n)\widetilde{\mathbf{x}}(n)$
3	$\widetilde{y}(n) = \widetilde{\mathbf{g}}^H(n-1)\widetilde{\mathbf{x}}(n)$
4	$e(n) = d(n) - \widetilde{y}(n)$
5	$\mathbf{g}_0(n) = \lambda\mathbf{r}(n-1) + e^*(n)\widetilde{\mathbf{x}}(n)$
6	$\mathbf{R}(n)\Delta\mathbf{g}(n) = \mathbf{g}_0(n) \xrightarrow[N_u]{\text{LSM}} \Delta\widetilde{\mathbf{g}}(n), \mathbf{r}(n)$
7	$\widetilde{\mathbf{g}}(n) = \widetilde{\mathbf{g}}(n-1) + \Delta\widetilde{\mathbf{g}}(n)$

II. The VR-WL-RLS-LSM Algorithm

The WL-RLS-LSM adaptive filter working in the SAEC configuration benefits from the simplifications provided by the WL model, and employs a single adaptive filter $\widetilde{g}(n)$ with $2L$ complex valued coefficients in order to estimate the four acoustic loudspeaker-to-microphone impulse responses, each with L real valued coefficients. The input information (corresponding to the two loudspeakers) is grouped in a single complex valued signal $x(n)$, respectively the outputs of the unknown echo paths are combined into $y(n)$. The complex valued microphone information, represented by $d(n)$, comprises the contribution of $y(n)$ cumulated with environmental noise. Consequently, the adaptive filter estimates the complex echo as $\widetilde{y}(n)$ and sends to the interlocutor the error $e(n)$, from which the value $\widetilde{y}(n)$ is subtracted. The WL-RLS-LSM general algorithm is presented in Table I, where $\mathbf{I}_{2L}$ denotes the identity matrix, and ^{H} is the Hermitian operator.

 ISBN: 978-1-68558-153-4

The first two steps of the algorithm are dedicated to the updates of the input vector and the correlation matrix, respectively. The output of the filter results in step 3, while the error signal is computed in step 4. Then, the residual component $\mathbf{g}_0(n)$ is evaluated in step 5. In step 6 of the algorithm, the method employs a complex valued LSM to solve an auxiliary system of equations and generate the *solution vector* $\Delta\mathbf{g}(n)$, which is used in step 7 to update the filter estimate.

A wide range of LSMs can be used for step 6 of the algorithm. For example, the complex valued Conjugate Gradient (CG) method has a complexity proportional to $4L^2$ real valued multiplications and attractive tracking capabilities. A low-complexity alternative for the CG are the Dichotomous Coordinate Descent (DCD) iterations [5], [6], which can solve the system in step 6 using only additions and bit-shifts. The DCD strongly relies on the statistical properties of the input signal as they are reflected in the correlation matrix $\mathbf{R}(n)$. The WL-RLS-DCD variant is more attractive for hardware implementations, because it can function with an overall arithmetic workload proportional to $2L$.

However, regardless of the algorithm used for the SAEC, the WL-RLS-LSM is still susceptive to DT scenarios, when the noise component of $d(n)$ is much higher than the contribution of $y(n)$. A solution proposed in [7] is to alter the purpose of the initialization constant Φ used to avoid the singular nature of $\mathbf{R}(n)$ during the initial iterations of the algorithm. We can compute a regularization coefficient at every time index as

$$\Phi(n) = 2L\widetilde{\sigma}_x^2(n)(1+\sqrt{1+\widetilde{\mathrm{ENR}}})/(\widetilde{\mathrm{ENR}}), \qquad (1)$$

where $\widetilde{\sigma}_x^2(n)$ and $\widetilde{ENR}$ are estimates for the variance of the input signal, respectively the Echo-to-Noise Ratio (ENR). Consequently, we can compare the performances of the WL-RLS-LSM using the CG, respectively DCD, methods with their variable regularized counterparts in DT situations. The complexities of the namely VR-WL-RLS-LSM algorithms remain similar to the original versions.

III. Simulation Results

In Figures 1 and 2, the WL-RLS-CG and WL-RLS-DCD methods were compared to the VR-WL-RLS-CG, respectively the VR-WL-RLS-DCD, by simulating a tracking scenario [2], followed by a DT occurence. The input signal is speech, and the ENR was set to 25 dB for Figure 1, respectively to 10 dB for Figure 2. We employed different number of N_u iterations for both algorithms [3], [4]. We can notice in Figure 1 that the VR versions of the algorithms perform better during the DT interval, with the compromise of slightly lower tracking speeds. Moreover, when the ENR is lower (Figure 2), the VR-WL-RLS-CG, respectively the VR-WL-RLS-DCD, clearly outperform their counterparts, including during the steady-state portions of the simulation.

IV. Conclusions

The VR approach for the WL-RLS-LSM improves the performance of the algorithm for smaller ENR values, respectively when DT situations occur. The increase of performances adds reasonable extra arithmetical effort, respectively slight losses in the tracking speeds. Considering the VR-WL-RLS-DCD has results similar to the VR-WL-RLS-CG, with less necessary arithmetic resources by an order of degree, the former is an attractive choice for practical applications. These regularized algorithms could be more suitable in echo cancellation and noise reduction scenarios, where the long length impulse responses and highly correlated input signals represent significant challenges for the LMS-based algorithms.

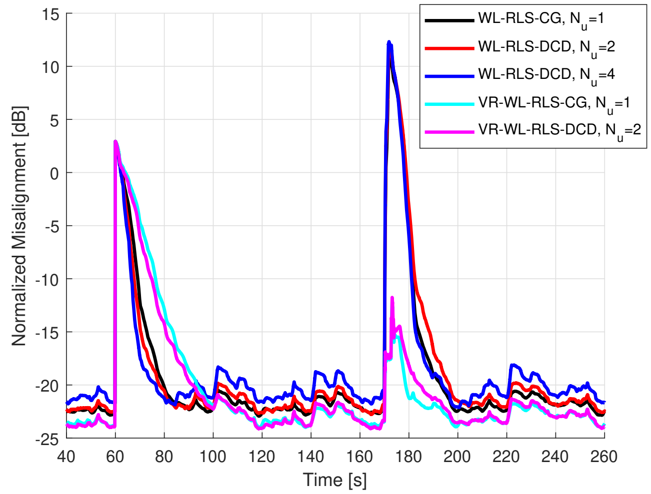

Figure 1. Normalized misalignment of the WL-RLS-CG, WL-RLS-DCD, VR-WL-RLS-CG, and VR-WL-RLS-DCD for different values of N_u. The four unknown echo paths have the length $L = 256$, and the ENR is experimentally set to 25 dB. The echo paths change at time index t_0 = 60 seconds, and a DT situation occurs in the time interval [170,174] seconds.

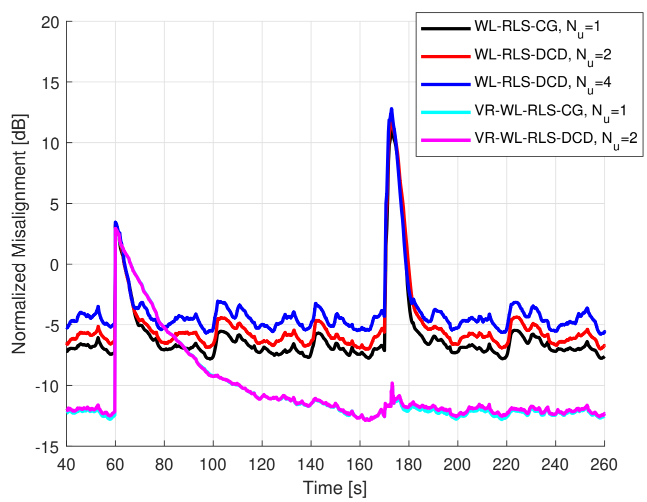

Figure 2. Normalized misalignment of the WL-RLS-CG, WL-RLS-DCD, VR-WL-RLS-CG, and VR-WL-RLS-DCD for different values of N_u. The four unknown echo paths have the length $L = 256$, and the ENR is experimentally set to 10 dB. The echo paths change at time index t_0 = 60 seconds, and a DT situation occurs in the time interval [170,174] seconds.

 ISBN: 978-1-68558-153-4

Acknowledgment

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS–UEFISCDI, project PN-III-P4-PCE-2021-0438, within PNCDI III.

References

[1] S. Haykin, *Adaptive Filter Theory – Fourth Edition*. Prentice Hall, 2002.
[2] J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, *A Perspective on Stereophonic Acoustic Echo Cancellation*, vol. 4. Germany: Springer-Verlag, 2011.
[3] J. Liu, Y. V. Zakharov, and B. Weaver, "Architecture and FPGA Design of dichotomous coordinate descent algorithms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 11, pp. 2425–2438, 2009.
[4] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, "A widely linear model for stereophonic acoustic echo cancellation," *Signal Processing*, vol. 93, no. 2, pp. 511–516, 2013.
[5] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3150–3161, 2008.
[6] Y. V. Zakharov and V. H. Nascimento, "DCD-RLS adaptive filters with penalties for sparse identification," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3198–3213, 2013.
[7] C.-L. Stanciu, C. Anghel, and C. Elisei-Iliescu, "Regularized RLS adaptive algorithm with conjugate gradient method," in *International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pp. 18–23, 2023.