Springer Proceedings in Advanced Robotics 33
Series Editors: Bruno Siciliano · Oussama Khatib

Cristian Secchi
Lorenzo Marconi *Editors*

European Robotics Forum 2024

15th ERF, Volume 2

spar

Springer

Springer Proceedings in Advanced Robotics 33

The Springer Proceedings in Advanced Robotics (SPAR) publishes new developments and advances in the fields of robotics research, rapidly and informally but with a high quality.

The intent is to cover all the technical contents, applications, and multidisciplinary aspects of robotics, embedded in the fields of Mechanical Engineering, Computer Science, Electrical Engineering, Mechatronics, Control, and Life Sciences, as well as the methodologies behind them.

The publications within the "Springer Proceedings in Advanced Robotics" are primarily proceedings and post-proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. Also considered for publication are edited monographs, contributed volumes and lecture notes of exceptionally high quality and interest.

An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by SCOPUS, SCIMAGO, WTI Frankfurt eG, zbMATH.

All books published in the series are submitted for consideration in Web of Science.

Cristian Secchi · Lorenzo Marconi
Editors

European Robotics Forum 2024

15th ERF, Volume 2

Springer

Editors
Cristian Secchi
University of Modena and Reggio Emilia
Reggio Emilia, Italy

Lorenzo Marconi
University of Bologna
Bologna, Italy

ISSN 2511-1256 ISSN 2511-1264 (electronic)
Springer Proceedings in Advanced Robotics
ISBN 978-3-031-76427-1 ISBN 978-3-031-76428-8 (eBook)
https://doi.org/10.1007/978-3-031-76428-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

These Volumes collects a series of scientific contributions presented at the European Robotics Forum (ERF) held in Rimini from 13 to 15 March 2024. The forum, which has been held on an annual basis since 2010 around Europe, is the reference event of the EuRobotics association. The latter is an association created with the aim of providing a common home for the main European stakeholders, both industrial, academic and institutional, with interests in robotics. One of EuRobotics' missions is precisely to facilitate interaction between the industrial and research worlds, promoting collaboration with the aim of providing answers to societal problems and identifying strategic areas useful for defining future research roadmaps.

In this context, the ERF is a strategic annual event whose programme aims to be of interest to both industrial and academic players and whose structure is specifically designed to facilitate interaction. ERF is not a trade fair, although there is regularly an exhibition area where industries and research centres can showcase the latest technical and technological advances, and it is not a conference, although there is always a rich programme of workshops and scientific sessions where current project activities and latest research results are presented. The event places itself 'in the middle' by providing a regular event that can be of interest to both industry and pure research with the aim, among other things, of promoting the field of robotics and related disciplines to an ever-wider audience.

The edition of ERF 2024 in Rimini was characterised by a particularly rich scientific programme, specifically designed to complement the spirit of the forum mentioned above. For the first time in the history of ERF, sessions with the presentation of scientific papers were promoted and organised, in the style of those characterising 'regular' conferences, called 'Industrial Scientific Sessions' and 'Insight Scientific Sessions'. The former gathered contributions, mainly from research facilities and universities, on macro-areas explicitly stimulated by industry. In the months leading up to the forum, in fact, a direct call was launched to the many industrial players who are members of EuRobotics and who were asked to specify particularly important areas of development according to their roadmap. The outcome of this survey was then used to calibrate a 'call for papers' on specific topics, thus forming "Industrial-driven" Scientific Sessions. The Insight Scientific Sessions, on the other hand, were an addition to the scientific workshops that have characterised the ERF programme since the first edition. ERF workshops have always been an opportunity to present ongoing projects (European and otherwise) and stimulate debate on topics of particular relevance in the field of robotics. With the 'insights', the aim is to provide a follow-up to these workshops, in which very often the more technical aspect is kept under wraps to the benefit of dissemination and inclusion of the broader community, by presenting more technical scientific insights on the topics that characterise the workshop.

These Volumes collects the 144 contributions presented in the Industrial and Insight Scientific Sessions of ERF 2024. Based on the topics covered, the papers have been divided into four chapters: the first chapter brings together contributions in the field of ‘Mechatronics and Algorithms for Robotics’, touching, among other things, on autonomous navigation, control for robotics and designing aware robots. The second chapter is instead more focused on aspects of “Artificial Intelligence and Robotics” where the emphasis is on the development of efficient and portable robot learning algorithms for real-world settings and also on aspects of computational hardware that enable the implementation of AI algorithms in real applications. The third chapter, on the other hand, contains contributions that fall under the umbrella of ‘Human robot interaction and collaboration’, touching on, among others, issues of trustability, dependability and acceptability of intelligent robots in smart societies, extended reality and safe adaptation for long-term autonomy in human-populated areas. Many applications in robotics are finally presented in the fourth chapter.

Like all large-scale events, ERF2024 saw the decisive contribution of many people and organisations to whom we extend our sincere thanks. We would particularly like to highlight the close and fruitful cooperation with the Eurobotics team represented by President Bernd Liepert, secretary general Reinhard Lafrenz, administrative team Inge Rehorst, Marie Fortems and Javier Lunque and all board members, co-organisers of ERF2024. Special thanks to Antonio Bicchi and Bruno Siciliano, co-general chairs of ERF2024, whose experience and vision contributed significantly to making the event scientifically relevant and internationally visible. Decisive was the work of Giovanni Berselli, Michele Focchi and Nicola Mimmo, to whom great thanks are due for the time and quality put into the organisation of the Robotics Challenges that distinguished ERF2024 and attracted the participation of many young researchers and students from all over Europe. We also acknowledge the patronage of the associations A.N.I.P.L.A. (Associazione Nazionale Italiana Per L’Automazione), I-RIM (Istituto di Robotica e Macchine Intelligenti) and SIRI (Associazione Italiana di Robotica e Automazione), which gave ERF significant visibility within their spheres. Decisive, finally, was the contribution of the PCO AIM group, with the great work done by Marienza Marguglio, Serena Lamarucciola and Federica Russo, and the local team at the Rimini Palacongressi, which saw Lara Sandre and Giulia Viscusi at the forefront, for all logistical and organisational secretarial aspects.

Lorenzo Marconi
Cristian Secchi

Contents

Robotic Applications

Human Robot Interaction and Collaboration

Kinematic and Muscular Assessment of an Active Hand Exoskeleton for Industrial Applications

Francesco Scotto di Luzio[1(✉)], Christian Tamantini[1], Ghita Boutaib[1], Chiara Carnazzo[2], Stefania Spada[2], and Loredana Zollo[1]

[1] Research Unit of Advanced Robotics and Human-Centred Technologies, Università Campus Bio-Medico di Roma, Rome, Italy
f.scottodiluzio@unicampus.it
[2] Stellantis SpA, Turin, Italy

Abstract. The introduction of new technologies to support work in industry has significantly improved working conditions. However, active devices are not yet adopted in assembly lines due to their complexity. This study provides a preliminary assessment of the impact of the IronHand active exoskeleton, analysing both kinematic and physiological aspects. The results show a positive impact of the exoskeleton in terms of physiological range of motion (ROM) and muscle activation. Such preliminary results underscore the need for a more comprehensive and systematic evaluation of active exoskeletal devices on workers.

Keywords: Exoskeletal devices · Worker Health · Kinematic and Muscular Assessment

1 Introduction

Active exoskeletons are conceived to provide benefits to workers across a wide range of industrial applications in performing their tasks [1]. An exoskeleton is composed of several key components conceived to provide external support to the wearer. The mechanical frame is typically made of lightweight and durable materials, such as aluminum, carbon fiber, or composite materials, houses different components and provides structural support. An exoskeleton can be passive, active and semi-passive. A passive exoskeleton provides support and assistance to the wearer without the use of powered actuators or motors, relying on mechanical structures, springs and other passive elements to reduce physical strain on the wearer. Active exoskeletons are powered wearable devices equipped with actuators. They can provide assistance by generating force to support lifting, walking, or other physical tasks. Semi-passive exoskeletons, also known as variable impedance exoskeletons, combine elements of both active and passive exoskeletons. These devices use passive mechanical structures along with adjustable damping or variable stiffness to provide assistance to the wearer [2].

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 3–7, 2024.
https://doi.org/10.1007/978-3-031-76428-8_1

Moreover, in the state of the art there are several examples of hand exoskeletons, such as ExoHand (Festo, USA) [3], which is worn by an operator like a glove and is designed to support the performance of typical assembly tasks carried out on the assembly line, RoboGlove (General Motors and NASA, USA) [4] and IronHand (Bioservo Technologies, Sweden) [5], designed to reduce muscle fatigue in workers, whose tasks are repetitive or intensive, by improving the strength exerted by the user's fingers and palm. In industrial sectors, such as manufacturing or logistics, hand exoskeletons are adopted to assist workers in manipulating heavy objects or reducing physical workload and can help to improve efficiency and reduce the risk of work-related musculoskeletal disorders (WMSDs). Hand exoskeletons have been extensively investigated in the literature especially for rehabilitation and assistance applications, but their use in occupational contexts appears still limited and not much investigated. As a result, they could potentially lead to a reduction in the symptoms of WMSDs and, conceivably, a decrease in the overall incidence rates of such disorders [3]. It's worth noting, however, that most studies conducted on small participant groups performing tasks not always aimed at the work task. This constraint makes it challenging to definitively ascertain the benefits of industrial exoskeletons, despite the optimistic expectations regarding their role in preventing workplace injuries. For these reasons, the aim of this work is to propose and preliminary validate an approach for kinematic and muscular assessment of active hand exoskeletons for manufacturing applications. Therefore, the proposed approach has been tested on a volunteer subject with and without the IronHand active exoskeleton while carrying out a typical work task, to evaluate any kinematic and muscular impact of the exoskeleton on the user.

2 The Proposed Approach

The proposed approach is based on the kinematic and muscular assessment of the user to quantify the biomechanical and muscular impact of the hand exoskeleton on the user. For the kinematic performance of the hand during the execution of the working gesture, a stereophotogrammetric system is adopted to perform a detailed assessment of joint movements and grip dynamics. Electromyographic (EMG) sensors are applied to the skin with double-layered adhesive tapes to monitor changes in EMG activation. EMG data, sampled at 1 kHz, filtered and normalised with respect to Maximum Voluntary Contraction (MVC), are then used to estimate the integrated muscle activity (iEMG), as

$$iEMG = \int_0^T \frac{EMG_{envelope}(t)}{T} \tag{1}$$

where T represents the final time of the task and $EMG_{envelope}(t)$ denotes the envelope of the EMG data for each muscular district. The proposed method was tested on IronHand (Bioservo Technologies, Sweden), an active hand exoskeleton designed to assist in complex tasks requiring several grips. The system comprises a glove with eight pressure sensors and five motors with artificial tendons attached to the fingers of the glove, a backpack and a remote control. The grip force can be adjusted by the user via the remote control or a customised profile and can automatically adapt its performance during the work cycle.

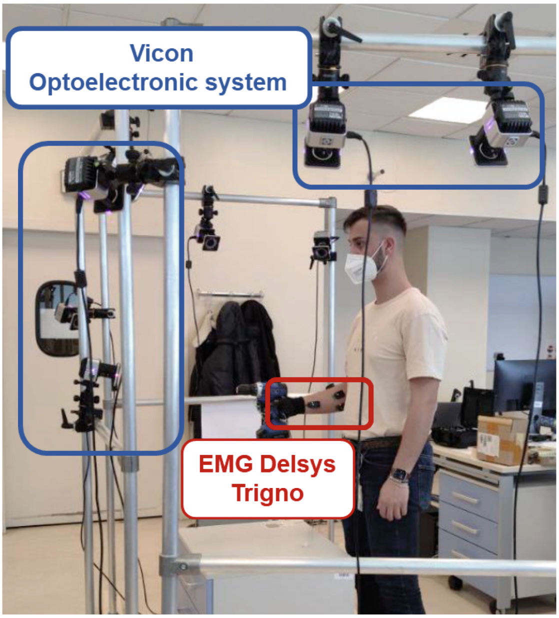


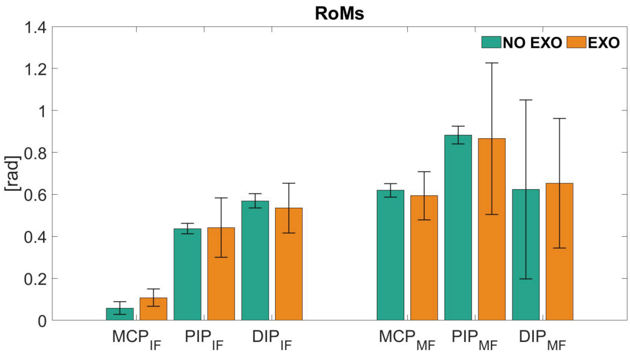
Fig. 1. The experimental setup.

2.1 Experimental Setup and Protocol

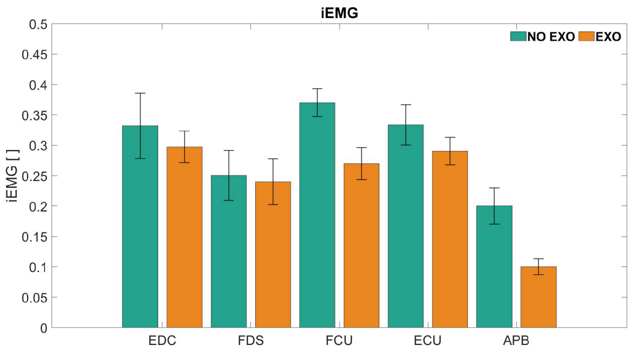
The proposed setup, shown in Fig. 1, is composed of: i) a VICON Nexus stereophotogrammetric system (8 optoelectronic cameras with a resolution of 2.2 MP and a frame rate of 330 Hz), along with a hand model integrated into the Vicon Nexus software. This system enables the reconstruction of the Range of Motion (ROM) of the Index (IF) and Middle Fingers (MF) and ii) Delsys Trigno Wireless system to acquire the EMG data. EMG electrodes were positioned on Extensor Digitorum Communis (EDC), Flexor Digitorum Superficialis (FDS), Flexor Carpi Ulnaris (FCU), Extensor Carpi Ulnaris (ECU) and Abductor Pollicis Brevis (APB). To monitor hand kinematics and muscle activation patterns, one volunteer (28 y.o.) performed five repetitions of a dynamic task. This task consisted of starting from a resting position, grasping an electric screwdriver, lifting it for 5 s to simulate a frontal plane task and then returning to the starting position. This assembly line task was chosen for its grip strength and constant load, making it ideal for assessing biomechanical and muscular risks such as fatigue, discomfort or injury. The task was carried out in two conditions: without the exoskeleton (NO EXO) and with the exoskeleton (EXO), to allow a comparative analysis. The volunteer performed the tasks while seated in a simulated working environment.

3 Results and Discussions

Figure 2a shows the Range of Motions (ROMs) computed in both NO EXO and EXO conditions, while performing the assigned task. Moreover, Fig. 2b reports the iEMG for each muscle in the assigned conditions. In terms of kinematics, it is evident that the ROM during task execution are largely similar between the two conditions, with minor discrepancies due to normal physiological fluctuations. Moreover, the adopted exoskeleton leads to a reduction in muscle activation under the EXO condition. This result is also confirmed with the user's subjective experience while wearing the glove. The obtained findings in both kinematic and muscular assessment pushes for further exploration, involving a larger sample of participants and a broader range of tasks to investigate its potential positive effects.



(a)



(b)

Fig. 2. (a) The RoM of MCP, PIP, and DIP joints of Index Finger (IF) and Middle Finger (MF) in both NO EXO ed EXO conditions. (b) The iEMG of EDC, FDS, FCU, ECU and APB muscles in both NO EXO ed EXO conditions.

4 Conclusions

The adoption of active exoskeletons can offer substantial advantages to workers in the manufacturing contexts. The initial assessment of both kinematic and muscular aspects conducted in this study has yielded interesting findings. Specifically, it has been observed that the same Range of Motion (ROM) can be achieved also with the exoskeleton, along with a reduction in muscular activation. These promising findings highlight the need for further investigation into the ergonomic benefits of these devices. This should involve a larger group of volunteers and an assessment of potential secondary effects, such as work cycle duration and the incidence of muscle fatigue or wrist/hand issues among workers. While these exoskeletons offer comprehensive solutions addressing safety, compliance, cost-effectiveness, and ongoing support, it is crucial to evaluate their usability and acceptance among end users.

References

1. Govaerts, R.: Prevalence and incidence of work-related musculoskeletal disorders in secondary industries of 21st century Europe: a systematic review and meta-analysis. BMC Musculoskelet. Disord. **22**(1), 1–30 (2021)
2. Voilqué, A.: Industrial exoskeleton technology: classification, structural analysis, and structural complexity indicator. In: 2019 Wearable Robotics Association Conference (WearRAcon), pp. 13–20. IEEE (2019)
3. Du Plessis, T.: A review of active hand exoskeletons for rehabilitation and assistance. Robotics **10**(1), 40 (2021)
4. Diftler, M.: RoboGlove-a robonaut derived multipurpose assistive device. In: International Conference on Robotics and Automation, no. JSC-CN-30211 (2014)
5. IronHand, BioServo. https://www.bioservo.com/press-releases/ironhand-2-0-the-new-version-of-award-winning-exoskeleton-glove. Accessed 27 Oct 2023

A Functional Approach for High-Velocity Walk-Through Programming

Matteo Ragaglia[1(✉)], Mattia Bertuletti[1], Simone Di Napoli[1], Mattia Gambazza[1], Cesare Fantuzzi[2], and Federica Ferraguti[2]

[1] Gaiotto Automation S.p.a (SACMI Group), Piacenza, Italy
matteo.ragaglia@sacmigroup.com
[2] University of Modena and Reggio Emilia, DISMI, Reggio Emilia, Italy

Abstract. In recent years, robots have emerged as indispensable tools for improving the competitiveness and adaptability of small and medium sized enterprises. Introducing robotics into the companies brings to well established challenges, including the intricate and time-intensive process of robot programming. To overcome these obstacles, this paper introduces a novel control architecture for Walk-Through Programming, specifically crafted to program industrial manipulators at high speeds and facilitate seamless integration with any generic closed-loop robotic controller.

Keywords: Physical Human-Robot Interaction · Cooperative Robotics · Admittance Control · Walk-Through Programming

1 Introduction

Traditionally, programming industrial robots was performed by highly specialized workers who write code using robot-specific programming languages and utilize teach pendants to guide the robot through desired waypoints. The complexity and time-consuming nature of this process significantly hindered the adoption of industrial robots by companies requiring flexibility to adapt to limited production volumes and rapidly changing product demands. A strategy that addresses these limitations is the so-called "walk-through programming" (WTP) Ragaglia et al. (2016), where a human operator manually guides the robot's end-effector to the desired locations to teach the robot the trajectory. In this programming approach, the human operator is not required to know the robot's programming language. During the teaching phase, the robot controller records the operator's movements, allowing the manipulator to replicate the motion subsequently. This type of interaction is enabled through the implementation of interaction control strategies such as admittance or impedance control, utilizing

This work is partly supported by Regione Emilia-Romagna under the agreement "Innovazione della value proposition di Gaiotto Automation 2021–2025" (grant number/CUP: E32C21001090009).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 8–13, 2024.
https://doi.org/10.1007/978-3-031-76428-8_2

a Force/Torque (F/T) sensor to measure the forces exerted by the operator. These measurements are then processed and converted into displacements in the operational space. The standard implementation of WTP often presents well-documented challenges, such as limited reachable velocities during the teaching phase, the need to compensate for dynamic effects introduced by the tool Talignani Landi et al. (2016), potential instability due to the operator's variable stiffness Ferraguti et al. (2019), and the requirement for an "open" controller to directly command joint positions (and possibly velocities) to the axis low-level control loops Ferraguti et al. (2017).

To overcome these issues, novel approaches have been proposed in Ferraguti et al. (2023), di Napoli et al. (2023). The novel WTP architecture enables the recording of continuous trajectories at high velocities, accurately compensates for tool dynamics, detects and avoids instability in the system's dynamic response, and ensures straightforward integration with a generic closed robotic controller. This proposed solution has been validated on an industrial manipulator.

2 Functional Logic Design

The proposed control architecture is illustrated in Fig. 1. During the teaching phase the working tool (e.g. the spray gun for a painting robot) is directly mounted on the robot end-effector. The architecture consists of:

- A tool compensation algorithm that accounts for the tool's influence. The algorithm leverages the forces measured from the F/T sensor F_{ext} to separate the component due to the tool F_{nc} from the component due to the human interaction F_c.
- An admittance controller based on the law (1). The controller generates the desired Cartesian pose x_{ref} that the robot must follow through the low-level position controller.
- An algorithm for detecting rising oscillations and adapting the interactive behavior of the robot. Indeed, potential instabilities may arise from human-robot interaction. The algorithm detect and mitigate the oscillations in the robot's behavior by adjusting the admittance control parameters (inertia $M(t)$ and damping $D(t)$) in order to guarantee stability.

The output of the control architecture is used to move a Dynamic Target Reference Frame (DTRF), which serves as the Cartesian position reference for the robot's Tool Center Point (TCP). The default trajectory interpolation algorithm calculates the corresponding joint-space reference, which is then sent to the low-level position controller. A Singularity Avoidance block is incorporated to prevent the robot from reaching singular configurations that could lead to errors in the interpolation algorithm.

$$F_{ext} = F_{nc} + F_c \qquad M(t)\ddot{x}_{ref}(t) + D(t)\dot{x}_{ref}(t) = F_c(t) \tag{1}$$

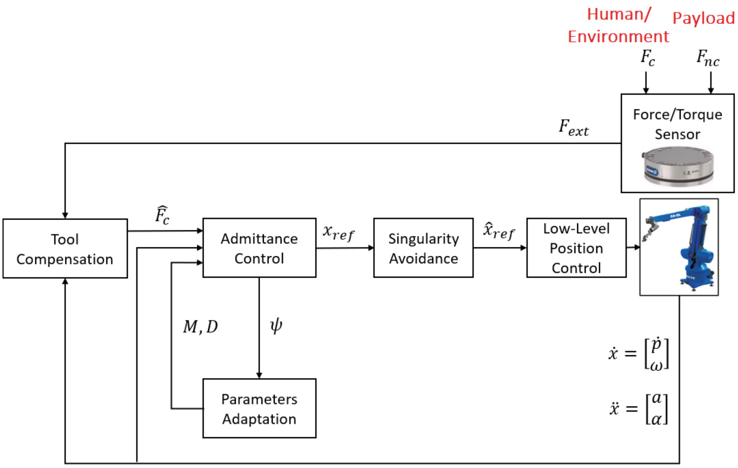


Fig. 1. Overall control architecture.

3 Implementation

The proposed control architecture was implemented on the Gaiotto GA-OL manipulator, a 6-DOF industrial robot designed for spraying applications. The software architecture consists of the following primary tasks. A **WTP** task that implements the walk-through programming strategy discussed earlier and generates the output pose x_{ref} that the robot must follow through the low-level position controller. To prevent the robot from reaching singular configurations, a **Singularity Avoidance** task ensures that the robot avoids the GA-OL manipulator's two distinct singularities, both characterized by significant values of the fifth joint $q_5 = \pm\frac{\pi}{2}$. This can be effectively achieved by imposing the following position limits on the fifth joint: $q_5 \in \left(-\frac{\pi}{2}, +\frac{\pi}{2}\right)$. Finally, a **DTRF** task initializes the Dynamic Target Reference Frame (DTRF) with the Cartesian coordinates of the robot's Tool Center Point (TCP) and synchronizes the DTRF with the pose output of the admittance controller, ensuring the robot follows the desired trajectory.

4 Experimental Results

An extensive experimental evaluation was conducted at Gaiotto SpA to: i) verify the effectiveness of the proposed control architecture in adapting admittance parameters; ii) assess the performance of the implementation and compare it to a commercially available solution featuring mechanical compensation systems for passive WTP; iii) evaluate the usability of the proposed architecture in real industrial scenarios. The experimental campaign involved several employees and was structured around four different experiments. Specifically: i) executing a linear trajectory with a fixed orientation (see Fig. 2(a) and 2(d)); ii) executing the same linear trajectory with a variable orientation (see Fig. 2(e)); iii) simulating the glazing of a floor-mounted flush toilet (see Fig. 2(b)); iv) simulating the glazing of a wall-hung flush toilet (see Fig. 2(c)). From a verification perspective, Figs. 3 and 3(a) demonstrate that whenever instability is detected, the algorithm promptly increases the linear inertia value to stabilize the system. A forgetting factor is implemented to gradually decrease the inertia after the desired behavior is restored. To compare the proposed solution to WTP approaches based on mechanical solutions, a benchmark was conducted on 40 different programs recorded using a back-drivable and mechanically compensated manipulator. Figure 4 shows the comparison between this benchmark and the approximated distribution of maximum linear TCP speed values recorded during the experimental validation campaign. In conclusion, the proposed solution can deliver a performance level comparable to the much more expensive existing mechanical solutions.

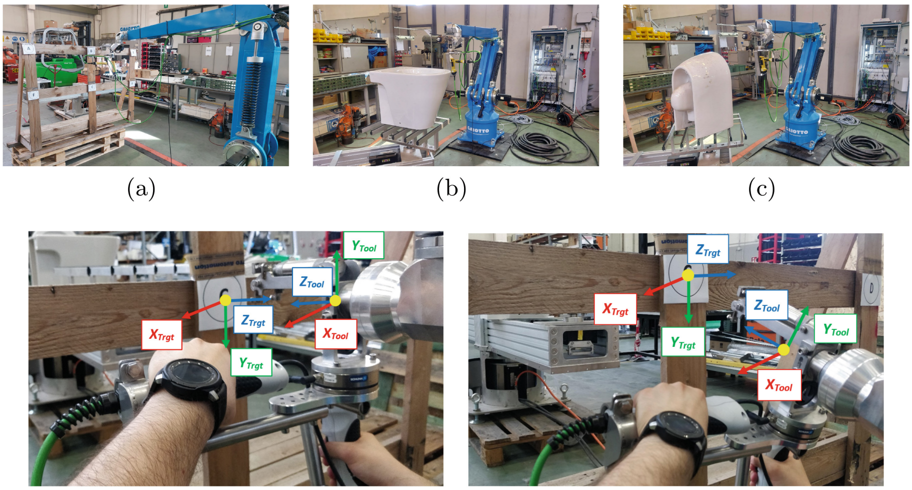
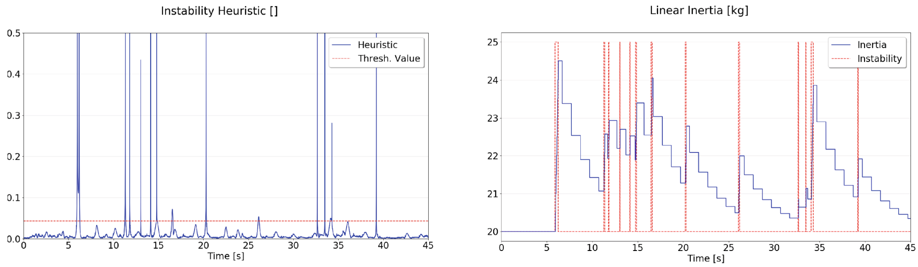


Fig. 2. 2(a) Experimental setup: linear trajectory with enumerated targets “A” to “F”; 2(b) floor-mounted toilet; 2(c) wall-hung toilet; 2(d) orientation of labeled targets during the first experiment; 2(e) orientation during the second experiment.



(a) (b)

Fig. 3. Adaptation of admittance parameters in presence of detected oscillations.

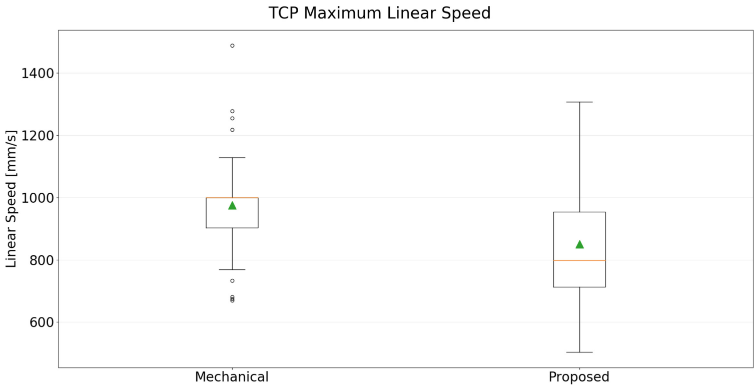


Fig. 4. Maximum TCP linear speed reached during teaching. Traditional solution benchmark (left), versus proposed solution (right).

References

Ferraguti, F., Talignani Landi, C., Secchi, C., Fantuzzi, C., Nolli, M., Pesamosca, M.: Walk-through programming for industrial applications. Proc. Manuf. **11**, 31–38 (2017)

Ferraguti, F., Landi, C.T., Sabattini, L., Bonfè, M., Fantuzzi, C., Secchi, C.: A variable admittance control strategy for stable physical human-robot interaction. Int. J. Robot. Res. **38**(6), 747–765 (2019)

Ferraguti, F., Bertuletti, M., Gambazza, M., Ragaglia, M.: High-velocity walk-through programming for industrial applications. Robot. Comput.-Integr. Manuf. **81**(102), 505 (2023)

di Napoli, S., Bertuletti, M., Gambazza, M., Ragaglia, M., Fantuzzi, C., Ferraguti, F.: High-velocity walk-through programming for industrial applications: a safety-oriented approach. In: International Conference on Informatics in Control, Automation and Robotics (ICINCO) (2023)

Ragaglia, M., Zanchettin, A.M., Bascetta, L., Rocco, P.: Accurate sensorless lead-through programming for lightweight robots in structured environments. Robot. Comput.-Integr. Manuf. (RCIM) **39**, 9–21 (2016)

Landi, C.T., Ferraguti, F., Secchi, C., Fantuzzi, C.: Tool compensation in walk-through programming for admittance-controlled robots. In: IEEE Annual Conference of the Industrial Electronics Society (IECON), pp. 5335–5340 (2016)

A Safety-Oriented Controller for High-Velocity Walk-Through Programming

Matteo Ragaglia[1(✉)], Simone Di Napoli[1], Mattia Bertuletti[1], Mattia Gambazza[1], Cesare Fantuzzi[2], and Federica Ferraguti[2]

[1] Gaiotto Automation S.p.a (SACMI Group), Piacenza, Italy
matteo.ragaglia@sacmigroup.com

[2] University of Modena and Reggio Emilia, DISMI, Reggio Emilia, Italy

Abstract. Despite the potential benefits for small-to-medium-sized companies, the widespread adoption of robots is hampered by the complexities of traditional programming methods. To address these challenges, this paper presents a safety architecture for walk-through programming of industrial manipulators. This approach aims to enable high-speed robot operations while prioritizing operator safety.

Keywords: Physical Human-Robot Interaction · Cooperative Robotics · Admittance Control · Walk-Through Programming

1 Introduction

Traditional robot programming methods, requiring specialized skills and time-consuming processes, prevent the widespread adoption of industrial manipulators, since small-to-medium sized companies require high flexibility to manage limited production volumes and rapidly changing product specifications. To address this, "walk-through programming" approaches have been proposed Bascetta et al. (2013). These strategies utilize sensor systems to measure forces and torques exerted by the human during the interaction with the robot and admittance control algorithms to ensure appropriate robot responses. However, safety standards, such as the 250 mm/s limit on the TCP Cartesian velocity limit for Human-Robot Interaction (HRI), can be too restrictive for recording high-velocity continuous trajectories. While this limitation is suitable for applications where only via-points need to be memorized Ragaglia et al. (2016), it may prevent the use of walk-through programming in applications where the recording of high-velocity continuous trajectories is required (e.g. spraying applications). This paper introduces a novel safety control architecture that combines traditional

This work is partly supported by Regione Emilia-Romagna under the agreement "Innovazione della value proposition di Gaiotto Automation 2021–2025" (grant number/CUP: E32C21001090009).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 14–18, 2024.
https://doi.org/10.1007/978-3-031-76428-8_3

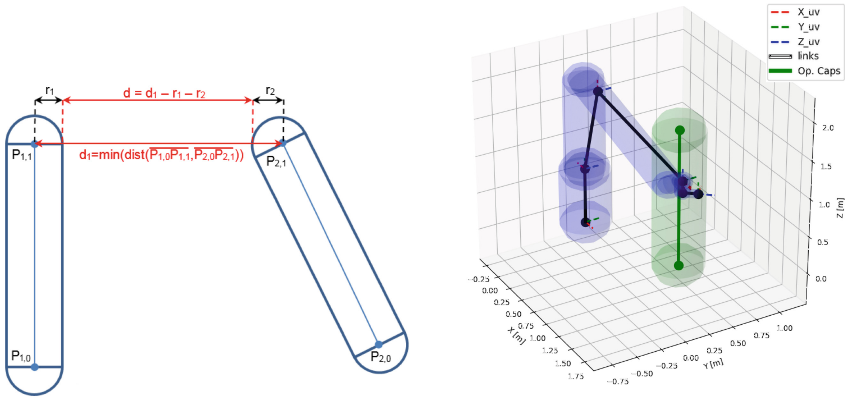
safety checks with advanced monitoring functionalities to enable safe recording of high-velocity trajectories during walk-through programming. For a comprehensive explanation of the proposed solution, refer to the detailed descriptions in the cited works Ferraguti et al. (2023), di Napoli et al. (2023).

2 Safety Logic Design

The proposed safety controller employs a two-fold approach. It first defines **Basic Safety Functions** as upper bounds for dynamic quantities in both joint and Cartesian spaces. Then, **Dynamic Human-Robot Monitoring** functionalities are introduced to continuously track the relative distance and velocity between the robot and the human operator. In particular, for basic safety functions, three cases are considered: i) maximum allowed Cartesian Linear Speed at TCP ($\dot{x}_p^{UB}$); ii) maximum allowed Joint Velocities ($\dot{q}_k^{UB}$); iii) maximum allowed Joint Accelerations ($\ddot{q}_k^{UB}$). The threshold for $\dot{x}_p^{UB}$ can be determined based on task requirements. By considering a reference configuration of the manipulator ($\overline{q}$) and its Jacobian matrix ($J(\overline{q})$), the angular velocity value ($\dot{q}_k^{UB}$) for each joint corresponding to $\dot{x}_p^{UB}$ can be determined. Subsequently, $\ddot{q}_k^{UB}$ can be computed using the controller's cycle time (Δt).

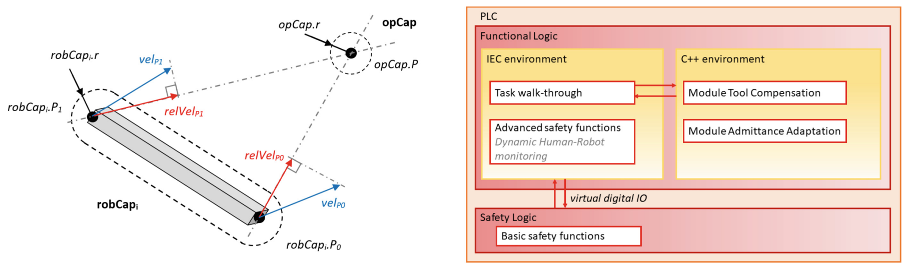
$$\dot{q}_k^{UB} \leftarrow \min\left(\dot{q}_k^{Max}, \frac{\dot{x}_p^{UB}}{\sqrt{\sum_{w=1}^{3} \left(J\left(\overline{q}\right)\right)_{w,k}^{2}}}\right) \qquad \ddot{q}_k^{UB} \leftarrow \min\left(\ddot{q}_k^{Max}, \frac{\dot{q}_k^{UB}}{\Delta t}\right) \tag{1}$$

Beyond the basic safety functions, the proposed safety controller monitors the real-time relative distance and velocity between the robot and the human operator. Building upon previous work in safe Human-Robot Interaction (HRI) Ragaglia et al. (2015), Ragaglia et al. (2018), capsule-based geometry models are used to represent the space occupancy of both the robot and the human operator Fig. 1(b). During the teaching phase, the safety controller updates capsule positions, calculates relative distances between robot and operator capsules Fig. 1(a), and ensures that the minimum relative distance do not exceeds a predefined threshold. By setting this threshold based on the robot's maximum achievable distance at $\dot{x}_p^{UB}$ within the emergency stop time (Δt), the safety controller effectively prevents collisions between the robot and the operator.
Several studies in safe HRI Ragaglia et al. (2014), Ferraguti et al. (2020) have explored the use of relative human-robot velocity and distance for safety. This work monitors relative human-robot velocity to identify situations where an emergency stop may not prevent a collision. A suitable threshold value is the maximum allowed linear speed ($\dot{x}_p^{UB}$). The maximum relative velocity is calculated between the endpoints of the robot capsules and the operator's capsule Fig. 2(a). The final output of the dynamic human-robot monitoring algorithm is the logical conjunction of the conditions related to relative distance and relative velocity.



(a) (b)

Fig. 1. 1(a) Representation of two separate capsules and of the minimum distance between them; 1(b) example of capsules modelling an operator and a 6 DoF manipulator.



(a) (b)

Fig. 2. 2(a) Representation of the relative velocity between the end-points of a generic robot link (solid grey bar) and the operator's capsule; 2(b) General control architecture.

3 Implementation

The proposed control architecture was implemented on the Gaiotto GA-OL manipulator, a 6-DOF industrial robot designed for spraying applications. The safety logic is integrated between the safety and functional environments, as illustrated in Fig. 2(b). Basic safety functions are implemented within the "Safety Logic" domain, while dynamic human-robot monitoring functionalities are executed within the "Functional Logic" domain. These domains communicate through boolean variables exchanged via shared memory. The Safety Logic gathers relevant information and can deactivate the Safe Torque Off (STO) signal, cutting off electric power to the actuators.

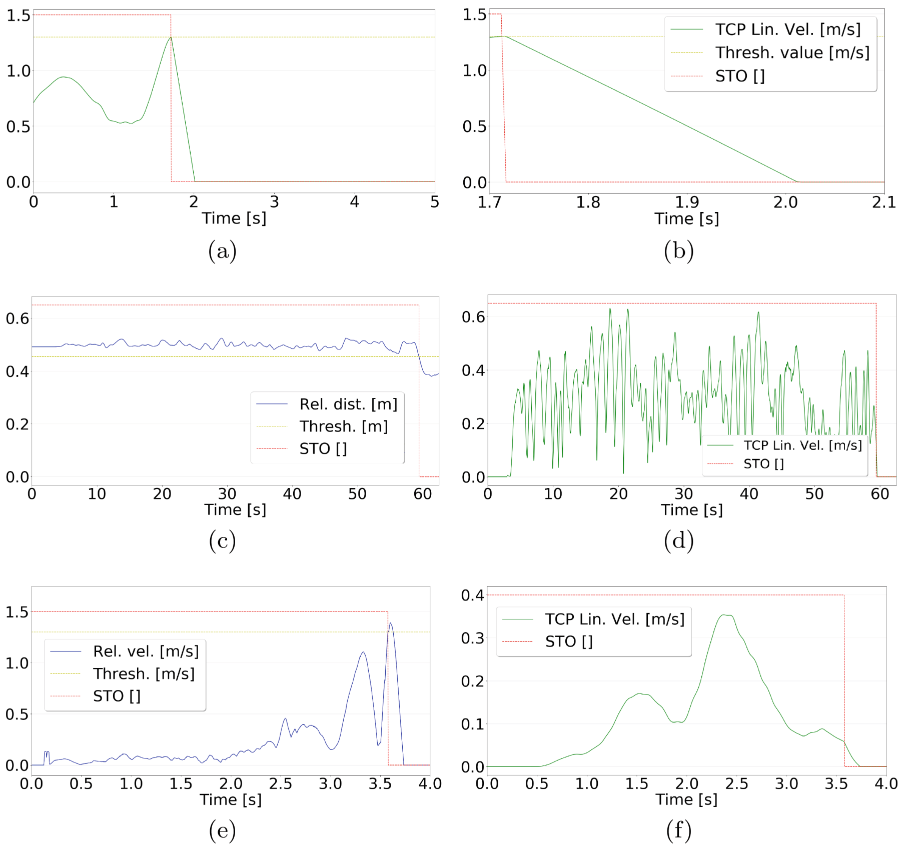


Fig. 3. 3(a)–3(b) Violation of TCP linear velocity limit and corresponding variation of scaled STO signal, with corresponding magnification; 3(c)–3(d) violation of the relative human-robot minimum distance with TCP linear velocity and corresponding variation of scaled STO signal, with corresponding magnification; 3(e)–3(f) violation of the relative human-robot maximum velocity with TCP linear velocity and corresponding variation of scaled STO signal, with corresponding magnification.

4 Experimental Results

An experimental campaign was conducted to demonstrate the effectiveness of the proposed architecture, and the results are presented in this section. First, Figs. 3(a)–3(b) show that when the TCP linear velocity limit ($\dot{x}_p^{UB} = 1.30\,\text{m/s}$) is exceeded, the STO signal (properly scaled) is disabled, and the robot stops within the declared stopping time of 0.35 s. Regarding violations of the dynamic human-robot monitoring conditions, Figs. 3(c)–3(d) (3(e)–3(f)) show the deactivation of the STO signal when there is a violation of the minimum relative distance (maximum relative velocity) condition.

References

Bascetta, L., Ferretti, G., Magnani, G., Rocco, P.: Walk-through programming for robotic manipulators based on admittance control. Robotica (ROB) **31**, 1143–1153 (2013)

Ferraguti, F., et al.: Safety barrier functions and multi-camera tracking for human-robot shared environment. Robot. Auton. Syst. **124**(103), 388 (2020)

Ferraguti, F., Bertuletti, M., Gambazza, M., Ragaglia, M.: High-velocity walk-through programming for industrial applications. Robot. Comput.-Integr. Manuf. **81**(102), 505 (2023)

di Napoli, S., Bertuletti, M., Gambazza, M., Ragaglia, M., Fantuzzi, C., Ferraguti, F.: High-velocity walk-through programming for industrial applications: a safety-oriented approach. In: International Conference on Informatics in Control, Automation and Robotics (ICINCO) (2023)

Ragaglia, M., Bascetta, L., Rocco, P., Zanchettin, A.M.: Integration of perception, control and injury knowledge for safe human-robot interaction. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 1196–1202 (2014)

Ragaglia, M., Zanchettin, A.M., Rocco, P.: Safety-aware trajectory scaling for human-robot collaboration with prediction of human occupancy. In: International Conference on Advanced Robotics (ICAR), pp. 85–90 (2015)

Ragaglia, M., Zanchettin, A.M., Bascetta, L., Rocco, P.: Accurate sensorless lead-through programming for lightweight robots in structured environments. Robot. Comput.-Integr. Manuf. (RCIM) **39**, 9–21 (2016)

Ragaglia, M., Zanchettin, A.M., Rocco, P.: Trajectory generation algorithm for safe human-robot collaboration based on multiple depth sensor measurements. Mechatronics (MECH) **55**, 267–281 (2018)

Audio-Based Analysis of Child-Robot Interactions in the Wild

Xela Indurkhya[1](✉) and Gentiane Venture[2]

[1] Tokyo University of Agriculture and Technology, Tokyo, Japan
xelaindurkhya@gmail.com
[2] University of Tokyo, Tokyo, Japan

Abstract. We describe 2 in-the-wild child-robot interaction English-language group experiments conducted in India, using the social robot Vector. We analyze the children's engagement with the robot for the duration of each interaction by analyzing their vocal behavior using 3 approaches: the gendering of the robot by the children, a behavioral categorization system, and a qualitative breakdown of vocal behaviors observed. We highlight the potential of audio- and language-based behavioral analysis for human-robot interaction.

Keywords: child-robot interaction · in-the-wild study · behavioral study · vocalization behavior · human-robot interaction · behavioral coding

1 Introduction

In-the-wild interaction studies provide valuable insights into real-world behavior of people when interacting with robots outside of controlled environments [5]. There have been several such studies conducted across a variety of different cultures, using a variety of methods, from visual behavioral analysis to post-interaction interview-based approaches [3, 5, 7]. A constraint often faced is logistical and ethical difficulties filming in public settings, which may be restricted or not possible at all [2]. Audio-based approaches are under-utilized in such scenarios. Some existing studies contain elements of vocal-behavior analyses, alongside body language and facial expression [5, 7]. We posit that just as visual behaviors can be analyzed without reliance on audio elements, audio behaviors can be analyzed without reliance on visual elements. While it is ideal to have access to both audio and visual elements for analysis, there are circumstances in which this is not possible, and thus it is important to highlight the possibilities of audio analysis so as to conduct more versatile and robust behavioral analyses [1].

2 Methods

Two in-the-wild interaction experiments were conducted in India in May 2019, in Bhopal and Hyderabad. Local children aged 4–8 were recruited by word of mouth. Parents were also present. In Bhopal, there were 4 children (2m, 2f), all in the target age range, and 4 adults present; in Hyderabad, there were 9 children in the target age range (6m, 3f), 1 teen, and 4 adults. We used the social robot Vector by Anki, Inc.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 19–23, 2024.
https://doi.org/10.1007/978-3-031-76428-8_4

Each interaction consisted of 2 stages: a remote-controlled interaction stage (RC), and a factory mode stage (FM). In RC, the robot was set on a table in front of the children and they interacted with it freely while the researcher controlled the robot, responding to the children's questions with responding questions. In FM, the children were shown some of Vector's basic functions: asking questions, requesting fistbumps, playing with its cube, and taking photos, as well as changing the robot's eye color through the app. In Bhopal, RC lasted 17 min, FM 28 min. In Hyderabad, RC lasted 19 min, FM 13 min. Hyderabad ended earlier due to the robot running out of battery. Each experiment was filmed by an observing adult, prioritizing audio over video.

2.1 Vocal Behavior Categorization

We applied a variation on a vocalization-based behavioral categorization that we have used previously, based on behavioral analysis methods applied in wildlife biology [4]. This method, referred to as behavioral coding, aims to broadly categorize 3–4 robot-related behaviors in children, which indicate different modes of engagement, as distinct from non-robot-related behaviors, which indicate disengagement (unpublished). We defined the categories in Table 1 and noted instances where we observed these behaviors throughout the interaction.

To make the categorization consistent, when several children were talking over each other, we only categorized the voices heard distinctly; when 2 or more voices were heard making simultaneous comments that belonged to the same category, no more than 2 instances of that behavior were noted. Only the vocal behaviors of child speakers were categorized. When a sequence of commands was required to execute one command to the robot, these were all categorized as a single "command" behavior unless someone else spoke between the sequence, in which case they would be counted separately.

Functional commands (to which the robot is programmed to respond) and language that we socially recognize as commands are sorted into the command category, while all other instances of children speaking directly to the robot were categorized as non-command. Where the children followed the format of a functional command, but neither used the phrasing to activate the robot (such as altering the "Hey Vector" into "Hey Matar" or "Vector?") nor phrased it as a social command, these were counted as non-command behavior.

Table 1. Definitions of the vocal behavior categories applied.

Behavior Category	Definition
To (command)	Directly addressing the robot; social/functional command
To (non-command)	Directly addressing the robot; not a command
Responsive	Direct responses to the robot (verbal responses only)
About	Directly speaking about the robot
Other	Non-robot-related conversation/comment

3 Results

3.1 Gendering the Robot

In Bhopal, Vector was consistently referred to with he/him pronouns, with only one instance of a child using "she". In Hyderabad, he/him and it/its pronouns were used. Despite the relatively consistent use of he/him in both the Bhopal and Hyderabad interactions, in both locations, during the RC stage, the children asked whether Vector was a boy or a girl. Vector would respond by asking the children what the difference was. In Bhopal, a child joked that "a boy is a girl who burps", to which Vector responded, "I don't burp, so I'm not a boy." Initially only hearing "I am not a boy," the children discussed that "he is a girl", continuing to use he/him pronouns. In Hyderabad, the children briefly debated the validity of asking a robot's gender amongst themselves and, having seemingly decided it was not important, instead began to ask Vector what it was, whether it was a robot, and whether it could fly.

In both Bhopal and Hyderabad, an adult used or suggested she/her pronouns for the robot, but the children did not appear swayed by this. The single "she" used by a child in the Bhopal experiment took place several minutes later in the interaction. Another child immediately corrected this back to he/him.

3.2 Behavior Categorization

The breakdown of the behaviors observed during each interaction is displayed in Fig. 1 as percentages of the total number of behaviors categorized in each interaction stage.

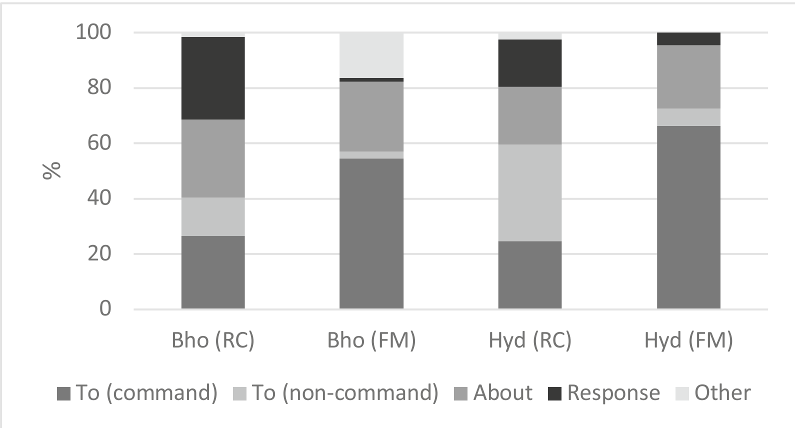


Fig. 1. Behaviors observed during each interaction, broken down by percentage of total behaviors categorized. Percentage is calculated out of a total of 121 behaviors categorized in Bhopal (RC), 226 in Bhopal (FM), 317 in Hyderabad (RC), and 175 in Hyderabad (RC).

Responses were most common during both RC stages; commands made up more than double the percentage of behaviors during FM stages. Non-robot-related behavior was most frequently seen in Bhopal's FM stage, where 2 of the 4 children disengaged from the robot in the last 12 min, returning for only 2 min toward the end to watch the robot play with its cube. Bhopal's RC stage saw the highest proportion of overall behaviors talking about the robot, while Hyderabad's RC stage saw the highest proportion of overall behaviors talking to the robot without commanding it. These results show a high level of engagement in all interactions, with distinctly different patterns between RC and FM stages, as well as a different style of interaction between the two groups in the RC stage.

3.3 Common Vocal Behaviors

Some common behaviors observed across multiple interaction stages are described broadly in Table 2. While the frequency of "Come here" could be attributed to the phrase being a functional command, the children used it most often during the RC stage of both interactions before learning to use it as a functional command in the FM stage.

Table 2. Number of occurrences of common robot-related vocal behaviors in each interaction. Bho = Bhopal, Hyd = Hyderabad

Vocal Behavior	Bho(RC)	Bho(FM)	Hyd(RC)	Hyd(FM)
"Come here"	11	9	31	5
Remarks on robot's apparent emotion	4	0	19	0
Giving the robot directions	1	0	10	14
Giving the robot random objects	5	0	0	1
Complimenting the robot	2	3	0	1

The robot's emotions were first mentioned in both interactions when the children perceived anger (remarking "he is angry" in Bhopal, and asking, "are you angry?" in Hyderabad), both interactions saw the children subsequently remark on the robot's happiness, with the children in Hyderabad going on to ask if it was sad. Consistently in Bhopal, the children remarked on these emotions in the third person, while in Hyderabad, the children directed the observations at the robot as a question. The Hyderabad group was also frequently heard giving the robot directions ("go straight", "turn around", "that way", etc.). The Bhopal group attempted to offer the robot a variety of objects (a card, a ring, and cookies), while the Hyderabad group attempted to offer it a toy ball. The Bhopal group complimented the robot as excellent and cute, while the Hyderabad group complimented it as intelligent.

In Bhopal's FM stage, the children had a heated discussion regarding what color the eyes of the robot ought to have. Throughout the RC stage, the eyes had been brown, and they commented that they were the color of chocolate. On learning that the colors could be changed, they were eager to do so, but the disagreement over the color of preference became very heated. One child threw a small tantrum and refused to interact with the robot for a short time even while being coaxed by a parent. The frustration was exacerbated when the robot would rarely respond to the children's verbal cues without assistance from a parent, which was also a difficulty in the Hyderabad interaction, though the children were less vocal about their frustration.

In the RC stage in Hyderabad, one of the children repeatedly mispronounced Vector's name as "matar" (green peas), and as a result, some of the other children repeatedly referred to Vector as "matar" through both RC and FM interactions (21 times in RC, 6 times in FM). Several minutes into RC, a different child threatened the robot, "If you don't tell us [your name], we'll turn you into a chutney". Following this, several children asked the robot "Are you a chutney?" 6 times through the rest of RC.

4 Discussion

Here we have showcased 3 approaches to audio-based analysis. A few other possible approaches include looking at responses to unexpected actions by the robot, analyzing non-verbal elements such as laughter [7]. Language is a crucial part of society and has great untapped potential in human-robot interaction analysis, but requires attention to culture and nuances of the language or dialect being analyzed [1, 6]. An audio analysis need not entirely exclude visual elements, as researchers could record certain visual elements during the experiments, such as how many children touched or avoided the robot. As an increasing number of robot interaction studies incorporate behavior based on audio [6, 7], this is a field that could greatly benefit from further study.

References

1. Boyd, R.L., Schwartz, H.A.: Natural language analysis and the psychology of verbal behavior: the past, present, and future states of the field. J. Lang. Soc. Psychol. **40**(1), 21–41 (2021)
2. Coronado, E., Indurkhya, X., Venture, G.: Robots meet children, development of semi-autonomous control systems for children-robot interaction in the wild. In: IEEE International Conference on Advanced Robotics and Mechatronics 2019, pp. 360–365 (2019)
3. Nomura, T., et al.: Why do children abuse robots? In: ACM/IEEE International Conference on Human-Robot Interaction 2015, pp. 63–64 (2015)
4. Mann, J.: Behavioral sampling methods for cetaceans: a review and a critique. Mar. Mamm. Sci. **15**(1), 102–122 (1999)
5. Ros, R., et al.: Child-robot interaction in the wild: advice to the aspiring experimenter. In: Proceedings of the ACM International Conference on Multimodal Interaction, pp. 335–342 (2011)
6. Sienkiewicz, B., Sejnova, G., Gajewski, P., Vavrecka, M., Indurkhya, B.: How language of interaction affects the user perception of a robot. In: Proceedings of the 15th International Conference on Social Robotics (2023)
7. Wróbel, A., Źróbek, K., Schaper, M.-M., Zguda, P., Indurkhya, B.: Age-appropriate robot design: in-the-wild child-robot interaction studies of perseverance styles and robot's unexpected behavior. In: Proceedings of the IEEE Ro-MAN (2023)

Human-Aware Motion Planner for Collaborative Transportation of Flexible Materials

Alberto Gottardi[1,2(✉)], Enrico Pagello[1], Emanuele Menegatti[2], and Stefano Tonello[1]

[1] IT+Robotics srl, 36100 Vicenza, Italy
alberto.gottardi@it-robotics.it
[2] Intelligent Autonomous System Lab, Department of Information Engineering, University of Padova, 35131 Padua, Italy
emg@dei.unipd.it

Abstract. In industrial environments like factories and warehouses, transportation of flexible materials that need the collaboration of several subjects is a typical activity. One instance is the handling of enormous fibre sheets in the fabrication of composite parts, which presents several difficulties, including handling flexible materials and needing to place the material with extreme precision. Recently, there has been a lot of interest in employing robots to help human workers carry such things. However, this typically entails the robot adopting a follower attitude just intended for passive support without fully using its accuracy and repeatability. To make the best possible use of the robot's and the operator's skills, it is necessary to use an intelligent motion planner that takes into account the ergonomics of the operator but at the same time ensures the precision required by the task. In this paper, we present a preliminary study for a Human-Aware Motion Planner for the cooperative transportation of materials.

Keywords: Human-robot cooperation · human-aware motion planner · flexible material transportation · composite parts manufacturing

1 Introduction

One of the most critical activities in the textile and composite part manufacturing landscape is the Human-Robot cooperative transportation of flexible materials [5]. Large pieces are typically utilized to shape a component, and in current production lines, they are handled manually by two or more workers: they pick up the piece by hand off a table and bring it to the tooling station. The accurate placement of the cut-piece is a demanding requirement due to the high positioning requirements and the size of the piece. This results in non-ergonomic handling positions for the workers and high process times, motivating the introduction of a robotic partner to alleviate operator fatigue and improve efficiency. The robot has to act as a smart assistant, following the user while ensuring an ergonomic position and preventing material damage. At the same time, it must be able to take an active role by guiding the user when high accuracy is needed (for instance, to guarantee the proper fiber orientation).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 24–28, 2024.
https://doi.org/10.1007/978-3-031-76428-8_5

Fig. 1. An example of human-robot cooperative transport.

Many works in the literature focus on the co-transportation of flexible materials where the robot only follows the operator [2,10,12]. In situations when precise material placement is necessary, a pure follower robot cannot offer the greatest support to the operator since it can only respond to human actions, and it may be challenging and unpleasant for the human operator to manoeuvre the robot into the desired position. To be able to respond to this problem, it is necessary to use a high-level planner who can help the two agents accomplish the task. Recent works on this type of motion planner are oriented towards maintaining the ergonomics of the operator, but they only considered rigid materials [7,13] or the reconfiguration of the robot itself [3,11]. However, these approaches do not completely satisfy the requirements of flexible material transport in composite parts manufacturing. To overcome the limitations and improve these approaches for a more challenging scenario, we propose in this paper a preliminary study for a Human-Aware Motion Planner able to guarantee ergonomic transportation of flexible material and a final place precision. This challenge is addressed within the EU project DrapeBot (https://www.drapebot.eu/), which attempts to create an HRC system capable of aiding an operator working on carbon fibre draping. Figure 1 depicts an example of the human-robot cooperative transport to be achieved in the DrapeBot project.

2 System Concept

As described in the previous paragraph, the proposed method aims to go beyond the current approaches. The robot must not only passively support the user but also actively direct him to the desired position while avoiding collisions with obstacles in the environment and keeping to ergonomic boundaries for the operator. We propose a Human-Aware Motion Planner based on online and offline components to fulfil these demands. The offline module is responsible for computing an ergonomic path that serves as a high-level guideline for the robot and the operator. However well the calculated trajectory may satisfy ergonomic constraints, the operator may prefer not to follow this trajectory completely (as if he were a pure follower of the robot). For this reason, the

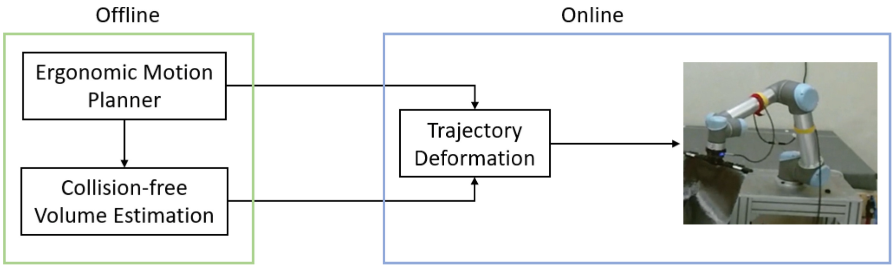


Fig. 2. Human-Aware Motion Planner overview.

online component is necessary: its task is to accommodate the deformations introduced by the user while ensuring the required accuracy in the place phase. However, deformations introduced by the operator can lead to collisions with environmental obstacles. This makes using a third module necessary to ensure the robot does not collide. A deformation volume estimation algorithm is therefore needed for this purpose. The Human-Aware Motion Planner uses information from an external vision system that perceives the environment and the user's movements and models them via a skeleton. Thanks to the skeleton, the ergonomic path (offline module) and the online deformation can be computed (Fig. 2).

2.1 Ergonomic Motion Planner

The planner must consider human constraints to determine the best possible trajectory. Only user arm position restrictions are considered in this first phase regarding ergonomic limitations. This suggests that the user's effort is considered for the transportation job. For example, when the user raises his arms over his head, he exerts more effort than when he raises his arms at hip level. Based on this idea, it is feasible to compute the user's height and the maximum and least accessible height of the arms using the skeleton tracker given by the vision camera system. Upper and lower achievable height limitations are produced based on this information. The limits are generated using values along the z-axis (the axis that points up). The outside zone is eliminated as non-ergonomic, while the inner zone is split further and connected with a cost function. As the limitations are approached, the cost increases linearly. The ultimate goal is reducing the cost function to achieve a low-effort path representing the best possible ergonomic path.

2.2 Collision-Free Volume Estimation

Given the nominal ergonomic path obtained by the motion planner, it is possible to forward it directly to the online deformation module. However, a collision-free region around the nominal trajectory is estimated to increase safety during deformation [8]. The collision-free volume is a part of the space where the robot changed its nominal path in reaction to human activities without colliding with the other items in the work-cell. The algorithm that computes the deformation boundaries has a linear computational complexity and will be entirely performed offline in the joint space, thus without occupying computational time for the online control during execution. The overall concept is to generate a volume which synthesizes the whole workspace but is left free in the space surrounding the nominal path and near the goal. Then, the boundaries defining the final intervals are found by observing for which joints the robot collides with such volume values (Fig. 3).

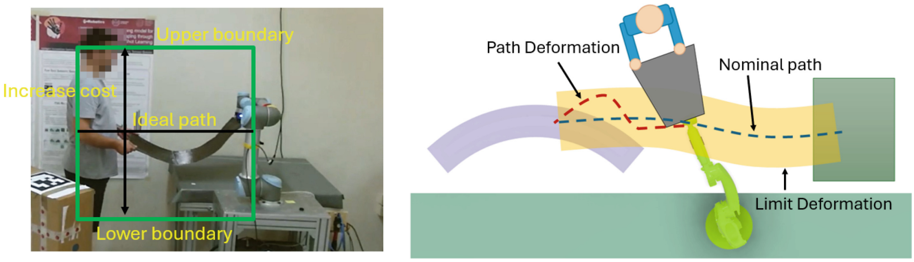


Fig. 3. Representation of the Ergonomic Motion Planner on the left and Deformation boundaries limits on the right.

2.3 Trajectory Deformation

Various approaches can be followed to deform the motion of a robot. [1,9] focused on how the material deforms during the motion of the robot and operator. Other techniques focused on the controller robot side like [4,6]. We believe that an approach based on Artificial Potential Fields (APFs) [6] is the best one to meet the needs of cooperative human-robot transportation. In addition, the combination of the offline approach with the APF enables the robot to move independently, adjusting to the dynamic environment and successfully reacting to unanticipated changes (i.e., moving obstacles) while assuring the operator's safety and high-level job precision and accuracy.

3 Conclusions

This paper presents a preliminary study about Human-Aware Motion Planner for the Human-Robot cooperative transportation of flexible materials. This approach is under development in a real industrial application like draping fiber carbon pieces within the EU project DrapeBot. Preliminary results are showing good results both in terms of offline module computation. The offline modules are evaluated by the planning time less than $500ms$ and a collision-free confidence factor equal to 92%. The confidence factor aims to express the quality of the boundaries produced. Concerning the online component of the framework, no quantitative results have yet been achieved to validate the approach.

Acknowledgment. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101006732, "DrapeBot - collaborative draping of carbon fiber parts".

References

1. Andronas, D., Kampourakis, E., Bakopoulou, K., Gkournelos, C., Angelakis, P., Makris, S.: Model-based robot control for human-robot flexible material co-manipulation. In: 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE (2021)
2. De Schepper, D., Moyaers, B., Schouterden, G., Kellens, K., Demeester, E.: Towards robust human-robot mobile co-manipulation for tasks involving the handling of non-rigid materials using sensor-fused force-torque, and skeleton tracking data. Proc. CIRP **97**, 325–330 (2021)
3. Ferraguti, F., Villa, R., Landi, C.T., Zanchettin, A.M., Rocco, P., Secchi, C.: A unified architecture for physical and ergonomic human-robot collaboration. Robotica **38**(4), 669–683 (2020)
4. Franceschi, P., Pedrocchi, N., Beschi, M.: Adaptive impedance controller for human-robot arbitration based on cooperative differential game theory. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 7881–7887. IEEE (2022)
5. Ghidoni, S., et al.: A smart workcell for human-robot cooperative assembly of carbon fiber parts. In: 3rd Italian Conference on Robotics and Machine Intelligence, I-RIM, Rome (2021)
6. Gottardi, A., Tortora, S., Tosello, E., Menegatti, E.: Shared control in robot teleoperation with improved potential fields. IEEE Trans. Hum.-Mach. Syst. **52**(3), 410–422 (2022)
7. Kim, W., Lee, J., Peternel, L., Tsagarakis, N., Ajoudani, A.: Anticipatory robot assistance for the prevention of human static joint overloading in human-robot collaboration. IEEE Robot. Autom. Lett. **3**(1), 68–75 (2017)
8. Miotto, N., Gottardi, A., Castaman, N., Menegatti, E., et al.: Collision-free volume estimation algorithm for robot motion deformation. In: Proceedings of IEEE 21st International Conference on Advanced Robotics (ICAR) (2023)
9. Nicola, G., Villagrossi, E., Pedrocchi, N.: Human-robot co-manipulation of soft materials: enable a robot manual guidance using a depth map feedback. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 498–504. IEEE (2022)
10. Nicola, G., Villagrossi, E., Pedrocchi, N.: Co-manipulation of soft-materials estimating deformation from depth images. Robot. Comput.-Integr. Manuf. **85**, 102630 (2024)
11. Peternel, L., Kim, W., Babič, J., Ajoudani, A.: Towards ergonomic control of human-robot co-manipulation and handover. In: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), pp. 55–60. IEEE (2017)
12. Sirintuna, D., Giammarino, A., Ajoudani, A.: An object deformation-agnostic framework for human–robot collaborative transportation. IEEE Trans. Autom. Sci. Eng. (2023)
13. van der Spaa, L., Gienger, M., Bates, T., Kober, J.: Predicting and optimizing ergonomics in physical human-robot cooperation tasks. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 1799–1805. IEEE (2020)

Enhancing Noise Robustness of Speech-Based Human-Robot Interaction in Industry

Stefano Bini(✉), Alessia Saggese, and Mario Vento

University of Salerno, 84084 Fisciano, SA, Italy
{sbini,asaggese,mvento}@unisa.it

Abstract. In the industrial environment, Human-Robot Interaction is a cornerstone for enhancing productivity and safety. This paper introduces a real-time speech-command recognition (SCR) system designed for running directly on board of embedded devices mounted on board of the robot. Our SCR, based on a ResNet architecture, employs dynamic and domain-specific data augmentation. Our approach significantly enhances accuracy in a noisy environment, achieving a +16.2 improvement over non-augmented training and a +3.6 gain over a non-dynamic approach. These advancements are obtained with low memory usage (1.76 MB) and good processing times (14.9 ms).

Keywords: speech-command recognition · domain-specific data augmentation · dynamic data augmentation · human-robot interaction

1 Introduction

Human-Robot Interaction plays a key role in industrial environments, addressing the growing need for enhanced productivity and safety in manufacturing. Within this context, virtual assistants have been proposed to integrate speech assistants into real manufacturing. However, challenges such as high computational demands and consistent noise levels in industrial settings prevent traditionally existing systems from obtaining good performance (Schmidt et al. 2018). Early attempts, like the one proposed by Mishra et al. (2015), used smartphones for remote voice commands, but faced concerns about latency and Bluetooth networking. Li et al. (2023) introduced Max, powered by BERT, via Wi-Fi, addressing some issues but still faced computational demands and latency. Conversely, Bingol and Aydogmus (2020) connected a high-performance laptop to the robot to reduce latency, exploring Convolutional Networks as an option.

These approaches often rely on complex Automatic Speech Recognition and Natural Language Understanding architectures, suitable for social robotics but

This work has received funding from the EU Horizon 2020 programme under GA No. 101017151 FELICE.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 29–33, 2024.
https://doi.org/10.1007/978-3-031-76428-8_6

ill-suited for industrial scenarios where interactions mainly involve short commands (Aydogmus et al. 2023). They also face challenges in noisy industrial environments, introducing communication latency and privacy concerns due to server dependencies (Schmidt et al. 2018). In response, a command-based approach, namely a Speech-Command Recognition (SCR), emerged: the idea is to have an end-to-end approach, classifying audio chunks directly into predefined commands (Aydogmus et al. 2023). In order to face the noise, Speech Enhancement modules are often used. Anyway, they risk to introduce artifacts and add computational burden (Greco et al. 2021). Data augmentation techniques could be thus a less invasive but still effective solution. In speech applications, noise addition is the commonly employed technique. Typically, noise samples from a specific domain are gathered and statically applied to the audio dataset. This static application implies that the model encounters the same noisy data throughout the entire training process. A notable concern is the risk of overfitting, stemming from the model becoming overly specialized on specific types of noise in particular positions within the audio samples (Park et al. 2019). This risk is particularly pronounced when the speech and noise datasets have limited size.

Our research addresses these challenges by introducing an advanced command-based robotic speech assistant, designed for noisy industrial settings. Operating in real-time without external server dependencies, it employs a ResNet architecture for SCR. A novel approach based on Dynamic Data Augmentation is proposed: the idea is to introduce industrial noise during training in a dynamic way, by modifying noise types and intensities to increase versatility and then improve the robustness of the system to the noise.

2 The Proposed Method and Architecture

The proposed system, based on a ROS architecture, is shown in Fig. 1. A dedicated microphone (a ReSpeaker Mic Array V1.0) for speech acquisition and an embedded device (an NVIDIA Jetson Xavier NX) for on-robot deployment have been integrated inside the robot. The ROS architecture includes four nodes, combining synchronous and asynchronous operations. Low-level communication uses the asynchronous mechanism of Publisher-Subscriber: (1) *acquisition_node* captures audio from a microphone and publishes it on the *audio_data* topic; and (2) *vad_node* subscribes to *audio_data* and employs a WebRTC Voice Activity Detection (VAD) (wiseman 2016) for speech detection. Higher-level communication uses two services: (3) *manager_service* on the *manager_node* processes newly detected speech and enhances independence by separating acquisition and classification components. (4) *scr_service* on the *scr_node* classifies voice data from the *manager_node* using the proposed SCR system.

In more detail, our SCR takes as input raw waveform data and exploits an intermediate Mel spectrogram-based representation; the choice is justified by the fact that it mimics the human auditory system, is robust to noise, and reduces spectral data dimensionality (Gold et al. 2011). The SCR returns as output the detected command, thus can be formulated as a classification problem. Let be D

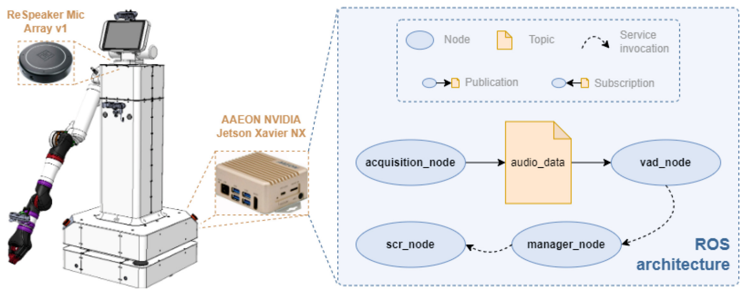


Fig. 1. Architecture of the proposed ROS-based SCR, integrated as a speech assistant in the robot developed within the FELICE H2020 project (FELICE Project 2022).

the audio dataset including pairs of audio samples (x_i) with the corresponding labels (y_i), and let be $r_i \in R$ the set of background noises. To increase noise robustness, we dynamically generate the sample data s_i to be used for feeding the network by adding to the original sample x_i the noise r_i. In this way, the generic x_i can be combined with multiple noise types and with different Signal-to-Noise-Ratio (SNR). To control the SNR variation, we introduce a gain function ($gain(\cdot)$) inversely proportional to the SNR (snr), applied to the noise, resulting in $s_i = x_i + gain(snr) \cdot r_i$. In each training step, the algorithm randomly selects the SNR, noise sample, and, if needed (when noise duration exceeds speech duration), a subsection of noise to apply to the speech sample. This dynamic process enhances data variability, contributing to a more diverse and robust training dataset and improving the model's generalization capabilities in handling noisy conditions. The classifier we employ is based on a ResNet-8.

3 Experimentation

In our experimentation, we used two datasets in the Italian language, namely Mivia Industrial Speech Commands (MISC) (Bini et al. 2023) and Mivia in-Plant Industrial Speech Commands (MiPISC), each containing 31 command classes. MISC comprises "non-noisy" samples, obtained from individuals via a Telegram bot and automatically generated through Text-To-Speech services. This dataset includes 6,650 samples (5,696 from real speakers). We divide the dataset into 80% training, 10% validation, and 10% test sets. MiPISC consists of 992 samples collected from a noisy environment at an Italian car manufacturer's site. We use this dataset only to test the system in a real industrial environment. To facilitate noise introduction during training and validation, we created a noise dataset, incorporating factory noise and various industrial noise types sourced from the FreeSound dataset (Fonseca et al. 2017).

In our experimentation, we compare three approaches: (a) "None": Training without any data augmentation. (b) "Static": Training with domain-specific but static data augmentation. For each sample, we generate noisy versions at SNRs of 0, 5, 10, 15, 20, 25, 30, 35, and 40, effectively expanding the dataset size by a factor of 9. (c) "Proposed": Training with the proposed domain-specific and dynamic data augmentation (Table 2).

Table 1. Percentage accuracy on the test set of MISC dataset varying the SNR.

Data Augmentation	0	5	10	15	20	25	30	35	40	Average
None	19.78	40.66	70.33	93.40	98.90	**100.00**	**100.00**	**100.00**	**100.00**	80.34
Static	**86.81**	**95.60**	**98.90**	98.90	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	97.80
Proposed	**86.81**	**95.60**	**98.90**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**97.92**

Table 2. Percentage accuracy on MiPISC dataset.

Data Augmentation	Accuracy
None	70.46
Static	83.06
Proposed	**86.69**

In Table 1, the results on the test set of the MISC dataset are presented, showcasing the impact of varying the SNR applied to the speech samples. Notably, data augmentation, encompassing both the "Static" and the "Proposed" dynamic approaches, results in a significant increase in accuracy, enhancing the baseline with no data augmentation by 17.5%. While the proposed approach may not exhibit substantial improvements in an artificially constructed dataset (MISC), its effectiveness becomes evident in real industrial environments (MiPISC) with authentic noise conditions. Indeed, the proposed approach enhances system accuracy by 3.6%. It's worth mentioning that the proposed data augmentation significantly reduces the training time, by a factor of 3.5. Furthermore, the proposed system operates in real-time on an NVIDIA Jetson Xavier NX with a remarkably low memory footprint of just 1.76 MB. Computation involves 8.2 ms for preprocessing tasks such as windowing, Mel spectrogram computation, and VAD, followed by 6.7 ms for model inference.

4 Conclusion

In this paper, we propose a speech assistant within the manufacturing sector, characterized by particularly challenging noise conditions. To tackle this noisy environment, we propose the utilization of a dynamic and domain-specific data augmentation approach. Unlike traditional methods, this technique circumvents the upfront creation of an augmented dataset, resulting in accelerated training (×3.5) and improved accuracy (+3.6%), with a low memory footprint and good execution time of the entire pipeline.

References

Schmidt, B., et al.: Industrial virtual assistants: challenges and opportunities. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pp. 794–801 (2018)

Mishra, A., Makula, P., Kumar, A., Karan, K., Mittal, V.: A voice-controlled personal assistant robot. In: 2015 International Conference on Industrial Instrumentation and Control (ICIC), pp. 523–528. IEEE (2015)

Li, C., Chrysostomou, D., Yang, H.: A speech-enabled virtual assistant for efficient human-robot interaction in industrial environments. J. Syst. Softw. **205**(111), 818 (2023)

Bingol, M.C., Aydogmus, O.: Performing predefined tasks using the human-robot interaction on speech recognition for an industrial robot. Eng. Appl. Artif. Intell. **95**(103), 903 (2020)

Aydogmus, O., Bingol, M.C., Boztas, G., Tuncer, T.: An automated voice command classification model based on an attention-deep convolutional neural network for industrial automation system. Eng. Appl. Artif. Intell. **126**(107), 120 (2023)

Greco, A., Roberto, A., Saggese, A., Vento, M.: DENet: a deep architecture for audio surveillance applications. Neural Comput. Appl. **33**(17), 11, 273–284 (2021)

Park, DS., et al.: SpecAugment: a simple data augmentation method for automatic speech recognition. In: Interspeech 2019 (2019)

FELICE Project: FELICE Project Newsletter - Q1 2022 (2022). https://www.felice-project.eu/wp-content/uploads/2022/05/Newsletter-Q12022.pdf. Accessed 15 Oct 2023

wiseman: py-webrtcvad - WebRTC VAD for Python (2016). https://github.com/wiseman/py-webrtcvad. Accessed 15 Oct 2023

Gold, B., Morgan, N., Ellis, D.: Speech and Audio Signal Processing: Processing and Perception of Speech and Music. Wiley (2011)

Bini, S., Percannella, G., Saggese, A., Vento, M.: A multi-task network for speaker and command recognition in industrial environments. Pattern Recogn. Lett. **176**, 62–68 (2023). https://doi.org/10.1016/j.patrec.2023.10.022. https://www.sciencedirect.com/science/article/pii/S0167865523002945

Fonseca, E., et al.: Freesound datasets: a platform for the creation of open audio datasets. In: Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017), Suzhou, China, pp. 486–493 (2017)

Towards Robotization of Foraging Wild Fruits Under Canopy - A Multi-camera Drone-Borne Berry Mapping

Paweł Trybała[1(✉)], Luca Morelli[1,2], Fabio Remondino[1], and Micael S. Couceiro[3]

[1] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy
{ptrybala,lmorelli,remondino}@fbk.eu
[2] Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy
[3] Ingeniarius Lda, Alfena, Portugal
micael@ingeniarius.pt

Abstract. The rapid technological advancements are allowing not only the automatization of work, but often also its facilitation through human-robot collaboration. This topic is investigated in the FEROX project, which addresses the underexplored domain of improving the process of wild berry harvesting in Northern European forests with robotics and AI. This paper investigates the integration of multi-camera drone technology for under-canopy mapping in the context of wild berry location mapping. Our proposed methodology lays a groundwork for utilizing AI methods to provide georeferenced maps of berries' locations in forest areas, inherently characterized by an unreliable GNSS signal. We carry out initial tests in a forest in eastern Finland with a custom hexacopter, proving the suitability of our approach for retrieving a geographical position of detected fruits with the tested sensor configuration, enabling further processing to supply foragers with wild fruit yield heat maps on a per-species basis.

Keywords: UAV · robotics · visual SLAM · 3D · wild fruit · foraging

1 Introduction

Drones, or unmanned aerial vehicles (UAVs), and AI are gradually revolutionizing our practices, offering various advantages for daily practices, reducing efforts and providing undeniable benefits. In the context of Northern European forests, where wild fruits like blueberries and cloudberries grow abundantly, these technological advancements hold great promise. However, despite the natural abundance, it is estimated that less than 10% of the total annual wild berry yield is harvested from these forests. The main challenge in collecting wild berries lies in the manual harvesting, namely pickers' working conditions. Due to short seasons, most of the work is conducted by foreigners with limited knowledge of the language, practices and the area [1]. Previous studies examining the feasibility of providing berry yield maps to support pickers have underscored the need to carry

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 34–38, 2024.
https://doi.org/10.1007/978-3-031-76428-8_7

out a practical pilot study [2]. The EU FEROX[1] project aims to develop innovative solutions based on the integration of drones, AI, computer vision and 3D mapping to monitor harvesting activities in under canopy environments. In the long term, this will incorporate improvements in the working conditions of berry pickers and mitigate the risk factors and threats to their safety during berry picking activities.

2 Related Works

UAVs have found extensive applications among professionals and enthusiasts across diverse industries, including construction, city planning, cultural heritage preservation, mining, and geology [3]. While the autonomy of flight in open areas with reliable access to GNSS-RTK positioning has reached a mature state, more intricate environments, such as under canopy flights in the forests, still present formidable challenges.

Several studies have delved into the exploration of deploying drones beneath forest canopies. The predominant method for mapping and navigation involves the use of LiDAR sensors. Hyyppä et al. [4] introduced an approach to forest mapping using a commercial simultaneous localization and mapping (SLAM) for extracting valuable information for forest inventory purposes. Wang et al. [5] extended the work by incorporating a high-quality Riegl LiDAR into the methodology. Tian et al. [6] employ LiDAR sensors in collaborative SLAM, but for the main purpose of navigating a UAV fleet during search and rescue operations.

Conversely, several studies have explored visual sensors in mapping forests from the under-canopy. Krisanski et al. [7] conducted an analysis of tree diameter measurements based on images captured using a small UAV. Similarly, Zhang et al. [8] investigated the potential use of 3D reconstruction results of drone-based imagery for biomass measurements. Notably, both studies employed manually flown drones and processed the data using Agisoft Metashape [9].

3 Proposed Methodology

3.1 Data Processing Pipeline

In this paper, we propose a novel data processing pipeline for obtaining berry yield maps. The method utilizes a drone, equipped with 2 stereo cameras and a GNSS-RTK receiver.

In the first step, COLMAP-SLAM[2] [10, 11] is used for processing the high-frequency front-facing stereo images. Based on matching of stereo pairs, the camera trajectory is retrieved. Then, camera timestamps are matched using the nearest neighbor approach to the timestamps of GNSS positions. All positions obtained with too low precision are discarded. A 2D Helmert transformation is used to transform the front-facing camera trajectory from local to global coordinate frames.

[1] This work has been partly funded by the EU FEROX project (https://ferox.fbk.eu/) which received funding from the European Union's Horizon Europe Framework Programme under the GA No. 101070440.

[2] https://github.com/3DOM-FBK/COLMAP_SLAM.

In the next step, the external orientation parameters of both stereo cameras are employed to transform the georeferenced set of forward-looking camera poses into the poses of the downward-facing camera. The parameters for this transformation are assumed to be known from the field calibration. This step is necessary due to the low overlap of the nadir images. A nearest neighbor approach is used, once again, to match the timestamps of the front- and down-facing cameras. Finally, the georeferenced nadir images are transferred to Agisoft Metashape [9], where the generation of an orthomosaic is carried out. The results can be used to apply AI-based methods to obtain the location of berries, either directly from the subset of the orthomosaic, or by performing the detection on single images and finding the corresponding position in the orthomosaic.

3.2 Drone Setup

For testing the proposed methodology, a custom hexacopter *Scout* developed by Ingeniarius Lda (Fig. 1) has been employed. The drone is equipped with 2 ZED X stereo cameras (nadir- and front-facing), an RTK-enabled Emlid Reach GNSS receiver, an inertial measurement unit, Pixhawk flight controller and onboard Jetson Orin AGX with Robot Operating System (ROS; [12]) Noetic for data collection. To ensure stable framerates, they were limited to 15Hz and 0.5Hz for the front- and down-facing cameras.

Fig. 1. Custom drone *Scout* developed by Ingeniarius Lda (left). View from a front-facing stereo camera of the drone inside the forest (right).

4 Experiments

4.1 Data

Data were collected during the summer field work in June 2023 in Ilomantsi, in eastern Finland (Fig. 2) with the *Scout* UAV presented in Sect. 3.2. For the purpose of testing the data processing pipeline, a sequence with 2,081 stereo-pairs from the forward-looking camera and 234 stereo pairs from the nadir camera was collected. In total, 171 valid GNSS positions were used: the rest (around 95%) have not passed the quality control criteria of fixed position with a horizontal precision at most equal to 1 m.

4.2 Results

The drone trajectory and a sparse point cloud of tie points can be seen in Fig. 2. Notably, tree trunks are clearly identifiable, without shadowing effects of odometry errors, despite a complex trajectory.

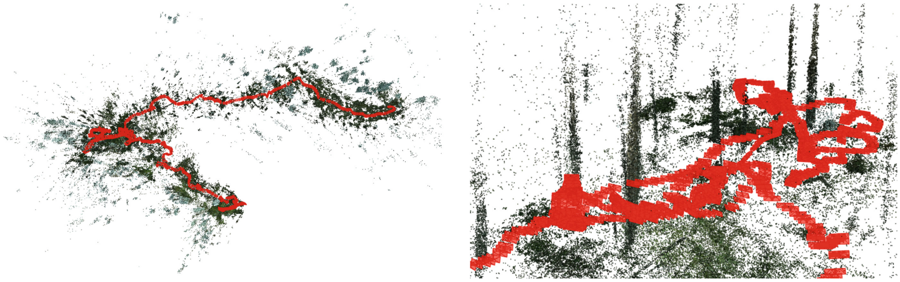


Fig. 2. A sparse point cloud and the front-facing camera trajectory (in red) reconstructed with COLMAP-SLAM [11]: an orthographic top view (left) and close-up view (right).

Then, the georeferencing procedure has been performed. The root mean squared error (RMSE) of the 2D alignment between the trajectory and sparse GNSS positions was equal to 1.8m, which, considering the cutoff of 1 m for the GNSS position precision, was deemed acceptable. In the next step, an orthomosaic was generated (Fig. 3) and a berry detection was simulated. Since the AI models for fruit detection are not yet available, a pair of berries are annotated by hand on the image to showcase the berry geolocation function of the workflow (Fig. 4).

Fig. 3. Orthomosaic generated in Agisoft Mehashape.

5 Conclusions

In this study, a novel pipeline for UAV data processing has been established, enabling the creation of georeferenced wild fruit maps based on under-canopy surveys. An example dataset obtained in a Finnish forest during the summer season has been processed and the workflow has been successfully tested. In further developments, focus will be on integrating the described methodology with AI-based fruit detection models. Additionally,

Fig. 4. Example results of the proposed pipeline. Berry locations marked on the nadir drone image (left) and their respective positions in a global reference frame on an orthomosaic (right).

further tests will be conducted with an improved hardware setup to select optimal data acquisition parameters and verify the robustness of the method on different datasets.

References

1. Turtiainen, M., Nuutinen, T.: Evaluation of information on wild berry and mushroom markets in European countries. Small-Scale For. **11**(1), 131–145 (2011)
2. Hujala, T., Kurttila, M., Store, R.: Mobile forest berry map service: co-creating value from open public data. In: Toivonen, M. (ed.) Services and New Societal Challenges: Innovation for Sustainable Growth and Welfare. Proceedings of the 24th Annual RESER Conference, 11–13 September 2014, Helsinki, pp. 593–600. RESER & VTT (2014)
3. Nex, F., Remondino, F.: UAV for 3D mapping applications: a review. Appl. Geomat. **6**, 1–15 (2014)
4. Hyyppä, E., et al.: Under-canopy UAV laser scanning for accurate forest field measurements. ISPRS J. Photogramm. Remote Sens. **164**, 41–60 (2020)
5. Wang, Y., et al.: Seamless integration of above- and under-canopy unmanned aerial vehicle laser scanning for forest investigation. Forest Ecosyst. **8**(1) (2021)
6. Tian, Y., et al.: Search and rescue under the forest canopy using multiple UAVs. Int. J. Robot. Res. **39**(10–11), 1201–1221 (2020)
7. Krisanski, S., Taskhiri, M., Turner, P.: Enhancing methods for under-canopy unmanned aircraft system based photogrammetry in complex forests for tree diameter measurement. Remote Sens. **12**(10), 1652 (2020)
8. Zhang, Y., Onda, Y., Kato, H., Feng, B., Gomi, T.: Understory biomass measurement in a dense plantation forest based on drone-SfM data by a manual low-flying drone under the canopy. J. Environ. Manag. **312**, 114862 (2022)
9. Agisoft LLC: Agisoft Metashape (2.1) [Software] (2023)
10. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113 (2016)
11. Morelli, L., Ioli, F., Beber, R., Menna, F., Remondino, F., Vitti, A.: COLMAP-SLAM: a framework for visual odometry. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **XLVIII-1/W1-2023**, 317–324 (2023)
12. Quigley, M., et al.: ROS: an open-source robot operating system. In: ICRA Workshop on Open Source Software 3, 5, Kobe, Japan (2009)

PROPHET: PReference-Based OPtimization for Human-cEnTric Visual Inspection

Marco Maccarini[1(✉)], Alberto Gottardi[2], Dario Piga[1], and Loris Roveda[1]

[1] Istituto Dalle Molle di studi sull'intelligenza artificiale (IDSIA USI-SUPSI), DTI, Scuola universitaria professionale della Svizzera italiana, 6962 Lugano, Switzerland
{marco.maccarini,dario.piga,loris.roveda}@supsi.ch
[2] IT+Robotics srl, 36100 Vicenza, Italy
alberto.gottardi@it-robotics.it

Abstract. Nowadays, complex inspection processes rely heavily on human operators, while automatic systems handle simpler tasks. However, these tasks are highly repetitive and demand consistent high-quality performance throughout, leading to significant stress for human workers. In contrast, automatic systems can help alleviate this burden. Nevertheless, configuring automatic inspection systems is challenging due to numerous parameters that require extensive time and trial-and-error adjustments. To address these issues, this project aims to introduce an optimization approach based on user preferences for configuring visual inspection systems. Preference-based optimization is a potent method for enhancing system performance in an intuitive manner. This methodology enables the resolution of optimization problems when the decision-maker cannot directly assess the objective function tied to the problem at hand. Instead, they can only express preferences, such as "this option is better than that" when comparing different decision choices.

Keywords: Human-centric production · human-robot collaboration · human-robot interaction · knowledge transfer · preference-based optimization · artificial intelligence

1 Introduction

Robotics is increasingly being used to assist humans in various fields such as rehabilitation, rescue missions, and medical applications. In these applications, robotic systems need to support and enhance human tasks by utilizing intelligent tools to understand human needs, requirements, instructions, feedback, and satisfaction [1,3].

One significant domain where robotics plays a crucial role is industrial workplaces, particularly in the context of Industry 4.0 and its evolution into Industry 5.0 [4,6]. Industry 5.0 aims to improve the working conditions of industrial operators while harnessing human knowledge, adaptability, and intelligence. A central goal of this new industrial revolution is to shift the focus from connected cyber-physical systems to the operators themselves, creating human-centric production environments that leverage human expertise to enhance industrial processes [5]. Instead of performing repetitive manual tasks, humans are placed at the core of the production environment, taking on a leading

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 39–44, 2024.
https://doi.org/10.1007/978-3-031-76428-8_8

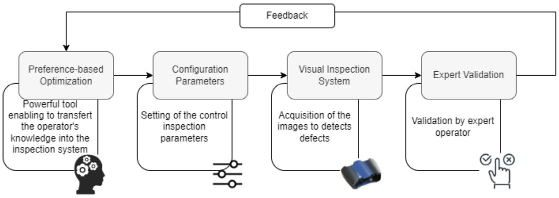


Fig. 1. PROPHET overview.

role. They are involved in high-value operations where their expertise can optimize the production process, leaving easily automatable tasks to robotic systems [7]. Thus, tools are needed to effectively and naturally transfer operator expertise to robotic systems.

The primary objective of the PROPHET project is to introduce and validate a human-centric approach for transferring human knowledge of a task to robotic systems, relying solely on qualitative feedback. This approach utilizes preference-based optimization (PBO) to enable non-expert programmers to impart their expertise to robotic systems, specifically for fine-tuning a target task. The project focuses on a visual inspection task, aiming to develop a system capable of automatically configuring control parameters by learning from the choices made by expert operators. The resulting system facilitates the adoption of automated visual inspection, even in complex tasks, by reducing setup time through the incorporation of skilled operators' inspection results and their valuable experience.

2 Methodology

As previously discussed, the goal of this project is to enhance the efficiency of inspection software by automating the parameter optimization process. Up until now, this task had been performed by a skilled user who relied on a trial-and-error approach to fine-tune the machine's settings until reaching the optimal configuration. This process solely relied on the user's expertise, lacking any mathematical criteria for selecting the parameters to test. Additionally, this operation was quite time-consuming, taking the user approximately 1-working day to identify the machine's optimal parameters. The solution chosen for this problem involves employing preference-based optimization (PBO). This optimization technique leverages a surrogate function to approximate the behaviour of the inspection machine, guided by user preferences. In simpler terms, it helps determine the best machine parameters by considering user experience and preferences. An overview of the PROPHET architecture is depicted in Fig. 1. The optimization algorithm utilized in this study is founded on the methodology outlined in a prior work by one of the authors [2], where the GLISp algorithm was introduced. The algorithm is not summarized in this paper, for a comprehensive understanding please consult the original paper [2].

3 Use Case

An evaluation of the preference-based visual inspection system was conducted using three real-world case studies, each involving a dataset of workpieces with and without

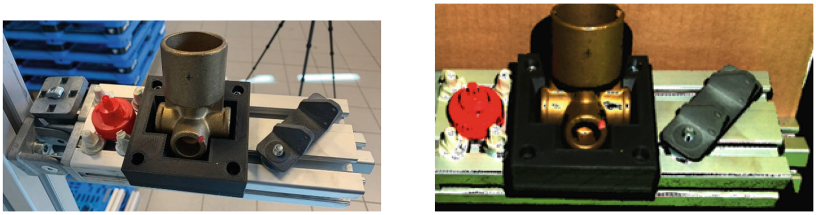


Fig. 2. Workpieces used for evaluating the system. On the left the RGB image, on the right the PointCloud.

defects. Figure 2 depicts an example of the three objects used to evaluate the system: use case 1 is a brass component (object in the center), while cases 2 and 3 are plastic pieces (objects on the side). The goal was to assess the system's performance in detecting various types of defects, including obvious and subtle ones. The evaluation protocol for each case study follows:

- **Configuration of Vision System:** The first phase involved setting up the vision system, including configuring camera parameters and acquiring a reference image of a defect-free workpiece (referred to as the "master"). This master serves as the benchmark for identifying defects in other workpieces. Additionally, masks were created to define specific Regions of Interest (ROIs) within the master image. These masks are used to align the master with the current workpiece image for defect detection.
- **Optimization of Inspection Parameters:** In the second phase, a workpiece with significant defects was chosen to fine-tune the inspection parameters using the preference-based optimization module. Initially, a set of parameter guesses was selected, and operators used the preference-based module to refine these parameters. Once optimization was completed, the best parameters were chosen based on operator preferences.
- **Validation of Parameters:** The third phase involved validating the optimized parameters using a variety of workpieces, both defective and defect-free, to ensure that the system accurately detected defects.

This project was focused on optimizing three key inspection algorithm parameters:

- **Control Threshold:** This parameter determines the minimum defect surface area (in square millimeters) that the system should identify.
- **Correlation Threshold:** It represented a multiplier applied to the resolution, setting the minimum distance (in millimeters) from the master image that a point must be to be considered a defect.
- **Cluster Defects Resolution Multiplier:** This parameter acted as a multiplier factor to the resolution, enabling the system to identify two nearby points as part of the same defect cluster.

To conduct the experiments, a workcell was set up in the IT+Robotics laboratory, consisting of the EyeT+ Inspect vision system and a template for positioning the three types

Table 1. Technical KPIs and results obtained.

KPI No	Description of Technical KPI	Starting value	Results
KPI-1	Configuration time	1 working day	[0.5, 1] h
KPI-2	Inspection mistakes	0.1%	0.02%
KPI-3	Productivity	–	500%
KPI-4	Reduced number of optimization iterations	–	20 iterations

of workpieces. The template ensured that experiments were repeatable, replicable, and comparable.

4 Result

The KPIs described in Table 1 are used to evaluate the preference-based visual inspection system. Since three different use cases are used for the validation phase, each case study is evaluated through the KPIs. Following the protocol and steps described in the previous paragraphs, this section presents the results achieved at the end of the project.

Configuration time (**KPI-1**) is one of the most important KPIs because it describes the amount of time spent configuring the inspection parameters of the system. At the beginning of the project, the time spent on this activity was 1 working day, i.e., 8 h. At the end of the project, the time was reduced between 30 min and 1 h at most for each use cases. In particular, use case 1 and 3 required at most 30 min to find a suitable set of values, while use case 2 needed at most 1 h. During the experiments, it was noted that these values depended on both the size of the chosen domain and the optimiser mode, i.e., whether continuous or discrete mode is used. A large domain with the continuous mode requires lots of time, while a small domain with discrete mode requires less time. In fact, for the evaluation a continuous mode with a large domain [0, 10] were used. In order to improve the system, the discrete mode was added and tested with a small domain [0, 5] and incremental step 0.1. The number of iterations is partially related to the optimization time and the domain size and mode used: with a continuous mode at least 50 iterations were needed, while with a discrete mode only 15/20 iterations were needed. The difference between the continuous and discrete modes highlights that the discrete optimization approach was able to converge on the optimal solution more efficiently than its continuous counterpart. This knowledge can be leveraged to guide future optimization efforts and enhance the efficiency of optimization algorithms. The best parameters obtained from the optimiser were then used to inspect all the other objects in the dataset and to evaluate **KPI-3** and **KPI-4**. As summarized in Table 1, both KPIs are achieved. These results show that the preference-based optimization module can best configure the parameters of the inspection system, and the values obtained are more accurate than those found through trial and error. Figure 3 depicts an example for the use cases: the non-defective object and two pieces with defects. The non-defective object is on the right of the image, the heavy fault object is in the middle, and the lighter defective object is on the left.

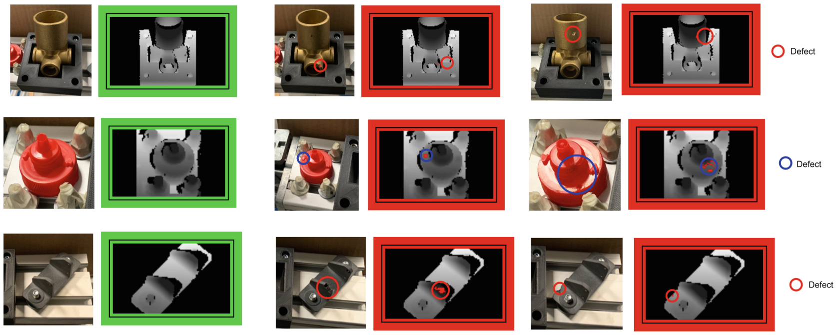


Fig. 3. Example of defects and non-defect recognition.

5 Conclusions

Preference-based Optimization represents a highly potent methodology for optimizing variables across various applications, including industrial machines and robots. One key advantage of PBO is its ability to leverage user experience as an added value to the optimization process. This aspect is crucial for enhancing efficiency and from a social perspective, as human-centric considerations are always paramount. By prioritizing the human element and considering the user's preferences, PBO can optimize the outcomes in a manner that is not only highly effective but also highly ethical, thus providing a more holistic approach to Optimization that considers both the technological and social aspects of the problem at hand. The application of preference-learning technologies in visual inspection has a strong impact from many points of view, such as simplifying and accelerating the deployment, allowing non-skilled operators to train the system, and increasing production efficiency, enabling wide adoption of such technology. The proof-of-concept developed within this project demonstrates the applicability of the PBO component to a visual inspection system. The designed component could be easily adaptable to various applications and rapidly transferred to other industrial fields, thus significantly improving its potential impact.

References

1. Abdelaal, A.E., Mathur, P., Salcudean, S.E.: Robotics in vivo: a perspective on human-robot interaction in surgical robotics. Ann. Rev. Control Robot. Auton. Syst. **3**, 221–242 (2020)
2. Bemporad, A., Piga, D.: Global optimization based on active preference learning with radial basis functions. Mach. Learn. **110**(2), 417–448 (2021)
3. Dalla Gasperina, S., Roveda, L., Pedrocchi, A., Braghin, F., Gandolla, M.: Review on patient-cooperative control strategies for upper-limb rehabilitation exoskeletons. Front. Robot. AI **8** (2021)
4. ElFar, O.A., Chang, C.-K., Leong, H.Y., Peter, A.P., Chew, K.W., Show, P.L.: Prospects of industry 5.0 in algae: customization of production and new advance technology for clean bioenergy generation. Energy Convers. Manag.: X **10**, 100048 (2021)
5. Nahavandi, S.: Industry 5.0-a human-centric solution. Sustainability **11**(16), 4371 (2019)

6. Oztemel, E., Gursev, S.: Literature review of industry 4.0 and related technologies. J. Intell. Manuf. **31**(1), 127–182 (2020)
7. Vicentini, F., et al. PIROS: cooperative, safe and reconfigurable robotic companion for CNC pallets load/unload stations. In: Caccavale, F., Ott, C., Winkler, B., Taylor, Z. (eds.) Bringing Innovative Robotic Technologies from Research Labs to Industrial End-Users. STAR, vol. 136, pp. 57–96. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-34507-5_4

Unlocking the Potential of Human-Robot Synergy Under Advanced Industrial Applications: The FEROX Simulator

Beril Yalcinkaya[1,2](✉), André Araújo[1], Micael Couceiro[1], Salviano Soares[2], and António Valente[2]

[1] Ingeniarius, Ltd., R. Nossa Sra. Conceição 146, 4445-147 Alfena, Portugal
{beril,andre,micael}@ingeniarius.pt

[2] School of Sciences and Technology-Engineering Department, University of Trás-os-Montes and Alto Douro (UTAD), Quinta de Prados, 5000-801 Vila Real, Portugal
{salblues,avalente}@utad.pt

Abstract. Human-Robot Collaboration (HRC) in advanced industrial scenarios has emerged as a transformative force. Modern robots, infused with artificial intelligence (AI), can enhance human capabilities, offering a wide spectrum of opportunities in agriculture, forestry, construction and many other domains. However, the complex nature of HRC demands realistic simulators to bridge the gap between theory and practice. This paper introduces the FEROX Simulator, purpose-built for robot-assisted wild berry collection. We briefly delve into the simulator's capabilities to showcase its potential as a platform to develop HRC systems. Our research underscores the need for simulators designed for challenging HRC contexts and aims to inspire further advancements in this domain.

Keywords: human-robot collaboration · robotic simulators · advanced industrial applications

1 Introduction

Robots have transcended the realm of science fiction to become indispensable in our daily lives, assisting us in tasks ranging from the mundane to the complex. Yet, their true potential lies not in isolation, but in collaboration with humans [1]. This is the essence of Human-Robot Collaboration (HRC) - a burgeoning field where humans and robots join forces to achieve feats beyond their individual capabilities. In an era marked by artificial intelligence (AI) and modern

This work has been partly funded by the EU FEROX project (https://ferox.fbk.eu/) which received funding from the European Union's Horizon Europe Framework Programme under Grant Agreement No. 101070440. Views and opinions expressed are however those of the authors only and the European Commission is not responsible for any use that may be made of the information it contains.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 45–49, 2024.
https://doi.org/10.1007/978-3-031-76428-8_9

robotics, imbuing robots with intelligent algorithms for HRC has become a pivotal research focus. The fusion of human prowess with robotic precision and endurance opens doors to a spectrum of tasks performed more efficiently.

1.1 Research Question and Objective

Human-Robot Collaboration (HRC) holds immense promise across various sectors, including agriculture, forestry, and construction. However, designing effective HRC systems is challenging. Robots must not only excel at tasks but also understand and respond to human behaviour, which is dynamic and unpredictable. To address these complexities, researchers turn to advanced virtual environments like simulators and digital twins. These tools provide controlled spaces for developing, testing, and optimizing robotic systems, all while exploring human-robot interactions without the need for physical equipment [10].

However, the current shortage of simulators for Human-Robot Collaboration (HRC) is a significant challenge. Existing frameworks such as Gazebo [4], MORSE [3], USARSim [2], CoppeliaSim [9], ARGoS [7], RoboNetSim [5], primarily focus on robot-only applications, leaving a gap in simulating modern collaborative robots. This paper introduces the FEROX Simulator, tailored for robot-assisted wild berry collection, offering a comprehensive virtual environment for effective HRC. Our research aims to inspire further innovation in the pivotal domain of HRC by offering a realistic platform to test and optimise systems under such challenging applications.

2 The FEROX Simulator

FEROX, a Horizon Framework Programme initiative funded by the EU, harmoniously integrates human elements with cutting-edge robotic technologies, aiming to revolutionise the wild berry industry[1]. This innovative approach reimagines berry picking, mitigating the physical and mental strains on pickers, while enhancing safety and efficiency in HRC.

2.1 General Architecture

The FEROX simulator accommodates interactions involving both single and multiple humans and robots, supporting various short and long-term HRC scenarios. The FEROX Simulator comprises two distinct types of instances: (i) **Human Instance**, which may comprehend multiple instances, each enabling the control of a single virtual human agents; and (ii) **Robot Instance**, which is a single instance running on a server machine, put together to manage robotic agents and other mission- and environmental-related conditions as shown in Fig. 1.

The multiple human instances coexisting within the same network seamlessly communicate with the robot instance via the Robot Operating System

[1] https://ferox.fbk.eu/.

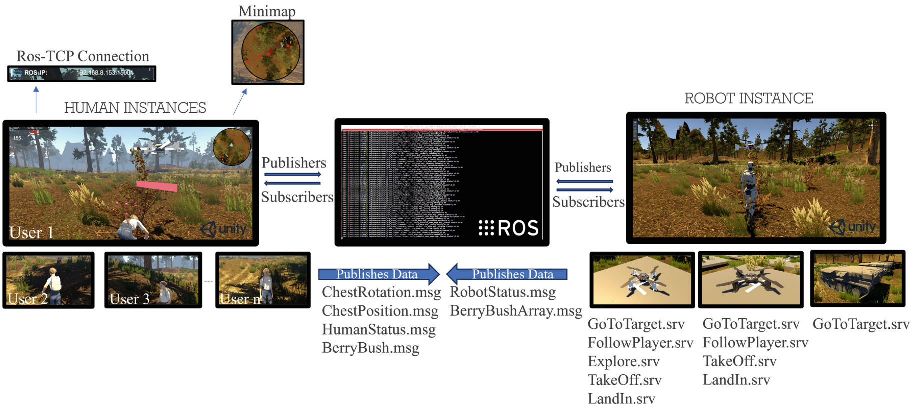


Fig. 1. The general overview of the Ferox Simulator.

(ROS), utilising topics for data exchange and action servers to initiate specific robotic tasks [8]. The simulator's architecture forms a robust foundation for its diverse range of features, which we will delve into in subsequent sections.

2.2 Ecosystem

In the FEROX Simulator, two pivotal components come together to create a versatile platform for simulating complex HRC scenarios: Unity and ROS. Unity, as a leading game engine, provides the foundation for creating a realistic and immersive simulation environment, including a powerful physics engine and sophisticated rendering capabilities. ROS, on the other hand, serves as the backbone of the FEROX Simulator [6]. Its adoption also ensures that the simulator can seamlessly transition into the real world, as it operates within the same ROS ecosystem used by robots. This means that the robot instance can be readily replaced with real robots, facilitating a smooth transition from simulation to practical deployment.

The ROS TCP Connector[2] plays a crucial role in seamless integration, serving as the linchpin connecting human and robot instances. It orchestrates synchronization through ROS topics and ROS services, facilitating the coordination of human operator movements from the Human Instance to the robot instance. It also triggers ROS services in the Robot Instance, replicating them in the Human Instance, ensuring fidelity and accuracy in the FEROX Simulator.

2.3 Environments

In the FEROX Simulator, Unity engine capabilities enable easy customization of immersive scenarios for advanced applications without extensive programming skills. Currently, scenario replication is needed for both robot and human

[2] https://github.com/Unity-Technologies/ROS-TCP-Connector.

instances. Future developments aim to simplify this process by enabling content transfer from the robot instance to the human instance, similar to multiplayer online games.

Interactable objects, like berries in robot-assisted picking, are managed by the robot instance within virtual environments. For example, the robot instance generates random berry locations and broadcasts them to connected human instances. Human-initiated actions, such as berry picking, occur in the human instance and are transmitted to the robot instance, updating berry states across all human instances. This bidirectional communication ensures a dynamic and realistic HRC environment with immediate and coherent consequences for both human and robot actions.

2.4 Robots

FEROX Simulator seamlessly integrates with ROS for access to its ecosystem, including navigation and SLAM packages. A unique feature is its user-friendly robot control, allowing researchers to define the robot as a Unity NavMesh Agent with built-in path planning and navigation. This abstraction simplifies interaction with the robot, much like a non-player character (NPC) in a video game. Especially when its capabilities align with those of a virtual agent, ensuring smoother control.

At its current state, the FEROX Simulator currently supports three types of robots, each providing distinct ROS services. This includes small Unmanned Aerial Vehicles (UAVs) for forestry exploration and mapping, along with large UAVs and Unmanned Ground Vehicles (UGVs) for berry transportation. These diverse platforms are designed for various wild berry collection scenarios, enabling researchers to optimize robot behaviors in challenging real-world conditions, ultimately enhancing human-robot collaboration.

2.5 Humans

The FEROX Simulator stands out by enabling users to control avatars similar to video game characters. Each human instance operates on a separate machine, creating a third-person human picker multiplayer simulator. This feature distinguishes it from other simulators that focus mainly on the robotic perspective and plays a crucial role in HRC development.

Users can perform actions like requesting specific robotic services, such as tasking a UAV for mapping designated regions or arranging rendezvous with it. They can also interact with the environment by collecting berries from bushes and loading them onto the robots. These activities closely resemble real-world tasks, allowing researchers to test and refine HRC strategies. Furthermore, The simulator further engages users through gamification. A scoring system awards points based on the number of berries they collect, creating a "coopetitive" environment where users can collaborate or compete in berry picking. This not only adds enjoyment but also explores various collaborative and competitive dynamics in HRC scenarios.

3 Conclusion

The FEROX Simulator stands as a tool for modern HRC in advanced industrial applications. It addresses the need for realistic simulators in modern robotic domains, where AI-driven robots can significantly augment human capabilities. This simulator, purpose-built for robot-assisted wild berry collection, offers a comprehensive platform to explore and optimise multi-human multi-robot collaboration. Future work involves expanding the simulator's capabilities and facilitating seamless content sharing between human and robot instances, further solidifying its position as a valuable asset in advancing HRC research and development. A stable version of the FEROX simulator is expected to be made freely available in Q1 2024.

References

1. Ajoudani, A., Zanchettin, A.M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., Khatib, O.: Progress and prospects of the human-robot collaboration. Auton. Robot. **42**, 957–975 (2018)
2. Carpin, S., Lewis, M., Wang, J., Balakirsky, S., Scrapper, C.: USARSim: a robot simulator for research and education. In: Proceedings 2007 IEEE International Conference on Robotics and Automation, pp. 1400–1405. IEEE (2007)
3. Echeverria, G., Lassabe, N., Degroote, A., Lemaignan, S.: Modular open robots simulation engine: MORSE. In: 2011 IEEE International Conference on Robotics and Automation, pp. 46–51. IEEE (2011)
4. Koenig, N., Howard, A.: Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566), vol. 3, pp. 2149–2154. IEEE (2004)
5. Kudelski, M., Gambardella, L.M., Di Caro, G.A.: RoboNetSim: an integrated framework for multi-robot and network simulation. Robot. Auton. Syst. **61**(5), 483–496 (2013)
6. Mizuchi, Y., Inamura, T.: Cloud-based multimodal human-robot interaction simulator utilizing ROS and unity frameworks. In: 2017 IEEE/SICE International Symposium on System Integration (SII), pp. 948–955. IEEE (2017)
7. Pinciroli, C., et al.: ARGoS: a modular, multi-engine simulator for heterogeneous swarm robotics. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5027–5034. IEEE (2011)
8. Quigley, M., et al.: ROS: an open-source robot operating system. In: ICRA Workshop on Open Source Software, Kobe, Japan, vol. 3, p. 5 (2009)
9. Rohmer, E., Singh, S.P., Freese, M.: V-REP: a versatile and scalable robot simulation framework. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1321–1326. IEEE (2013)
10. Yalçinkaya, B., Couceiro, M.S., Soares, S.P., Valente, A.: Human-aware collaborative robots in the wild: coping with uncertainty in activity recognition. Sensors **23**(7), 3388 (2023)

Towards Explainable Human Motion Prediction in Collaborative Robotics

Michael Vanuzzo[1(✉)], Francesco Borsatti[2], Marco Casarin[1], Mattia Guidolin[1], Monica Reggiani[1], and Stefano Michieletto[1]

[1] Department of Management and Engineering (DTG), University of Padova, Stradella S. Nicola, 3, 36100 Vicenza, Italy
{michael.vanuzzo,marco.casarin.4}@phd.unipd.it,
{mattia.guidolin,monica.reggiani,stefano.michieletto}@unipd.it

[2] Department of Information Engineering (DEI), University of Padova, Via Gradenigo, 6/B, 35121 Padova, Italy
francesco.borsatti.1@phd.unipd.it

Abstract. Predicting human motion is challenging due to its complex and non-deterministic nature. This is particularly true in the context of Collaborative Robotics, where the presence of the robot significantly influences human movements. Current Deep Learning models excel at modeling this complexity but are often regarded as black boxes. Explainable Artificial Intelligence (XAI) offers a way to interpret these models. In this work, we introduce an XAI approach to identify key features in a Human Motion Prediction (HMP) system. Additionally, we semantically associate action labels to the joint rotations representing human motion to further improve the interpretability and precision of the model. We evaluated our system using the AMASS dataset and BABEL labels. Experimental results demonstrated the importance of specific action-related features, enhancing prediction accuracy compared to the Zero-Velocity baseline model.

Keywords: Human Motion Prediction · Human Action Semantic · Explainable AI · Collaborative Robotics

1 Introduction

The anticipation of human motion is a multidisciplinary task that involves the capability of interpreting and understanding human body dynamics. The complexity and the non-deterministic nature of human behavior makes predicting human body poses a very difficult challenge. This is particularly true in a Collaborative Robotics context, where the presence of the robot must be considered due to its strong relation with human behavior. Currently, state-of-the-art approaches are based on Deep Learning (DL) techniques, including Recurrent Neural Network (RNN), Graph Convolutional Network (GCN), Generative Adversarial Network (GAN), and the*attention mechanism* to model human

M. Vanuzzo and F. Borsatti—These authors have contributed equally to the work.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 50–54, 2024.
https://doi.org/10.1007/978-3-031-76428-8_10

motion [3]. These approaches are very powerful in modeling complex data, however they are very difficult to interpret and generally used as black-boxes in an end-to-end procedure [7].

XAI [5] refers to the ability to understand and interpret the decisions made by Artificial Intelligence (AI) systems. It consists of providing insights about why a given model makes a specific decision, helping to understand potential biases and to identify errors inside the model. Moreover, XAI plays a crucial role in the identification of the most relevant features for making predictions, with the aim of simplifying the dataset through feature selection. Bento et al. [1] exploited an XAI approach in a classification task based on a DL model to identify and remove the undesired information from input images. Javed et al. [2] proposed a multi-sensor Human Activity Recognition (HAR) system based on the most important features selected through XAI approaches. The experiments showed performance improvements in multiple Machine Learning (ML) and DL approaches.

In this work, we introduce action semantics as additional information for the task of Human Motion Prediction (HMP)and we propose an XAI approach to identify the most relevant semantic features. Action semantics can significantly help to improve predictions in a Human-Robot Collaboration (HRC) context given that the tasks performed by the user are often known in advance. Moreover, highlighting the importance of semantic features is more intuitive and consistent for human interpretability compared to joint angle rotations.

The remainder of the paper is organized as follows. Section 2 describes the methodology used in this work to perform Human Motion Prediction. Section 3 reports on the experiments carried out to evaluate the proposed system, as well as the obtained results. Finally, Section 4 concludes the article.

2 Methods

Our approach consists of predicting future poses of an individual while gaining insight into the relevance of the semantic features used to describe human motion. Through this approach, we aim to create a system that is more comprehensible and thus easier to expand. The input of our system consists of both data related to the person's past poses and an embedding representing the semantic information of the action being performed. The person's pose is represented as a sequence of j rotations applied to its 3D skeletal structure, uniquely defining the spatial arrangement of the body elements. To describe a sequence of n frames, these poses can be concatenated into a one-dimensional vector of size $3 \cdot \text{ j} \cdot n$. The semantic information is a sentence describing the action performed. It is encoded using a one-hot encoding with p elements. Therefore, the overall input consists of a vector of size $3 \cdot j \cdot n + p$. The predictive architecture employed for this task is a random forest [2], that enables the analysis of the importance of the different input features to the system.

The predicted sequences can be quantitatively evaluated using the Mean Angle Error (MAE) metric, that can be computed as the Euclidean distance between the predicted and the ground truth pose vectors:

$$\mathrm{MAE_t} = \frac{1}{K}\sum_{k=1}^{K} \|\hat{x}_{k,t} - x_{k,t}\|_2$$

In the provided context, $\hat{x}_{k,t}$ and $x_{k,t}$ represent vectors associated with predicted and ground truth sequences, respectively. Both vectors are defined at frame t and encompass the corresponding j rotations expressed as Euler angles following the "XYZ" sequence. Here, K represents the number of motion sequences under consideration. Once the values of $\mathrm{MAE_t}$ are determined for each frame, it is possible to compute the cumulative MAE by summing the obtained values for each frame in the predicted sequence. To assess the effectiveness of the proposed system, we compared it with the Zero-Velocity baseline. In this model, the predicted sequence consists of a repetition of the last ground truth frame. Although this approach is straightforward, it proves effective in establishing a baseline for comparisons that is commonly employed in the field of HMP [4]. Finally, the significance of a feature is assessed through a metric known as Gini importance or Mean Decrease in Impurity (MDI) [6]. It quantifies the total decrease in node impurity weighted by the probability of reaching that node and averaged over all trees of the ensemble. The probability of reaching a specific node is approximated by the proportion of samples reaching that node.

3 Experimental Results

To evaluate the applied approach, experiments were performed using a subset of the Archive of Motion Capture of Surface Shapes (AMASS) dataset. AMASS is a comprehensive repository of human motion data, unifying various optical marker-based motion capture datasets within a common framework and parameterization. The framework employed is Skinned Multi-Person Linear Model (SMPL), which includes a skeletal representation comprising 24 keypoints. Each keypoint describes the relative rotation of a specific joint, except for the first, which represents the pelvis absolute rotation with respect to a global reference system. Each rotation is denoted in the axis-angle representation, which is characterized by a triplet of values. Additionally, the coordinates of the root joint concerning a global reference system are provided, describing the subject's trajectory within the environment during the action. The frame rate of these recordings is 120 Hz, which was reduced to 10 Hz to extend the temporal horizon without significantly increasing the dimensions of input and output information. The input sequence consists of 8 frames, corresponding to the preceding 0.8 s relative to the current moment. Then, the representation of poses during the action comprises $8 \cdot 25 \cdot 3$ values, that are the concatenation of the poses (24 joints plus translation) over the 8 frames used as input.

To obtain a description of the action in each sequence, the Bodies, Action and Behavior with English Labels (BABEL) dataset was used. BABEL represents a comprehensive resource with language labels describing the actions performed in numerous motion-capture sequences from AMASS. To construct a one-hot encoded vector, we cataloged all words from the action labels and marked those present in the respective sequence. This approach led to the creation of a sparse vector with a length of 612 elements.

For training the random forest, we used 25 estimators with a maximum depth of 10, selecting the parameters in an empirically fashion. The results obtained on the evaluation set were then compared with the Zero-Velocity model.

The outcomes on the test set exhibit that the random forest architecture achieves superior results compared to the baseline Zero-Velocity model. Specifically, the random forest architecture attains a MAE of 16.22, whereas the Zero-Velocity baseline yields a result of 20.55. Additionally, it is noteworthy that even over extended temporal horizons, the prediction error after 1.2s is 1.70 for the random-forest-based model and 1.98 for the Zero-Velocity baseline.

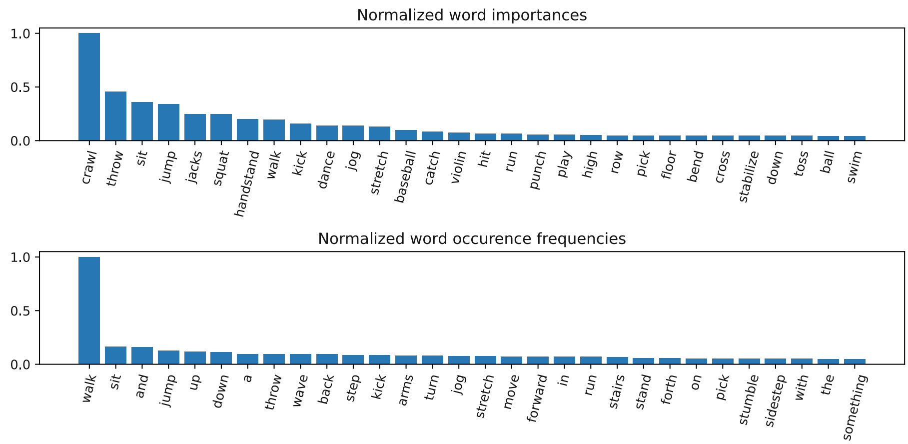


Fig. 1. Comparison between the normalized number of occurrences of words and the importance of them considered as semantic features for Human Motion Prediction

Subsequently, the importance associated with each semantic feature was computed using the Gini importance metric. The most important features were compared with the results obtained from the analysis of the word frequencies in the labels associated with the motion data. As depicted in the Fig. 1, the words that occur most frequently are often prepositions, articles, and conjunctions. However, these words did not have significant relevance for motion prediction and were not prominent in terms of importance. On the contrary, looking at the feature importance, the most crucial terms were associated with specific actions, providing valuable information about the future pose's evolution. It is also important to note that the most frequently occurring word, namely *walk*, appears in multiple descriptions due to the dataset nature. While this term is present on the feature importance scale, its rank is lower due to its significantly reduced discriminating capability in this context.

4 Conclusions

In this paper, we proposed an architecture for HMP that incorporates semantic information about the performed actions within the sequence, demonstrating prediction performance and interpretability of the solution. This strategy enables an analysis of the importance of the utilized semantic features, paving the way for adopting XAI methods in the filed of HMP. XAI is a crucial concept providing better insights into complex problems and being potentially pivotal both in designing improved modeling architectures and in processing input data. In particular, applying XAI into the most recent DL architectures could prove particularly effective when integrated into HRC systems, which strongly benefit from the comprehension of the underlying logic guiding motion prediction. An innovative design of this nature will provide superior understanding of human intentions, enabling rich and precise interaction between human and robot agents, all while ensuring the system's overall safety.

Acknowledgment. This study was partially funded by Next-GenerationEU (Italian PNRR - M4 C2, Invest 1.3 - D.D. 1551.11-10-2022, PE00000004) within the MICS (Made in Italy - Circular and Sustainable) Extended Partnership, CUP: C93C22005280001 and by the PRESENCE (anticiPatoRy bEhaviors for Safe and Effective humaN-robot CoopEration) project (BIRD221598).

References

1. Bento, V., Kohler, M., Diaz, P., Mendoza, L., Pacheco, M.A.: Improving deep learning performance by using Explainable Artificial Intelligence (XAI) approaches. Discover Artif. Intell. **1**, 1–11 (2021)
2. Javed, A.R., et al.: Toward Explainable AI-empowered cognitive health assessment. Front. Public Health **11**, 1024195 (2023)
3. Lyu, K., Chen, H., Liu, Z., Zhang, B., Wang, R.: 3D human motion prediction: a survey. Neurocomputing **489**, 345–365 (2022)
4. Martinez, J., Black, M.J., Romero, J.: On human motion prediction using recurrent neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2891–2900 (2017)
5. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. LNCS, vol. 11700. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28954-6
6. Nembrini, S., König, I.R., Wright, M.N.: The revival of the Gini importance? Bioinformatics **34**(21), 3711–3718 (2018)
7. Yang, B., et al.: Human pose prediction using interpretable graph convolutional network for smart home. IEEE Trans. Consum. Electron. (2023)

Semantic-Based Loco-Manipulation for Human-Robot Collaboration in Industrial Environments

Federico Rollo[1,2,3(✉)], Gennaro Raiola[1], Nikolaos Tsagarakis[2], Marco Roveri[3], Enrico Mingo Hoffman[4], and Arash Ajoudani[2]

[1] Autonomous Systems and Robotics, Leonardo Labs, Genoa, Italy
{federico.rollo,gennaro.raiola}@leonardo.com
[2] HHCM & HRII, Istituto Italiano di Tecnologia, Genoa, Italy
{federico.rollo,nikolaos.tsagarakis,arash.ajoudani}@iit.it
[3] Industrial Innovation, DISI, Università di Trento, Trento, Italy
{federico.rollo,marco.roveri}@unitn.it
[4] Université de Lorraine, CNRS, Inria, LORIA, Villers-lès-Nancy, France
enrico.mingo-hoffman@inria.fr

Abstract. Robots with a high level of autonomy are increasingly requested by smart industries. A way to reduce the workers' stress and effort is to optimize the working environment by taking advantage of autonomous collaborative robots. A typical task for Human-Robot Collaboration (HRC) which improves the working setup in an industrial environment is the *"bring me an object please"* where the user asks the collaborator to search for an object while he/she is focused on something else. As often happens, science fiction is ahead of the times, indeed, in the *Iron Man* movie, the robot *Dum-E* helps its creator, *Tony Stark*, to create its famous armours. The ability of the robot to comprehend the semantics of the environment and engage with it is valuable for the human execution of more intricate tasks. In this work, we reproduce this operation to enable a mobile robot with manipulation and grasping capabilities to leverage its geometric and semantic understanding of the environment for the execution of the *Bring Me* action, thereby assisting a worker autonomously. Results are provided to validate the proposed workflow in a simulated environment populated with objects and people. This framework aims to take a step forward in assistive robotics autonomy for industries and domestic environments.

Keywords: Semantic Loco-manipulation · Human-Robot Collaboration · Semantic Mapping · Assistive Robotics e

1 Introduction

Industries are increasingly demanding robotization to improve efficiency and simplify the complex tasks performed by human operators *e.g.* by cooperation and collaboration. Numerous examples of Human-Robot Collaboration tasks can

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 55–59, 2024.
https://doi.org/10.1007/978-3-031-76428-8_11

be found in the literature and industry [1,6]. For instance, in [4,5], the authors introduced a person-following application based on visual re-identification and gesture detection to assist humans. Building on the work presented in [3], we can detect and localize objects in the environment and utilize this map to interact with these artifacts.

The primary contribution of this paper is the proposal of an easy-to-use framework for a mobile robotic collaborator to complete the *"bring me"* service, leveraging semantic knowledge to accomplish the task. This work represents a step towards establishing a more comprehensive and adaptable application that can dynamically respond to the worker's needs to provide autonomous assistance.

The methodology is systematically presented in Sect. 2, and in Sect. 3 the experiments and results are exposed and discussed. Finally, in Sect. 4 the conclusive statements are reported along with some directions for future improvements.

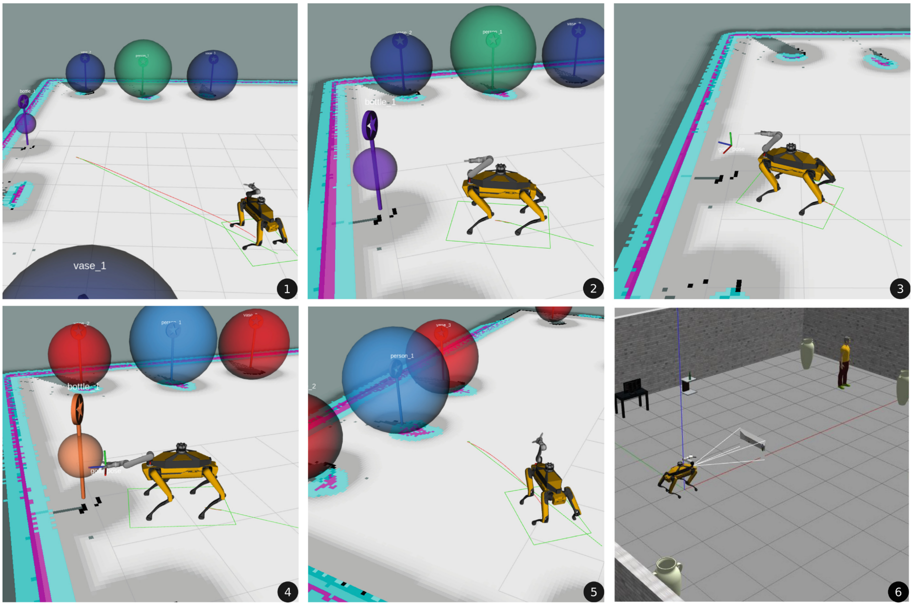


Fig. 1. Autonomous actions held during the execution of the application. Image 1 represents the navigation towards the requested object. In image 2 the object is approached, while in image 3 the object pose is estimated. In images 4 and 5 the object is picked and brought to the person. Image 6 shows the whole simulated world.

2 Method

To manage the application workflow we used a Behaviour Tree[1] (BT). The BT is composed of a sequence of 6 actions: *navigate_to_object*, *approach_object*,

[1] Behaviour tree library: https://py-trees.readthedocs.io/en/devel/.

pose_estimation, *pick_object*, *bring_to_user* and *release_object* and a safety action *abort* which is activated when one of the other actions fails. Most of these actions are represented in Fig. 1. The BT awaits till the user makes a request, which is defined as the string associated with the artifact, *e.g.*, bottle_1. After the request is received the BT starts ticking and executes the behaviours hereafter explained.

navigate_to_object: the rough positions of both the requested object and the target person are already known, as documented in the previous mapping process [3]. This action involves taking the object's position and sending a goal to the robot in an obstacle-free region in front of the required object, allowing the robot to calculate its path and navigate towards the object.

approach_object: the object approaching is necessary to facilitate a smooth picking action by refining the robot's position. This refinement is required due to the imprecise goal position received in the previous step. During this phase, the robot approaches the artifact, positioning it at the centre of the image and reducing the distance ensuring that the bottle falls within its manipulability space. To accomplish this, the robot employs an instance segmentation network to segment the object and uses the resulting mask to crop both the RGB and Depth images. Thereafter, a point cloud is built using the masked depth image and the camera's proprioceptive parameters. For each non-zero pixel of the masked depth, the translation vector ${}^{\mathbf{B}}\mathbf{t}_{\mathbf{p}}$ from the robot base frame $\mathbf{B}$ to the point $\mathbf{p}$ is computed as ${}^{\mathbf{B}}\mathbf{t}_{\mathbf{p}} = {}^{\mathbf{B}}\mathbf{t}_{\mathbf{C}} + {}^{\mathbf{C}}\mathbf{t}_{\mathbf{p}}$, where ${}^{\mathbf{B}}\mathbf{t}_{\mathbf{C}}$ is the translation vector from the base frame $\mathbf{B}$ to the camera frame $\mathbf{C}$ and ${}^{\mathbf{C}}\mathbf{t}_{\mathbf{p}}$ is the translation vector from camera frame $\mathbf{C}$ to the generic 3D point $\mathbf{p}$ and is computed as:

$$
{}^{\mathbf{C}}\mathbf{t}_{\mathbf{p}} = \begin{bmatrix} \frac{1}{f_x} & 0 & -\frac{p_x}{f_x} \\ 0 & \frac{1}{f_y} & -\frac{p_y}{f_y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} d_C \ , \tag{1}
$$

where f_x, f_y, p_x and p_y are the camera intrinsic parameters *i.e.*, the focal lengths and principal points of the camera along x and y image axis, u and v are the depth pixel position along x and y axis and d_C is the depth value of the considered pixel. With this point cloud, we can compute the object's centroid $\mathbf{o}$ expressed in the base frame $\mathbf{B}$, denoted as ${}^{\mathbf{B}}\mathbf{t}_{\mathbf{o}}$, and two key metrics: the 3D Euclidean distance to the robot base, denoted as d_{obj}, and the heading angle between the robot and the object, referred to as θ_{obj}, with the equation:

$$
\theta_{obj} = atan2(y, x) \tag{2}
$$

where x and y are the coordinates of the centroid ${}^{\mathbf{B}}\mathbf{t}_{\mathbf{o}}$ (the xy plane is parallel to the ground floor), and $atan2$ is the arc tangent function which takes into account the quadrant of the tangent. We then used two linear proportional controllers (*e.g.*, Eq. 3) to regulate the robot base position sending velocity commands (linear velocity along the x-axis v_x, angular velocity around the z-axis ω_z) using as reference the distance d_{opt} and the heading angle θ_{opt}.

$$
\xi = K(\chi - \chi_{opt}) \ , \tag{3}
$$

where ξ is the linear v_x or angular ω_z velocity output, K is the proportional gain, χ is the current distance d_{obj} or the heading angle θ_{obj} and χ_{opt} is the optimal distance d_{opt} or angle reference θ_{opt}.

pose_estimation: once the robot is well positioned, the homogeneous transformation $^{\mathbf{B}}\mathbf{H}_{\mathbf{G}}$ of the grasping pose $\mathbf{G}$ expressed in the base frame $\mathbf{B}$, is computed. Using the instance segmentation neural network and the RGB-D camera we compute the translation vector $^{\mathbf{B}}\mathbf{t}_{\mathbf{G}}$ from base frame $\mathbf{B}$ to the grasping pose $\mathbf{G}$ in the same way as $^{\mathbf{B}}\mathbf{t}_{\mathbf{o}}$ in the *approach_object* behaviour. We then set the rotation matrix $^{\mathbf{B}}\mathbf{R}_{\mathbf{G}}$ between the robot base $\mathbf{B}$ and the grasping pose $\mathbf{G}$ equal to an identity matrix $\mathbf{I}_{\mathbf{3}\times\mathbf{3}}$ to obtain the homogeneous transformation:

$$^{\mathbf{B}}\mathbf{H}_{\mathbf{G}} = \begin{bmatrix} ^{\mathbf{B}}\mathbf{R}_{\mathbf{G}} & ^{\mathbf{B}}\mathbf{t}_{\mathbf{G}} \\ 0\ 0\ 0 & 1 \end{bmatrix} \tag{4}$$

pick_object: in this phase, to move the robot gripper in the final pose, the robot receives the desired grasping homogeneous transformation $^{\mathbf{B}}\mathbf{H}_{\mathbf{G}}$ and roto-translates it in the end-effector frame $\mathbf{E}$ as:

$$^{\mathbf{E}}\mathbf{H}_{\mathbf{G}} = {^{\mathbf{E}}\mathbf{H}_{\mathbf{B}}}{^{\mathbf{B}}\mathbf{H}_{\mathbf{G}}} \tag{5}$$

In our case, the simulation uses an optimization-based whole-body inverse dynamics controller [2] to move the arm towards the target pose. Still, this step depends on the specific robot configuration used.

When the robotic end-effector is in the picking position, the gripper can grasp to pick up the artifact, and the arm can then move to a folded position for transportation.

bring_to_user: once the artifact is grasped, the robot brings it to the user navigating in the environment.

release_object: finally, when the robot reaches the user position it gives the requested object to the user.

abort: the abort status is essential for managing unexpected behaviours. All goals or requests sent are halted in this state, and the robot enters an idle state, awaiting further commands. The mission is deleted, and the robot promptly reports the error to the user.

3 Experiments

Experiments were conducted in simulation using WoLF [2], a framework that allows the simulation and control of a quadruped robot with an arm attached to it. The experiments were run on a notebook with an Intel® Core^{tm} i9-11950H processor and an NVIDIA Geforce RTX 3080 Laptop GPU. For the robot motion planning and navigation in the environment, we used the ROS Navigation Stack[2]

[2] ROS Navigation Stack: http://wiki.ros.org/navigation.

and YOLOv8[3] with pre-trained weights as the instance segmentation network. The code is implemented in *C++* and *Python* and integrated with *ROS*.

In the experiments, the robot is asked to bring an object to one of the people in the simulation (e.g. bring a bottle to the person with id 1). The experiment is repeated for different world configurations where the people and the objects are randomly moved (see an example in image 6 of Fig. 1). The proposed pipeline always completed the task following the steps presented in Sect. 2.

4 Conclusion

In this article, we have introduced a semantic loco-manipulation framework for object retrieval. This application uses high-level semantic scene understanding to enable a robot assistant to search for and bring objects that are not nearby and whose locations may be unknown to the user. It is a foundational platform for enhancing robotic assistance within industrial environments, where robots work alongside human operators and actively participate in the workplace community. Several future developments for this framework include the generalization of the pose estimation to handle more complex situations and objects, making experiments on a real robot and improving the reactivity of the BT, *e.g.* if the robot loses the grasp of the object the BT restart autonomously from the *pose_estimation* behaviour.

References

1. Lorenzini, M., Lagomarsino, M., Fortini, L., Gholami, S., Ajoudani, A.: Ergonomic human-robot collaboration in industry: a review. Front. Robot. AI **9**, 262 (2023)
2. Raiola, G., Focchi, M., Hoffman, E.M.: Wolf: the whole-body locomotion framework for quadruped robots. arXiv preprint arXiv:2205.06526 (2022)
3. Rollo, F., Raiola, G., Zunino, A., Tsagarakis, N., Ajoudani, A.: Artifacts mapping: multi-modal semantic mapping for object detection and 3D localization. In: 2023 European Conference on Mobile Robots (ECMR), pp. 1–8. IEEE (2023)
4. Rollo, F., Zunino, A., Raiola, G., Amadio, F., Ajoudani, A., Tsagarakis, N.: FollowMe: a robust person following framework based on visual re-identification and gestures. In: 2023 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO), pp. 84–89. IEEE (2023)
5. Rollo, F., Zunino, A., Tsagarakis, N., Hoffman, E.M., Ajoudani, A.: CARPE-ID: Continuously adaptable re-identification for personalized robot assistance. arXiv preprint arXiv:2310.19413 (2023)
6. Vysocky, A., Novak, P.: Human-robot collaboration in industry. MM Sci. J. **9**(2), 903–906 (2016)

[3] YOLOv8: https://github.com/ultralytics/ultralytics.

AR Solution for Indoor Drone Motion Forecasting

Imre Paniti[1,2](✉), János Nacsa[1,2], Erik Tóth[1], and József Tóth[3]

[1] EPIC Center of Excellence in Production Informatics and Control, HUN-REN SZTAKI, Kende Street 13-17, Budapest 1111, Hungary
{imre.paniti,janos.nacsa,erik.toth}@sztaki.hun-ren.hu
[2] Széchenyi István University, Egyetem Tér 1, Győr 9026, Hungary
[3] HEPENIX Ltd., Petőfi S. U. 39, Diósd 2049, Hungary
jozsef.toth@hepenix.hu

Abstract. In case of indoor drone applications three different areas are classified: inventory management, indoor intra-logistics and inspection & surveillance.

For each drone movement there is an optimal trajectory where the time for reaching point B from point A can be minimized. This optimal trajectory cannot be executed in case a human is blocking the route. In order to avoid collisions with the drone, either the drone trajectory has to be modified in real-time (which might cost additional time and energy in case the drone delivers a heavy object) or the human operator has to be warned with a pre-defined understandable signal so he/she can modify his/her movement in time. In this paper, the implementation of an Augmented Reality solution (previously tested in an industrial relevant environment on a collaborative robot) using a micro drone is presented.

Keywords: Augmented Reality · indoor drone application · motion tracking

1 Introduction

1.1 Motivation

According to MarketsandMarkets [1] the drone logistics and transportation market is projected to reach 17.881 Billion USD by 2030, at a compound annual growth rate (CAGR) of 55.1% during the forecast period. Solutions can be divided into warehousing, shipping, infrastructure and software. Furthermore, "The COVID-19 outbreak has proven to be a boon for the drone industry" [2]. The Warehouse Robotics Market was valued at 6.74 billion USD in 2023 and is expected to reach 15.22 billion USD by 2028 and grow at a CAGR of 17.70% over the forecast period (2023–2028) as stated in [3].

1.2 AR Solutions for Drones

The research in [4] introduced a methodology for remote planning and control of multiple unmanned aerial vehicles (UAVs) using a mobile AR application, enabling engineers

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 60–64, 2024.
https://doi.org/10.1007/978-3-031-76428-8_12

to set target points and optimize drone paths with an algorithm. The AR application visualizes UAV flight paths, identifies obstacles, and automates navigation, resulting in a 37% reduction in time compared to manual navigation tests in a machine shop setting. However, a key challenge is ensuring the safety of shop floor technicians, necessitating advanced techniques to consider human motions within the environment.

The “Drone Brush” application described in [5] introduces a mixed reality system for drone path planning through hand gestures using the HoloLens 2 (HL2). This interface enables the creation of drone paths with immersive hand gestures and utilizes the HL2 spatial map to preemptively detect potential collisions. The system supports dynamic path validation, potentially allowing simpler drones without real-time collision avoidance to execute pre-planned paths, with a proof-of-concept navigation system implemented for the Parrot ANAFI drone. The drawback of the system is that the full length of the pre-defined path remains visible for the user which is visually distracting.

The research in [6] describes the creation of a virtual and AR environment for aiding in the simple assembly, maintenance, and autonomous control of UAVs. The results demonstrate the effectiveness of integrating virtual and augmented reality systems in control processes to reduce errors resulting from insufficient knowledge about UAV assembly.

The work presented in [7] introduces ‘PinpointFly’, an egocentric drone interface that enables precise control of a flying drone through position-control interactions on a see-through mobile AR platform. Four motion-control techniques were designed and implemented in a proof-of-concept prototype using off-the-shelf devices and aruco markers.

2 System Description

In this research the focus is on a solution targeting warehouse drone applications where humans share the space with drones. Before tests in an operation or industrial relevant environment (Technology Readiness Level 7 or 6, respectively) experiments have to be carried out in a laboratory (TRL5). Previous experiments showed that a system using a digital twin of a human with a pre-played movement of a robotic arm model can warn the operator with a haptic signal to avoid real collisions after a virtual one as claimed in [8] and in [9]. The adaptation idea of this solution for drones was presented first at the Poster Session of the European Robotics Forum 2022, but a stable version was tested with a UR3e on TRL6 at Hepenix Ltd., an established integrator since 1991 and an experienced supplier in the automotive, consumer and nuclear industries. The architecture of the system is described in Fig. 1

The above mentioned solution can be used with Augmented Reality (AR) glasses like HoloLens 2 or with other devices using Unity v2021.3 and the Mixed Reality Toolkit. First a pre-defined path of the drone has to be created (e.g. using Skybrush, an open-source drone show and swarm management framework) for the drone controller and for the digital twin of the drone (in our case the Drone Model was given in FBX format). ROS serves was used with ROS nodes as a middleware to forward the position data of the drone from the OptiTrack Prime indoor motion tracking system. This system has a high 3D accuracy ($\pm$ 0.20 mm) with a high frequency sampling and 4.2 ms latency.

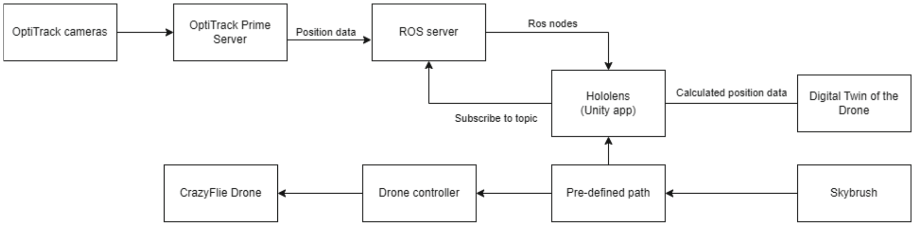


Fig. 1. Architecture of the System

The HoloLens 2 subscribes to the appropriate ROS topic through which it gets the data provided by the OptiTrack Prime, and uses it to move the virtual twin of the drone. This data is stored on the HoloLens 2 device to be used for movement prediction.

Tests above 1 m from the ground on a square trajectory (side length: 2 m) were carried out to obtain the accuracy of the applied controller, as in [10] where detailed description of the hardware and software elements for drone can be found. During the tests the maximum difference between expected and real positions were 0.063, 0.069 and 0.06 m respectively in x, y and z coordinates (see Fig. 2).

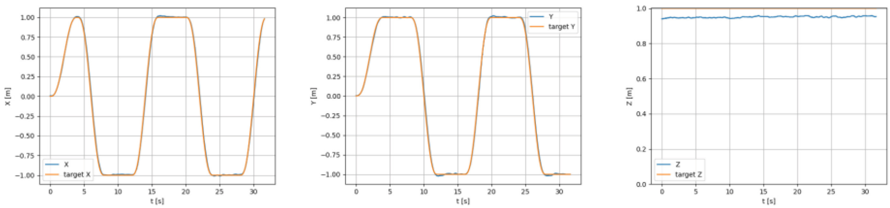


Fig. 2. Position error in x, y, z coordinates

Altogether this resulted in a maximum 0.11 m error in the position. The biggest deviation in the Z coordinates is due to the not insignificant weight of the SD card writing module mounted on the drone.

The system estimates the future position of the digital drone based on its previous positions and corresponding timestamps. It uses numerical methods to calculate derivatives and predict the object's position at a given time. The predictions consider the first and second derivatives, and there are conditions to handle edge cases and to ensure the accuracy of the predictions. The first derivate can be viewed as the speed of the drone and the second derivative as the acceleration, thus using the Taylor series [11] up to the second derivative provides sufficiently accurate predictions with an error that is approximately equal to the 3rd degree part of the series. The data is continuously updated as the drone moves, therefore the prediction is based on the most recent position of the drone. As the drone follows a predefined path which is available for the HoloLens 2 this prediction will be accurate unless the drone crashes or deviates from the path significantly.

3 Result of the Tests

For safety reasons the first tests of the system has been executed with a simple path (take off, horizontal movement and landing). A video in [12] shows the full test. The forecasted movement of the drone was started and visualized by the digital twin of the drone when the real CrazyFlie 2.1 took off and reached a high of 5 cm from the ground.

Uncertainties like stabilization movements from turbulences and correction sub-paths are not visualized as the forecasted movement of the drone is assumed to be ideal. However, each target point can be visualized in real-time and with prompt timing initialization the forecasted movement can be used to warn the workers to avoid potential collisions with the drone. Figure 3 shows a test with HoloLens 2, CrazyFlie 2.1 micro drone and OptiTrack Prime.

Fig. 3. Augmented Reality test with CrazyFlie 2.1 and OptiTrack Prime indoor motion tracking system mounted on the ceiling, where "A" marks the real drone, "B" the pre-played digital twin of the drone and "C" is the human observer.

After successful TRL5 tests the system is ready for further examination in an industrial relevant environment, also with bigger drones. Experiment with workers at the shop floor of Hepenix are planned where the drone must fly below 2 m from the ground to scan the codes of the stored items. The warehouse is in the same area as the assembly plant and it is not economical to empty the plant during each scanning process.

4 Conclusion

Autonomous indoor drone applications are unregulated and in most cases the flight area have to be clear from human workers. In the future the co-existence of drones and humans in the same workspace will be possible, but for that bi-directional signals and path modifications have to be used in real-time. In this work the authors presented a system which is capable to forecast the movement of a drone via an Augmented Reality device. Successful tests in a laboratory environment have been documented on video to show the maturity of the system.

Acknowledgement. This research has been supported by the ED_18-2-2018-0006 grant on a "Research on prime exploitation of the potential provided by the industrial digitalization".

The research in this paper was (partially) supported by the European Commission through the H2020 project EPIC (https://www.centre-epic.eu/) under grant No. 739592.

Furthermore, the research was supported by the European Union within the framework of the National Laboratory for Artificial Intelligence. (RRF-2.3.1-21-2022-00004).

References

1. Market forecast report of MarketsandMarkets, https://www.marketsandmarkets.com/Market-Reports/drone-logistic-transportation-market-132496700.html. Accessed on 14 January 2024
2. Report of Mordor Intelligence on drone logistics and transportation market, https://www.mordorintelligence.com/industry-reports/drone-logistics-and-transportation-market. Accessed on 14 January 2024
3. Report of Mordor Intelligence on warehouse robotics or material handling robotics market, https://www.mordorintelligence.com/industry-reports/warehouse-robotics-or-material-handling-robotics-market. Accessed on 14 January2024
4. Mourtzis, D., Angelopoulos, J., Panopoulos, N.: Unmanned Aerial Vehicle (UAV) path planning and control assisted by Augmented Reality (AR): the case of indoor drones. Int. J. Prod. Res. 1–22 (2023)
5. Angelopoulos, A., et al.: Drone brush: Mixed reality drone path planning. 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE (2022)
6. Misse, E.S., et al.: Augmented reality system for the assistance of unmanned aerial vehicles. 2020 15th Iberian Conference on Information Systems and Technologies (CISTI). IEEE (2020)
7. Chen, L., et al.: Pinpointfly: an egocentric position-control drone interface using mobile ar. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021)
8. Paniti, I., Nacsa, J., Kovács, P., Szűr, D.: VR and depth camera based human-robot collision predictor system with 3-finger gripper assisted assembly device. In: 23rd International Symposium on Measurement and Control in Robotics (ISMCR). IEEE, Piscataway (NJ), pp. 1–5 (2020). ISBN 9780738142692, https://doi.org/10.1109/ISMCR51255.2020.9263758
9. Paniti, I., Nacsa, J., Kovács, P., Szűr, D.: Human-robot collision predictor for flexible assembly. ACTA IMEKO **10**(3), 72–80 (2021). ISSN 0237-028X, https://doi.org/10.21014/acta_imeko.v10i3.107
10. Antal, P., Péni, T., Tóth, R.: Backflipping with miniature quadcopters by gaussian-process-based control and planning. IEEE Trans. Control Syst. Technol. **32**(1), 3–14 (2024). https://doi.org/10.1109/TCST.2023.3297744
11. Foy, W.H.: Position-location solutions by taylor-series estimation. In: IEEE Transactions on Aerospace and Electronic Systems vol. AES-12, no. 2, pp. 187–194 (1976). https://doi.org/10.1109/TAES.1976.308294
12. Video of a drone movement forecasting test: https://www.youtube.com/watch?v=lHQ05yhdRKA. Accessed on 14 January 2024

The Critical Role of Effective Communication in Human-Robot Collaborative Assembly

Davide Ferrari(✉) and Cristian Secchi

Department of Sciences and Methods of Engineering, University of Modena and Reggio Emilia, Reggio Emilia, Italy
{davide.ferrari95,cristian.secchi}@unimore.it
https://dismi.unimore.it

Abstract. In the rapidly evolving landscape of Human-Robot Collaboration (HRC), effective communication between humans and robots is crucial for complex task execution. Traditional request-response systems often lack naturalness and may hinder efficiency. This study emphasizes the importance of adopting human-like communication interactions to enable fluent vocal communication between human operators and robots simulating a collaborative human-robot industrial assembly. We propose a novel approach that employs human-like interactions through natural dialogue, enabling human operators to engage in vocal conversations with robots. Through a comparative experiment, we demonstrate the efficacy of our approach in enhancing task performance and collaboration efficiency. The robot's ability to engage in meaningful vocal conversations enables it to seek clarification, provide status updates, and ask for assistance when required, leading to improved coordination and a smoother workflow. The results indicate that the adoption of human-like conversational interactions positively influences the human-robot collaborative dynamic. Human operators find it easier to convey complex instructions and preferences, resulting in a more productive and satisfying collaboration experience.

Keywords: human robot communication · collaborative assembly · human robot collaboration

1 Introduction

The increasing integration of robots into industrial settings [1] has led to a more intricate landscape of collaboration between humans and machines [2]. This new era of human-robot interaction (HRI) presents numerous opportunities to enhance work efficiency and productivity. However, it also requires a thoughtful examination of how to optimize communication processes among the involved participants [3]. Effective communication between humans and robots,

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 65–69, 2024.
https://doi.org/10.1007/978-3-031-76428-8_13

which mirrors the importance of communication in Human-Human Collaboration (HHC) [4], represents a crucial factor in determining the success of these collaborative interactions.

Current approaches often fall short of achieving the dynamic, bidirectional, and proactive communication characteristic of human interactions; traditionally, in a collaborative assembly scenario within an industrial environment, humans and robots have predefined tasks and interactions, with communication limited to occasional user requests utilizing a simple request-response mechanism [5], providing direct commands to the robot. In [6] communication strategies have been employed to provide operators with insights into the intentions of robots to fosters coexistence and trust between humans and machines but also ensures that both parties remain well-informed about the planned actions of the robot. In [7] a multimodal communication approach is used to enable unidirectional commands from the operator to the robot. While the communication in these cases is primarily unidirectional, it helps to reduce the potential for misunderstandings or unsafe actions. However, it's important to note that unidirectional communication has limitations, as it does not facilitate the exchange of feedback or convey essential information that can be crucial for promoting both safe and efficient collaboration between humans and robots. Conversely, in [8], voice communication is structured to empower the robot with the capability to initiate dialogues using a bidirectional voice communication framework, in order to let the robot communicate some problems and error that may occur during the collaborative job.

While these approaches hold promise in enhancing human-robot communication, there are instances where relying solely on unidirectional or robot-initiated requests may prove to be limiting and inefficient, especially when confronted with complex and dynamic tasks. The main objective of this article is to implement a human-like [9] natural and bidirectional vocal communication architecture within a collaborative human-robot assembly task involving an industrial component, developing more complex and natural communication and seeking to emulate what occurs in collaborative human-human assembly, leading to improvements in task performance and collaboration efficiency. The robot's capacity to engage in meaningful vocal conversations allows it to seek clarification, offer status updates, and request assistance when necessary, resulting in enhanced coordination and a smoother workflow.

In the following sections of this paper, we will introduce our framework that incorporates natural communication within a collaborative assembly of an industrial component. Furthermore, we will conduct a comparative experiment to validate our proposal and illustrate how more sophisticated communication can bring benefits in collaboration and user experience. The contributions of this article encompass:

- An architecture for natural communication in a collaborative assembly task involving an industrial component.
- An experimental validation through a collaborative experiment to assess the validity of the proposed framework.

2 Natural Collaborative Assembly Framework

Let's consider a scenario of Human-Robot Collaborative Assembly involving an industrial component, where both the robot and the human operator are engaged in a complex task that requires shared knowledge and mutual assistance. In this context, effective communication between the operator and the robot takes on primary importance. They need to be able to exchange information regarding task progress, request tools or components, seek assistance or support, and engage in dynamic, context-aware conversations. Building upon this, our proposed architecture aims to incorporate advanced natural language processing (NLP) techniques, leveraging pre-trained deep reinforcement learning models capable of generating a coherent and natural dialogue flow, thus emulating human capabilities. This will enable the robot to understand and respond to user requests in a more natural and context-aware manner.

To facilitate this natural communication, our proposed architecture incorporates a commercial voice assistant into the Human-Robot Collaboration job. This integration enables reciprocal information exchange between the robot and the operator through a voice communication channel structured in the form of conversations. Each conversation, as depicted in Fig. 1, consists of a series of turn-taking dialogues, starting with the user's turn and then transitioning to the robot's turn. Within each dialogue, the user's requests are associated with predefined example phrases (utterance sets), which can be customized with variables or catalogs if required. The robot can respond directly, request additional information to fill in missing variable slots, or transmit the message to the back-end using a JSON-Request. Depending on the nature of the request, the system can call upon APIs that may reference external objects or trigger subsequent dialogues, allowing the conversation with the human to unfold in a contextually rich and meaningful manner.

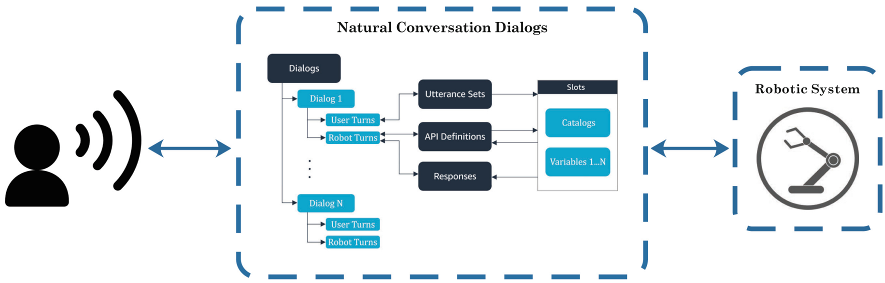


Fig. 1. Natural Conversation Framework

3 Experimental Validation

The experimental validation[1] involved a comparative experiment that simulated a collaborative assembly task. In this scenario, an operator worked together a UR10e collaborative manipulator to assemble an industrial planetary gearbox. The components and tools required for the assembly were strategically positioned within the shared workspace, accessible to both the robot and the operator. Leveraging voice communication, the operator had the capability to request tools, components, or assistance, while the robot, through dialogues, could provide details, address issues, propose alternatives, and provide assistance. The experiment, carried out on a sample of 10 participants, randomizing the execution order of the two experimental setups to minimize the potential influence of a learning effect, aimed to compare the proposed conversation collaborative assembly architecture with the traditional industrial assembly job based on fixed task and a minimal request-response communication structure.

3.1 Implementation Details

The architecture was implemented by integrating into the ROS framework a custom Amazon Alexa Conversations [10] skill, a deep learning-based approach that employs API calls to manage multi-turn dialogues between Alexa and the user, enabling more natural and human-like interactions. The skill's back-end was locally hosted, facilitating seamless integration with ROS by leveraging Microsoft Azure's HTTP Trigger Functions. Furthermore, to enable direct interaction with Alexa APIs, a Node-RED flow was developed, a web service for logical path programming that allows event management and the initiation of conversations by invoking specific dialogue APIs.

3.2 Analysis of the Results

To evaluate the effectiveness of our architecture, we measured execution times and robot downtimes, and we collected the user feedback via a questionnaire consisting in five ratings on a scale from 0 to 10, covering key aspects: *Clarity of Communication*, *Naturalness of Communication*, *Ease of Interaction*, *Stress During Communication*, and *Overall Satisfaction*. Figure 2 shows the results, with the comparative experiment in red and the proposed architecture in blue. The results underscore a notable difference: the proposed architecture received an average score of approximately 8.8/10, while the state-of-the-art approach averaged around 6.1/10. These findings indicate a significant enhancement in the user experience when utilizing our architecture, marked by improved clarity, ease of use, and reduced stress resulting from more natural communication. Moreover, the comparison of execution times and robot downtimes has demonstrated how the implementation of simple and natural communication within a collaborative assembly task can significantly reduce both execution times (22%) and robot downtimes (73%), resulting in increased efficiency and collaboration.

[1] Detailed video of the experiment at: https://doi.org/10.5281/zenodo.10105470..

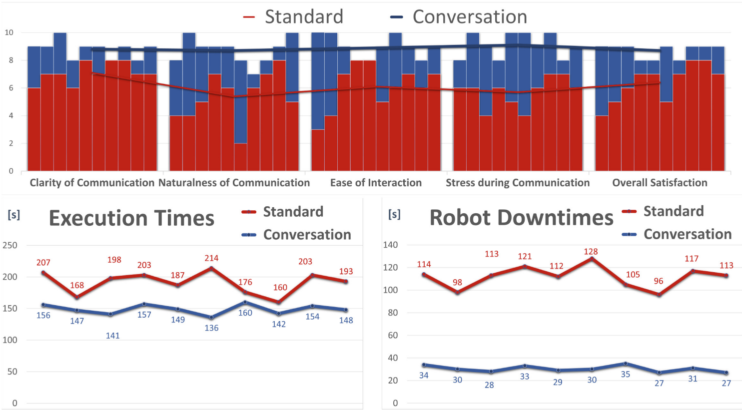


Fig. 2. Measured Times and Questionnaire Results

References

1. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human-robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics **55**, 248–266 (2018)
2. Baratta, A., Cimino, A., Gnoni, M.G., Longo, F.: Human robot collaboration in industry 4.0: a literature review. Procedia Comput. Sci. **217**, 1887–1895 (2023)
3. Mukherjee, D., Gupta, K., Najjaran, H.: A critical analysis of industrial human-robot communication and its quest for naturalness through the lens of complexity theory. Front. Robot. AI **9** (2022)
4. Mojtahedi, K., Whitsell, B., Artemiadis, P., Santello, M.: Communication and inference of intended movement direction during human-human physical interaction. Front. Neurorobot. **11** (2017)
5. Mavridis, N.: A review of verbal and non-verbal human-robot interactive communication. Robot. Auton. Syst. **63**, 22–35 (2015)
6. Grushko, S., Vysocký, A., Oščádal, P., Vocetka, M., Novák, P., Bobovský, Z.: Improved mutual understanding for human-robot collaboration: combining human-aware motion planning with haptic feedback devices for communicating planned trajectory. Sensors **21**(11) (2021)
7. Rosen, E., Whitney, D., Fishman, M., Ullman, D., Tellex, S.: Mixed reality as a bidirectional communication interface for human-robot interaction. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (2020)
8. Ferrari, D., Benzi, F., Secchi, C.: Bidirectional communication control for human-robot collaboration. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 7430-7436. IEEE Press (2022)
9. Marin Vargas, A., Cominelli, L., Dell'Orletta, F., Scilingo, E.P.: Verbal communication in robotics: a study on salient terms, research fields and trends in the last decades based on a computational linguistic analysis. Front. Comput. Sci. **2** (2021)
10. Acharya, A.: Alexa conversations: an extensible data-driven approach for building task-oriented dialogue systems (2021)

Time-Optimized Trajectory Planning for Non-prehensile Object Transportation in 3D

Lingyun Chen[1,2(✉)], Haoyu Yu[1], Abdeldjallil Naceri[1], Abdalla Swikir[1,2], and Sami Haddadin[1,2]

[1] Munich Institute of Robotics and Machine Intelligence (MIRMI), Technical University of Munich, Munich, Germany
lingyun.chen@tum.de

[2] Centre for Tactile Internet with Human-in-the-Loop (CeTI), Dresden, Germany

Abstract. Non-prehensile object transportation offers a way to enhance robotic performance in object manipulation tasks, especially with unstable objects. Effective trajectory planning requires simultaneous consideration of robot motion constraints and object stability. Here, we introduce a physical model for object stability and propose a novel trajectory planning approach for non-prehensile transportation along arbitrary straight lines in 3D space. Validation with a 7-DoF Franka Panda robot confirms improved transportation speed via tray rotation integration while ensuring object stability and robot motion constraints.

Keywords: Trajectory planning · non-prehensile object transportation

1 Introduction

With the advancement of robotic technology, robots are finding increasingly widespread applications in industrial production [1]. In the industrial production process, the transportation of objects is undeniably crucial. Transporting objects through non-prehensile manipulation, compared to grasping objects, offers several advantages, including simpler end-effector design, the ability to transport a wider range of objects, and improved transportation efficiency [2].

Transporting objects using non-prehensile manipulation requires a low center of gravity, a large base area, and minimizing acceleration during the transportation process for stability. To address the transportation challenges of unstable objects, in [3], researchers explored the limiting conditions for the maximum acceleration during the transportation of unstable objects on a tray-like end-effector. They achieved planar transportation through trajectory planning based on an S-curve. However, their work does not incorporate the rotation of the tray. Introducing tray rotation at different stages of transportation could evidently enhance the time efficiency of the process. Relevant studies include those in [4], which examine the contact model between the object and the tray, and [5], which

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 70–75, 2024.
https://doi.org/10.1007/978-3-031-76428-8_14

proposes a Model Predictive Control-based approach to track predefined transportation trajectories. The latter ensures non-sliding transportation of the object by considering the friction cone between the object and the tray as a constraint. Both papers discuss the rotation of the tray and the physical model of contact between the object and the tray. However, while similar studies introduce tray rotation, the majority focus their control objectives on enhancing tracking performance for given trajectories. In contrast, our work approaches the problem from a trajectory planning perspective, aiming to improve transportation speed by incorporating the rotational motion of the tray. The motion of the tray and the object in the desired trajectory is illustrated in Fig. 1.

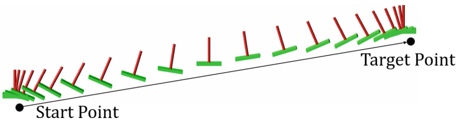


Fig. 1. Illustration of object and tray motion.

2 Method

In Subsects. 2.1 and 2.2, we discuss the physical model between the object and the tray and present the novel planning method for computing the transport trajectory, respectively.

2.1 Physical Modelling

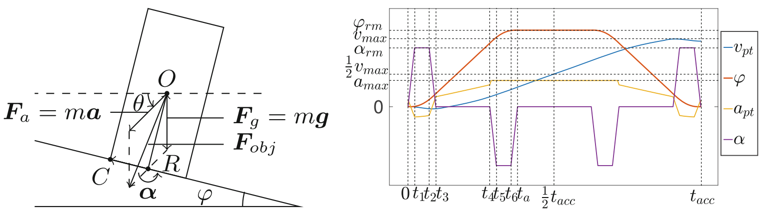
In our setting, it is presumed that both the object and the tray are rigid bodies. The object under consideration possesses a slender, uniform cylindrical morphology, and in the analysis, the effects of air resistance on its motion are disregarded. Furthermore, it is assumed that the object exhibits characteristics such as a high center of gravity and a minimal base area, consequently resulting in compromised stability. Under conditions of acceleration, the object is anticipated to undergo tipping prior to the onset of sliding between its base and the tray.

To ensure stability during the transportation process, we aim for no relative motion between the object and the tray. Let the tray rotate around the center of the object's base, and assume at this point there is a fixed joint between the object and the tray, connecting them. According to the 6D rigid body contact model proposed in [6], whether the object tilts or not depends on the location of the pressure center. To maximize the acceleration during the transportation process, we assume that the pressure center is located on the boundary of the contact surface between the object and the tray. Based on the rotational motion state of the tray, the torque $\boldsymbol{\tau} = I\boldsymbol{\alpha}$ acting on the virtual fixed joint can be calculated, where I is the moment of inertia, and $\boldsymbol{\alpha}$ is the rotational angular acceleration. Choosing the tray as the reference frame and analyzing the motion and forces acting on the object as shown in Fig. 2a, Where $\boldsymbol{a}$ is the object's translational acceleration, θ is the target direction, φ is the current rotation angle of the tray, O represents the object's center of gravity, R is the center of the object's base, and C is the chosen center of pressure position. The object is subjected to a resultant force, denoted as $\boldsymbol{F}_{obj}$, which is a vector sum of the forces due to gravity and acceleration, expressed as $\boldsymbol{F}_{obj} = -m\boldsymbol{a} + m\boldsymbol{g}$. The tray applies a force $\boldsymbol{F}_{tray}$ on the object, which is equal in magnitude but opposite in direction to $\boldsymbol{F}_{obj}$, thus $\boldsymbol{F}_{tray} = -\boldsymbol{F}_{obj}$. Additionally, the object is subjected to a centrifugal force $\boldsymbol{F}_r$, calculated as $\|\boldsymbol{F}_r\| = m\|\boldsymbol{\omega}\|^2 \frac{h}{2}$, where ω represents the rotational velocity, r is the radius of the cylinder, and h denotes the cylinder's height. The object's motion, synchronized with that of the tray and rotating around the object's base center, is facilitated by the collective influence of three forces. This dynamic interaction is encapsulated in the equation below:

$$\boldsymbol{\tau} = \overrightarrow{RO} \times \boldsymbol{F}_{obj} + \overrightarrow{RC} \times \boldsymbol{F}_{tray} + \overrightarrow{RC} \times \boldsymbol{F}_r \tag{1}$$

The objective is to determine the maximum translational acceleration of the object under the current tray motion. Therefore, by rearranging Eq. 1, we can obtain:

$$\|\boldsymbol{a}\| = \frac{\frac{I\|\boldsymbol{\alpha}\|}{m} + \frac{h\|\boldsymbol{g}\|}{2}\sin\varphi + r\|\boldsymbol{g}\|\cos\varphi - \frac{r\|\boldsymbol{\omega}\|^2 h}{2}}{\frac{h}{2}\cos\theta\cos\varphi - \frac{h}{2}\sin\varphi\sin\theta - r\sin\varphi\cos\theta - r\cos\varphi\sin\theta} \tag{2}$$



(a) Contact model (b) Acceleration phase of the trajectory.

Fig. 2. Physical model and trajectory illustration.

2.2 Trajectory Planning

In the domain of robotic motion planning, the constraints primarily encompass jerk, acceleration, and velocity parameters. These constraints remain uniform throughout the trajectory, represented as $j \in [-j_{max}, j_{max}]$, $a \in [-a_{max}, a_{max}]$, and $v \in [0, v_{max}]$. Similarly, rotational motion adheres to constraints denoted as $j_r \in [-j_{rm}, j_{rm}]$, $\alpha_r \in [-\alpha_{rm}, \alpha_{rm}]$, and $\omega_r \in [0, \omega_{rm}]$. It's worth noting that robots can achieve high jerks, facilitating rapid acceleration changes. However, since the rotation speed of the tray is comparatively slower, we assume that during the trajectory's acceleration phase, the maximum acceleration an object can reach is primarily dictated by the tray's rotation speed. Therefore, the trajectory planning strategy first generates the rotational motion trajectory of the tray using an S-curve, then calculates the corresponding translational motion trajectory using Eq. 2. The trajectory planning can be summarized in Table 1.

Table 1. Changes in motion states during the trajectory acceleration phase.

	$t < t_1$	$t_1 \le t < t_2$	$t_2 \le t < t_3$	$t_3 \le t < t_4$	$t_4 \le t < t_5$	$t_5 \le t < t_6$	$t_6 \le t < t_a$	$t_a \le t < \frac{1}{2}t_{acc}$
j_r	j_{rm}	0	$-j_{rm}$	0	$-j_{rm}$	0	j_{rm}	0
α	↑	α_{rm}	↓	0	↓	$-\alpha_{rm}$	↑	0
ω	↑	↑	↑	ω_{rm}	↓	↓	↓	0
φ	↑	↑	↑	↑	↑	↑	↑	φ_{rm}
a	Calculated based on (2)							a_{max}
v	Increasing and reaching $\frac{1}{2}v_{max}$ at time $\frac{1}{2}t_{acc}$							

The next step is to ensure that the trajectory does not exceed the robot's motion constraints. It can be observed that at time t_a, the trajectory's acceleration reaches its maximum, while the velocity is half of the maximum speed. Consequently, we can formulate the following conditions:

$$a_{pt}(t_a) \le a_{max}, \tag{3}$$

$$v_{pt}(t_a) \le \frac{v_{max}}{2}. \tag{4}$$

Finally, assuming there is no constant velocity phase in the trajectory, the velocity at time t_a precisely equals the average velocity of the trajectory. By knowing the durations of the acceleration phases t_a and deceleration phases t_b, the following conditions ensure that the trajectory does not exceed the target point.

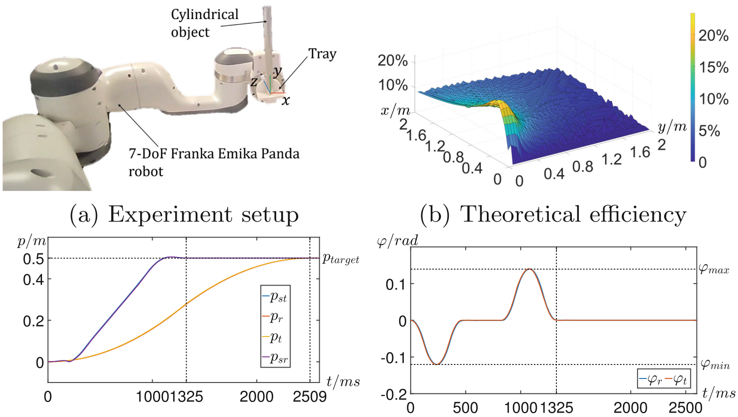
$$2v_{pt}(t_a)(t_a + t_b) \le p_t. \tag{5}$$

Adjusting the duration of the constant velocity phase based on the remaining distance between the trajectory and the target point completes the trajectory planning.

3 Experiment and Results

To test the effectiveness and time efficiency of our method, we conducted a series of experiments using a 7-DoF Franka Emika Panda robot. The setup for these experiments is depicted in Fig. 3a. The following conditions are selected as motion constraints in the experiment: $j_{max} = 6500\,\mathrm{m/s^3}$, $a_{max} =$

$13\,\text{m/s}^2$,$v_{max} = 0.6\,\text{m/s}$, $j_{rm} = 6000\,\text{rad/s}^3$, $a_{rm} = 9\,\text{rad/s}^2$,$v_{rm} = 2.61\,\text{rad/s}$. The tested object is a uniform aluminum cylinder with a radius of $r = 8$ mm and a height of $h = 0.2$ m. We performed theoretical calculations based on these conditions and compared the time taken by our proposed method with the time taken by the method that does not involve tray rotation. The results are shown in Fig. 3b, where x represents the horizontal displacement of the target, and y represents the vertical displacement of the target. Our proposed method can improve time efficiency by up to 25%.



(a) Experiment setup (b) Theoretical efficiency

(c) The target and actual trajectory (d) End-effector rotation

Fig. 3. Experiment results.

The experimental design sets the target distance to $p = 0.5$ m, with the target located above the starting point and orientation angle $\theta = \frac{\pi}{8}$. Due to the assumption in the physical model and control errors of the robot, the experiment gradually reduces the radius of the object's base until a stable trajectory is found. The time required for the trajectory is then compared with the time efficiency of the trajectory planned using S-curves under the same motion constraints.

In the experiments[1], the input radius for stable trajectories obtained are $r_o = 3$ mm without rotation and $r_r = 4$ mm with rotation. The trajectories of the target and the robot's motion are shown in Fig. 3c and Fig. 3d. It can be seen that our proposed new method reduces the time taken by 47.2% compared to the motion trajectory without rotation.

[1] Videos shown in the link.

4 Conclusion

This study introduces a novel trajectory planning approach, integrating tray rotation, for the non-prehensile transportation of unstable objects. Our method significantly accelerates straight-line transportation in three-dimensional space, and its efficacy has been demonstrated through practical implementation on a real robot. In future work, our aim is to enhance the proposed method to enable trajectory planning for the non-prehensile transportation of unstable objects along arbitrary paths.

Acknowledgement. The work was funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy - EXC 2050/1 - Project ID 390696704 - Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden. The work was also supported by the European Union's Horizon 2020 research and innovation programme as part of the project euROBIN under grant no. 101070596.

References

1. Hentout, A., Aouache, M., Maoudj, A., Akli, I.: Human-robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. Adv. Robot. **33**(15–16), 764–799 (2019)
2. Mason,M.T., Lynch, K.M.: Dynamic manipulation. In: Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 1993), vol. 1, pp. 152–159. IEEE (1993)
3. Acharya, P., Nguyen, K.-D., La, H.M., Liu, D., Chen, I.-M.: Nonprehensile manipulation: a trajectory-planning perspective. IEEE/ASME Trans. Mechatron. **26**(1), 527–538 (2020)
4. Tsuji, T., Ohkuma, J., Sakaino, S.: Dynamic object manipulation considering contact condition of robot with tool. IEEE Trans. Industr. Electron. **63**(3), 1972–1980 (2015)
5. Selvaggio, M., Garg, A., Ruggiero, F. Oriolo, G., Siciliano, B.: Non-prehensile object transportation via model predictive non-sliding manipulation control. IEEE Trans. Control Syst. Technol. (2023)
6. Bouchard, C., Nesme, M., Tournier, M., Wang, B., Faure, F., Kry, P.: 6D frictional contact for rigid bodies. In: Graphics Interface (2015)

Behavior Tree Based Robotic Skill Execution for Human Robot Collaboration in Industrial Settings

Sharath Chandra Akkaladevi(✉), Matthias Propst, Kapil Deshpande, Michael Hofmann, and Andreas Pichler

Robotics and Automation Systems, Profactor GmbH, 4400 Steyr-Gleink, Austria
sharath.akkaladevi@profactor.at

Abstract. This paper extensively explores the utilization of behavior tree-based robotic skill execution engines, focusing specifically on their application in industrial settings. By integrating behavior trees into the robotic framework, this research significantly contributes to enhancing the adaptability of robots in dynamic environments. The modularity and reactivity offered by behavior trees play a pivotal role in enabling robots to dynamically adjust their behaviors in response to unforeseen circumstances, especially in the context of extensive human-robot collaboration in industrial scenarios. The demonstrated application of this approach in a real-world assembly scenario utilizes a novel mobile collaborative manipulator with reduced computational power. This real-world implementation not only underscores the practical relevance of behavior tree-based execution engines but also highlights their applicability in human robot collaborative handling of objects. The behavior tree ticking mechanism, even on a less powerful PC, showcases the robustness of the proposed methodology. This work offers insights into the modularity and reactivity inherent in behavior trees, providing a promising avenue for addressing challenges in human-robot collaboration within industrial settings, even when operating under computational constraints.

Keywords: Robot task execution · Behaviour trees · Human Robot collaboration

1 Introduction

In recent years, the field of robotics has witnessed a paradigm shift towards collaborative frameworks, particularly in industrial settings, where human-robot collaboration has become increasingly prevalent. This shift necessitates a profound reevaluation of robotic systems, demanding heightened adaptability, reactivity, and efficiency in the face of dynamic and unpredictable environments. Behavior trees (BT), known for their modularity and reactivity, emerge as a promising solution to enhance the adaptability of robots in dynamic scenarios, especially those characterized by extensive collaboration with human counterparts. The fundamental premise lies in empowering robots to

This work is funded under the EU Project FELICE (EU grant agreement No 101017151).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 76–80, 2024.
https://doi.org/10.1007/978-3-031-76428-8_15

dynamically adjust their behaviors in response to unforeseen circumstances, a critical capability in the context of industrial scenarios marked by constant change. Behavior trees, with their structured and hierarchical representation of tasks, offer a systematic approach to decision-making and action execution. The modularity of behavior trees allows for the seamless integration of new behaviors and the modification of existing ones, facilitating agile responses to evolving situations.

This work contributes the following:

- Demonstrates the applicability of behavior tree-based execution engines in handling object manipulation for human-robot collaboration.
- The practical demonstration of this approach in a real-world assembly scenario involving a novel mobile collaborative manipulator with reduced computational power
- Showcases the modularity and reactivity inherent in behavior trees that can dynamically respond to changing collaborative environments.

2 Behaviour Trees Based Robot Task Execution

In the context of human-robot collaboration (HRC), to enable seamless interaction between humans and robots, the robot first **senses** its environment and the actions of human collaborators, **plans** its tasks accordingly, and **executes** those tasks while considering the dynamic nature of the collaborative environment. Figure 1a shows such the architecture implemented in this work. Given a high-level task plan, this work utilizes a skill-based task execution engine to execute them on the robotic platform. Skill-based robot task execution involves organizing and executing a robot's actions as a set of modular and reusable skills (see 2.1), enabling adaptability to different tasks and dynamic environments. Various methods exist for skill-based robot task execution engines, with common components including skill representation, execution modules, monitoring and verification, feedback systems, planning and sequencing, context awareness, error handling, human-robot interaction interfaces, task planning and scheduling, and execution control. These components work together to form a comprehensive skill-based robot task execution engine, crucial for successful human-robot collaborative scenarios in industrial settings.

Skill-based robot task execution involves organizing and executing a robot's actions as a set of modular and reusable skills, enabling adaptability to different tasks and dynamic environments. There are different techniques to model these skill-based robot task execution engines. Finite State Machines (FSMs) [3] model robot behavior with states and transitions, offering simplicity but limited scalability. Behavior Trees (BTs) [1] provide a hierarchical structure for modular, reactive, and flexible task handling. Other methods include Petri Nets [2], Hybrid Systems, Rule-Based Systems, Scripting Languages, and Machine Learning Approaches [4]. The choice of method depends on specific task requirements and environmental characteristics.

Behavior Trees (BTs) stand out as a favorable approach for human-robot collaborative assembly tasks in industrial applications. BTs' hierarchical and modular nature facilitates modularity, adaptability, and user interface integration. They excel in handling complex assembly scenarios by breaking tasks into manageable subtasks, promoting

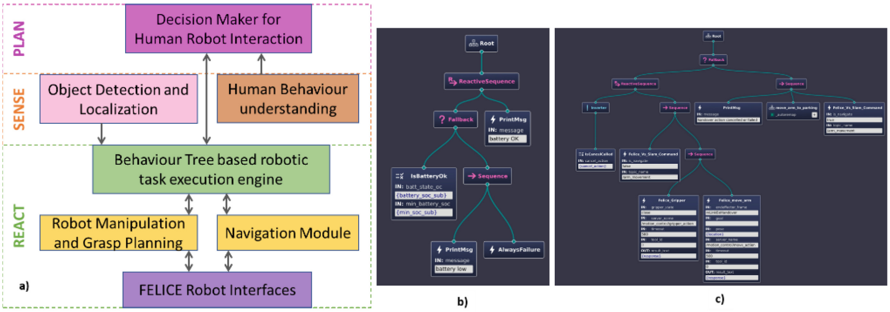


Fig. 1. a) Architecture for human robot collaborative assembly (based on the FELICE architecture [5]) b) basic BT structure c) BT structure for a handover skill

fault tolerance, and ensuring real-time adaptability. The behavior tree ticking mechanism enables continuous evaluation, aligning the robot's actions with the evolving collaborative context. In summary, behavior trees offer a versatile and efficient framework for skill-based robot task execution, particularly well-suited for the complexities of human-robot collaboration in industrial assembly scenarios.

Within the behavior tree structure, skills are meticulously mapped and organized through distinct **node** types. At the top level, the **root node** initiates the tree, representing the overarching robot objective. **Composite nodes**, including **Sequence and Selector nodes**, dictate the arrangement of actions and conditions. Sequence nodes execute child nodes sequentially until one fails, while Selector nodes proceed until one succeeds. **Task nodes**, positioned at the leaf level, encapsulate specific robot actions (skills) like "Grasp Object" or "Release Object.“ **Decorator nodes**, such as Repeater and Condition nodes, modify the behavior of child nodes. For instance, a **Repeater Node** repeats a child node's execution, and a **Condition Node** checks a condition before allowing execution. This structured organization caters to collaborative assembly intricacies, allowing task decomposition into manageable sub-tasks, ensuring modularity. Moreover, the reactive nature of behavior trees facilitates real-time adaptability to environmental or collaboration scenario changes, making them a valuable tool in the execution of complex tasks. Figure 1b and c shown an example of a basic BT and the BT structure for handover task respectively.

2.1 Executed Robotic Skills, Skill Mapping and Execution

Table 1 shows a sub set of skills handled by the BTE to facilitate seamless HRC.

Skills generalization: Basic tree is a subtree which is always present with any skill and each skill is modelled as per the actions they have to perform. Currently skills mentioned in Table 1 are implemented. All the actions that require a longer duration to execute are given inside the reactive sequence. This enables reactivity of the behavior tree for reacting to cancel actions from HRI-DM or when robot battery is low.

Skill Mapping: HRI-DM sends high level robot action to BTE, which checks if any other requests is running then accepts the requests and generates parameters for creating

Table 1. Subset of skills handled by the FELICE robotic platform.

Skill Type	Description
Localize	Trigger the ODL and extract the 6D pose of an object/tool of interest
Pickup	Skill combines localize, reaching for the object and grasping the object
Navigate	Triggering the NM to plan and move the robot from current position to next
Handover	Reach for a position ergonomically safe (w.r.t user) for the tool handover
Release	Release the tool from robot arm on trigger to complete collaborative handover

ODL: Object detection and localization; NM: Navigation Module (see Fig. 1a); High-level plans are sent by the Decision Maker for Human Robot Interaction module (**HRI-DM**) which are then converted to robot skill by the BT task execution engine (**BTE**) and executed on the FELICE robot

a behavior tree xml. The dynamic BT engine (a novelty) creates the blackboard variables required for running a skill and subscribes to condition nodes (e.g., battery status).

Skill Execution: BTE runs the tree ticking at a frequency of 4 Hz, BT executor will generate the new xml as per the params and start executing the requested skill tree. The integration of the modules (see Fig. 1a) is done on an Intel NUC 11 ProKit NUC11TNKv7 PC using Robot Operating Systems (ROS) actions and services to facilitate interaction between different components (see Fig. 1a), allowing the BTE to perform the requested skill on the FELICE robot.

3 Real-World Experiments and Discussion

For the demonstration scenario, a human-robot collaborative tool handover task in a real-world car door assembly line performed at CRF assembly pilot line located in Melfi (Italy) on the FELICE robotic platform (see Fig. 2) is chosen. The robot is equipped with 8 DOF arm with a three-finger gripper, attached to navigation unit and consists of various sensors (cameras, LIDARS). The modules in the plan and act/react layers (Fig. 1) are all implemented on a single NUC PC installed on the robot. Running all the layers on the NUC PC with limited resources (4 cores, 4.8 GHz) has an impact on the performance to counter this, only behavior trees required for execution were loaded dynamically. The sense layer (ODL) is implemented on a separate NUC AGX PC on the robot. As shown in Fig. 2, the high-level robot task of handover task is triggered by HRI-DM and the BTE maps it to a set of robotic skills (Navigate, Localize, Pickup Handover and Release) and executes them on the FELICE robotic platform. HRI-DM communicates the required tool for handover and the current human position, while BTE computes the other parameters required and performs the skill execution. Furthermore, experiments were carried out to activate robot's safe behavior by canceling current running actions and running out the battery (in simulation) thus proving reactivity of behavior trees. Similarly, inclusion of subtrees and fallback subtrees for handling deviations from the standard operations proves modularity of behavior trees.

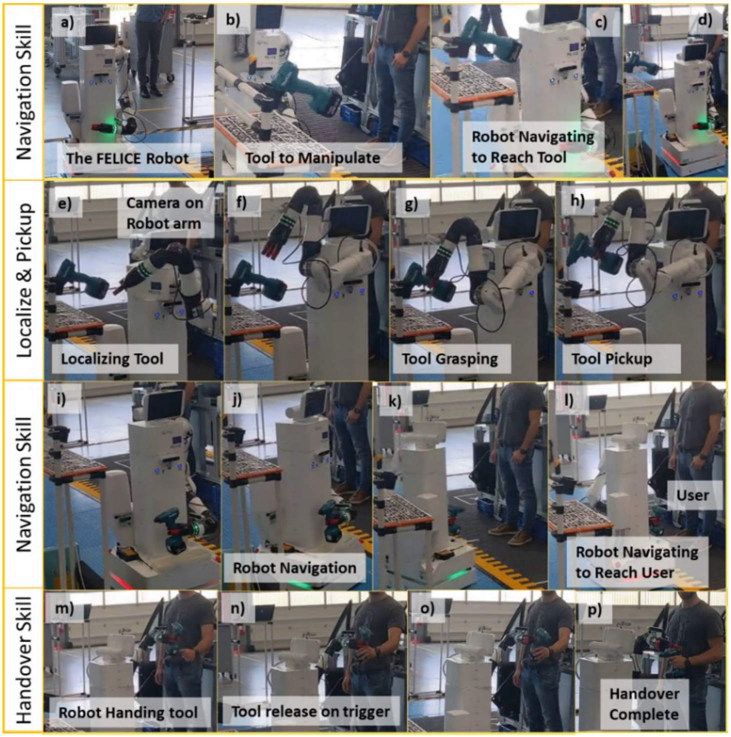


Fig. 2. The real-world experiments where the FELICE robot platform developed by ACCREA, performs the various skills including navigation, localization, pickup and handover tasks

4 Conclusion

The work demonstrates the applicability of BT-based robot task execution engines for object manipulation in human-robot collaboration, showcasing modularity and reactivity in dynamically changing environments, even with limited computational resources. The practical deployment of this approach in a real-world assembly scenario further solidifies its relevance and effectiveness.

References

1. Iovino, M., Scukins, E., Styrud, J., Ögren, P., Smith, C.: A survey of behavior trees in robotics and ai. Robot. Auton. Syst. **154**, 104096 (2022)
2. Casalino, A., et al.: Optimal scheduling of human–robot collaborative assembly operations with time petri nets. IEEE Trans. on Automation Sci. Eng. **18**(1), 70–84 (2019)
3. El Makrini, I., et al.: A hierarchical finite-state machine-based task allocation framework for human-robot collaborative assembly tasks. IROS, pp. 10238–10244. IEEE (2022)
4. Akkaladevi, S.C., et al.: Toward an interactive reinforcement based learning framework for human robot collaborative assembly processes. Front. in Robotics and AI **5**, 126 (2018)
5. FELICE Homepage: https://www.felice-project.eu/. Last accessed 10 November 2023

Test Methods for Passive Exoskeletons for Manufacturing Applications

Cecilia Scoccia(✉), Serenella Terlizzi, Samuele Tonelli, Daniele Costa, and Giacomo Palmieri

Università Politecnica delle Marche, 60131 Ancona, Italy
c.scoccia@staff.univpm.it

Abstract. This study explores the use of passive exoskeletons in modern industries to improve worker well-being and efficiency, particularly in tasks prone to musculoskeletal issues. A protocol has been defined to evaluate two passive exoskeletons (PAEXO back and PAEXO shoulder) through objective and subjective analyses. The first focused on muscle activity, kinematic movements and heart rate; the latter consisted of questionnaires to quantify user impressions. Results indicate that the PAEXO shoulder exoskeleton effectively reduces muscle strain and heart rate while maintaining user comfort. The PAEXO back exoskeleton showed potential but requires protocol refining and a longer familiarization period. Overall, passive exoskeletons hold promise for mitigating musculoskeletal risks and improving worker ergonomics.

Keywords: passive exoskeletons · human factor · motion capture

1 Introduction

Modern industries are adopting human-centred workplace designs and operator 5.0 concepts to customise work environments to improve operators' well-being and workforce efficiency [1]. Passive exoskeletons are emerging as valuable tools for improving the working environment, particularly in tasks involving repetitive or strenuous work that can lead to musculoskeletal issues [2,3], which afflict over 41% of the working population [4,5]. Recent studies show positive effects in the use of exoskeletons [6] for improving workers' biomechanics and researchers are exploring their benefits and developing protocols for their proper utilization [7,8].

In this preliminary study, the authors drafted a protocol to evaluate the effects of two passive exoskeletons during the execution of a repetitive lifting task and an assembly and screwing task by studying muscle activity, kinematic movements and heart rate alongside the subjective impressions of the participants.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 81–86, 2024.
https://doi.org/10.1007/978-3-031-76428-8_16

2 Materials and Methods

For the purposes of this study, two passive exoskeletons have been taken into account: PAEXO back and PAEXO shoulder, developed by Ottobock (Duderstadt, Germany). The first is designed to reduce the load on the lower back during heavy object lifting; the latter targets the user's upper limb strain during the execution of repetitive or prolonged movements, especially with raised arms.

2.1 Obejctive and Subjective Analysis

To explore the potential benefits of the two mentioned devices, an objective and subjective analysis was conducted. The objective analysis focused on three main aspects. Initially, an assessment was made to determine if there were differences in muscle contraction between task execution with and without the exoskeleton. Specifically, for the back the following muscles of interest were considered: the Erector Spinae Longissimus (ESL), Erector Spinae Ileocostalis (ESI) and the Semitendinosus (ST). On the other hand, for the shoulder, the Trapezius descendens (TD), the Biceps Brachii (BB), the Deltoid medial (DM) and again the Erector Spinae Longissimus (ESL) were observed. To quantify muscular contraction, electromyographic signals from the muscles were recorded using surface electrodes (sEMG) from the BTS FREEEMG 1000 system. The measured data were then filtered and processed, and finally the Root Mean Square (RMS) value of each signal is calculated. The second aspect assessed in the objective analysis was the subjects' heart rate (HR) in the two study scenarios, and these signals were acquired using the Empatica E4 wristband. In the final stage, an assessment was conducted to identify any potential alterations in movement or posture resulting from exoskeleton usage, with a specific focus on the analysis of flexion-extension angles for the hip and shoulders. The data in this case were acquired using a Motion Capture system (Optitrack). On the other hand, the subjective analysis consists of two separate questionnaires. The first assesses the subjects' perceived fatigue during task execution using the BORG CR-10 scale. The second questionnaire, which is filled in at the end of the two trials (with and without exoskeleton), consists of 16 different questions, to be rated on a 0–10 scale, and divided into four categories: Confidence (CO), Cognitive Load (CL), Functionality (FU) and Physical Effort (PE).

2.2 Testing Protocol and Experimental Set-Up

For the purpose of this analysis, 10 male participants were involved, without orthopaedic problems, being 29.6 years old in average. All participants provided informed consent and the study was approved by the ethics committee of the Universitá Politecnica delle Marche.

Figure 1 shows the steps of the testing protocol. Initially, each subject is given a period of time to become familiar with both exoskeletons and the test procedure are explained. Subsequently, the choice of the starting test protocol (back or shoulder) and whether to start with or without the exoskeleton (exo

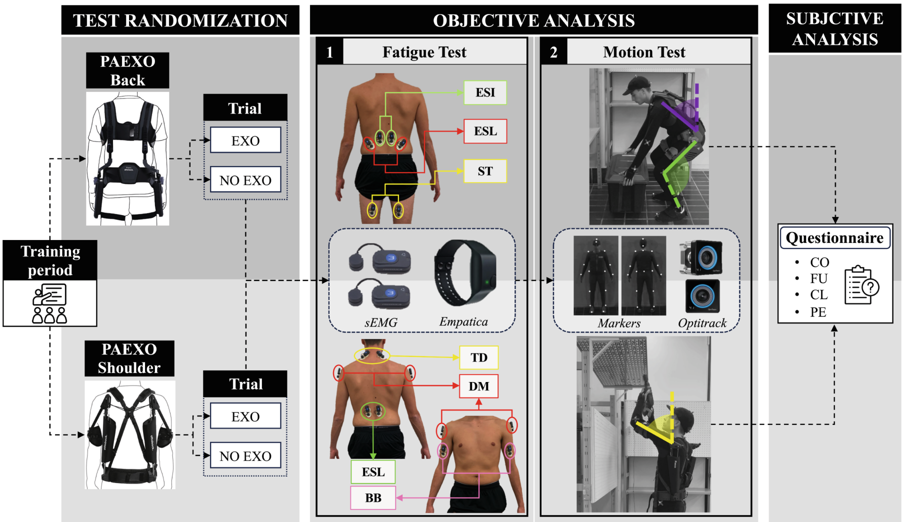


Fig. 1. Testing protocol steps: training period, test randomization, objective analysis (1. Fatigue Test; 2. Motion Test), subjective analysis. At the center, the used devices

or noexo) is randomized. For the objective analysis, the first step is the fatigue test, which involves the acquisition of EMG and HR signals and the perceived effort scores. For the back, the test consists of the cyclic repetition of a task that includes lifting and stowing a box. The initial weight of the box is 6 kg, but it is gradually increased to 12 kg and then to 20 kg. The exercise is repeated for 5, 4 and 3 min for each weight respectively and between each stage there is a 2-minute break. As for the shoulder, the test consists of three different tasks which require the arms to be kept raised: simulation of screwing, drilling and cabling. Each task lasts 2 min and the sequence is repeated twice. The second step is the motion test, aimed at the acquisition of joint angles with the motion capture system. For the back test, the previous exercise is repeated for 100 s for each weight. With regards to the shoulder test, only the frontal and overhead drilling simulation is performed for a duration of 100 s. At the end of the test, once it has been performed with and without the involved exoskeleton, the subject is asked to complete the post experimental questionnaire.

3 Results and Discussion

The muscular activity related to PAEXO back exoskeleton is summarized in Fig. 2(a). The most relevant result regards the ST, since for this muscle the plot shows a reduced muscular contraction while wearing the exoskeleton. On the other hand, from a postural point of view, Fig. 2(b) shows a significant reduction in the maximal hip flexion angle while using the exoskeleton, which indicates the maintenance of a better posture. However, no influence was found on the heart rate. This result is aligned with the subjective analysis (Table 1), in which the exoskeleton does not appear to provide relevant improvements in perceived fatigue [3] and shows an overall score of 5.3 in the post experimental questionnaire, which suggest a discomfort with the exoskeleton. Conversely, the PAEXO shoulder exoskeleton leads to a reduction in the contraction [8] noticeable for all the muscles involved in the analysis (Fig. 2(c)) and in the HR. Moreover, it can be observed that the shoulder flexion/extension angles (Fig. 2(d)) are comparable in the two tested configurations. These results demonstrate that in terms of fatigue and ergonomics, this exoskeleton proves useful in reducing physical effort while remaining comfortable to wear. This is also confirmed by the subjective analysis, which highlights the perception of an overall advantage in performing the task while wearing the exoskeleton (Table 1).

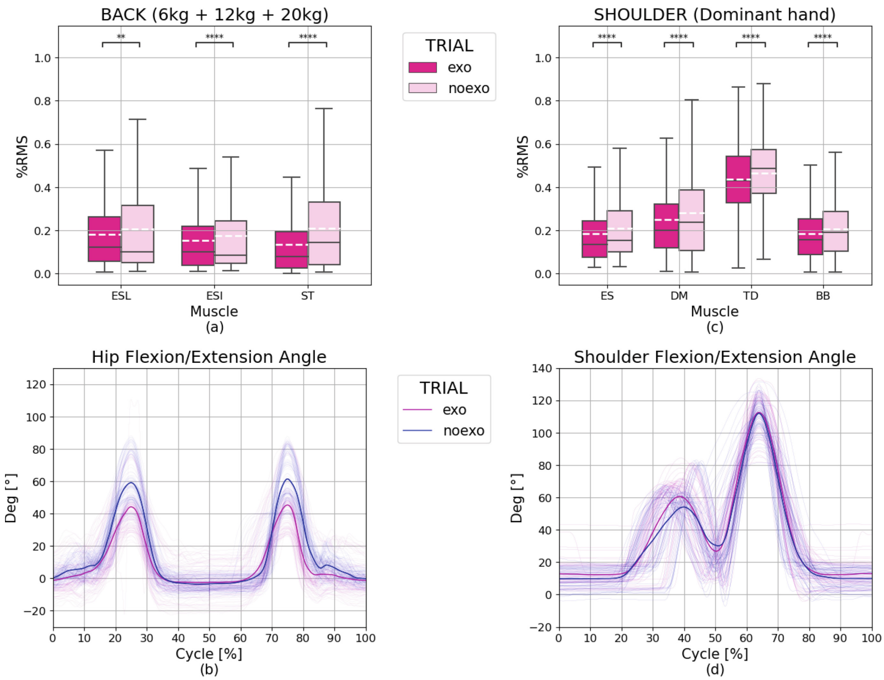


Fig. 2. Box plots of muscle activity (a, c), where the dotted lines represent the mean and the solid lines represent the median, stars indicate statistically significant differences resulting from Mann-Whitney-Wilcoxon two-sided test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$); hip and shoulder joint angles within a cycle (b, d), respectively lifting and stowing the box, and frontal and overhead drilling.

Table 1. Perceived fatigue (exo, noexo) and post-experimental questionnaires in four categories: Confidence (CO), Cognitive Load (CL), Functionality (FU) and Physical Effort (PE). Each result is expressed in Mean (sd).

	Perceived fatigue		Post experimental				
	exo	*noexo*	*CO*	*FU*	*CL*	*PE*	*TOT*
BACK	5 (2.0)	5.6 (2.1)	5.0 (1.7)	5.5 (1.3)	5.1 (1.2)	5.7 (1.3)	5.3 (1.0)
SHOULDER	4.5 (1.6)	6.25 (2.3)	6.9 (1.3)	6.8 (1.2)	6.7 (1.0)	6.7 (0.9)	6.8 (0.9)

4 Conclusion and Future Works

The presented work is a preliminary test to assess and refine the employed protocol, methods, and tools, in order to verify its transferability in industrial frameworks. From the initial results, the passive exoskeletons proved to be an effective measure for posture control and muscle overload, potentially reducing the risk of musculoskeletal injuries. The PAEXO back evaluation has highlighted some limits in the implemented protocol. From the users' feedback and from the objective analysis, these limits can be tackled by a longer familiarization period and by a simpler protocol that reproduces a real-case scenario. Conversely, the results concerning the PAEXO shoulder evaluation demonstrate that the use of the exoskeleton can help reduce muscle activity, consequently decreasing cardiac cost. Moreover, the subjects involved in the tests expressed positive effects on the reduction of perceived fatigue. This can be due to the greater intuitiveness of such an exoskeleton, which, compared to the PAEXO back, has been proven effective despite the short familiarization period. Future works will assess how comfortable and practical exoskeletons are when used during everyday non-task-specific movements (e.g. walking) and for performing tasks with collaborative robots.

Acknowledgements. The research has been funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 - VITALITY - CUP I33C22001330007.

References

1. Romero, D., Stahre, J.: Towards the resilient operator 5.0: the future of work in smart resilient manufacturing systems. Procedia CIRP **104**, 1089–1094 (2021)
2. De Bock, S., et al.: Passive shoulder exoskeleton support partially mitigates fatigue-induced effects in overhead work. Appl. Ergon. **106**, 103903 (2023)
3. van Sluijs, R.M., Wehrli, M., Brunner, A., Lambercy, O.: Evaluation of the physiological benefits of a passive back-support exoskeleton during lifting and working in forward leaning postures. J. Biomech. **149**, 111489 (2023)
4. Kok, J., et al.: Work-related MSDs: prevalence, costs and demographics in the EU. Publications Office of the European Union (2019)

5. Breque, M., De Nul, L., Petridis, A.: Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg (2021)
6. Lee, J., Kim, D., Ryoo, H.Y., Shin, B.S.: Sustainable wearables: wearable technology for enhancing the quality of human life. Sustainability **8**(5), 466 (2016)
7. Li-Baboud, Y.S., et al.: Evaluation methods and measurement challenges for industrial exoskeletons. Sensors **23**(12), 5604 (2023)
8. Brunner, A., et al.: Effect of a passive shoulder support exoskeleton on fatigue during working with arms over shoulder level. In: International Workshop on Medical and Service Robots, pp. 188–197. Springer, Cham (2023)

Towards Mastering Real-World Robot Benchmarking: Lessons Learned from the Robothon Grand Challenge

Peter So(✉), Ahmed Abdelrahman, Hoan Quang Le, Abdalla Swikir, and Sami Haddadin

Technical University of Munich, 80992 Munich, Bavaria, Germany
peter.so@tum.de
https://mirmi.tum.de

Abstract. In this paper, we delve into the pioneering insights gathered from the Robothon Grand Challenge (RGC) events of 2021–2023, focusing on the benchmarking of robotic manipulation skills. Central to our study is the innovative use of an internet-connected electronic task board (TB), a tool that allowed for the remote and comprehensive assessment of robot manipulation capabilities across varied research laboratories globally. We critically evaluate the pros and cons of this method, particularly in fostering reproducibility, enhancing transparency, and building trust in the results. Our findings not only contribute significantly to the field of robotics benchmarking but also pave the way for future advancements in the assessment and development of robotic systems under diverse conditions. This research underscores the evolving landscape of robotics, highlighting the role of smart technologies in shaping the future of robotics benchmarking and development.

Keywords: performance benchmarking · robotics · dexterous manipulation · internet-of-things · open-innovation

1 The Robothon Grand Challenge Competition Series

The Robothon Grand Challenge (*RGC*) is an annual competition event[1] which provides a benchmarking arena [1] for academia and industry to showcase progress in robot manipulation skill development. Following the example of existing robot system benchmarking practices [2,3], the competition centers around a real-world task artifact: an internet-connected, electronic task board (*TB*) invented specifically for our competition, which serves as a portable manipulation benchmark. The *TB* includes six increasingly difficult manipulation tasks. A program, known as the digital robot judge (*DR.J*) [4], running on a small

[1] Robothon Grand Challenge Competition at automatica event website https://www.robothon-grand-challenge.com.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 87–91, 2024.
https://doi.org/10.1007/978-3-031-76428-8_17

surface-mounted microcontroller, equipped with buttons, a screen, and a wireless radio on the *TB*, allows users to conduct their own trials and supervises their performance remotely. This feature enabled a novel decentralized competition with hardware-in-the-loop, see Fig. 1.

Qualified robot teams who applied to the competition received a *TB* by mail. The competition started after all teams have received their *TB* and the trial protocol is released during a video conference kickoff meeting. Then, teams have four weeks to develop an autonomous robot solution to the *TB*. At the end, teams submit a documentation package comprised of a code repository and a recorded video of their robot solution. Finishing teams present their solution to an expert jury over video conference to verify their results. The winners of the competition were decided by the jury using data collected with the *DR.J*.

From 2021 to 2023, the *RGC* competition series has received 85 team applications, with 48 competing teams from 24 countries across 190 developers and awarded 52,000EUR in cash prizes.

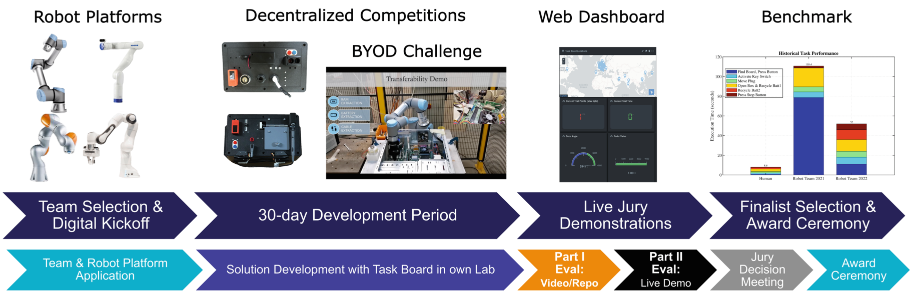


Fig. 1. Robothon Grand Challenge Competition Benchmarking Concept

2 Benchmarking with the Electronic Task Board

The *TB* serves as the real-world artifact for providing remote teams with a common test environment. This has the benefit that robot platforms and humans can be evaluated with identical hardware. A **Trial Protocol** instructs the user on the initial setup of the *TB* and indicates a series of tasks to be completed. **Task Performance Circuits** detect when a task is completed through sensors on the *TB*. A program running on the microcontroller mounted on the *TB* monitors trial attempts and regularly sends **Task Board Telemetry** to a web server over an internet connection with the *TB* state and trial execution times. A web application collects and renders usage statistics on a **Public Web Dashboard**.

Two design iterations of the electronic *TB* are in circulation. The first *TB* released in 2021 modeled disassembly and sorting of electronic waste and included a localization task, a peg-in-hole task, a key insertion task, and a battery box disassembly task. The same *TB* was used again in the 2022 competition to measure increases in performance. A new design iteration was introduced in 2023, which modeled an electrical inspection with a multimeter probe and incorporated a door-opening task and a cable-wrapping task.

2.1 Reproducibility and Best Practices

We claim all participants have the same test conditions by shipping an identical *TB* containing all objects required to execute the trial protocol to each team in the mail. The *TB*s were batch produced from an off-the-shelf microcontroller, sensors, and components mounted on a customized project box [4]. During the competitions, we followed these best known practices: (1) During the development period, keep the *TB* plugged in and connected to the internet to ensure data continuity; (2) All trial attempts were monitored by the *DR.J* and limited to a maximum of 10 min and scoring information displayed on the *TB* screen and on the public web dashboard; (3) Teams mounted the *TB* to a flat surface in their robot's workspace with Velcro and were required to physically pick up and replace the *TB* between trial attempts to showcase the localization ability of the robot; (4) Jury feedback from final demos was collected using a survey form with a numbered rating 1–5 scale to quickly generate a ranking report.

2.2 Technical Challenges in Robot Benchmarking

Managing the *TB*s after shipping them to the teams proved to be challenging. As a physical device, some *TB*s had delivery issues with customs controls. Robots broke parts on the *TB* during the development period. The design is made open source so teams could source or 3D print their own replacement parts. The compact size of the *TB* limited the type of tasks which could be evaluated. Relying on the *DR.J* to determine task completion required each task to be detectable with an electrical circuit which further limited task definitions. The *DR.J* could evaluate the metrics listed in Table 1 on the electronic *TB*. Challenges encountered in practice included intermittent telemetry and uncertainty in telemetry between robot and human trial attempts from telemetry data alone. These challenges were overcome by scheduling individual live demonstrations of finishing team with the organizers and jury over a video conference call to validate their results. Teams prepared two camera views for the jury demo and the web dashboard provided a common reference for task completion, see Fig. 2.

2.3 Assessing Transferability and Reusability of Robot Abilities

In conjunction with autonomously solving the trial protocol with the *TB*, teams were instructed to demonstrate their robot's manipulation skills on a similar but different object of their choice in the bring-your-own-device (BYOD) challenge. To encourage teams to develop modular robot skills, the execution order of the trial protocol was randomized during the live presentation with the jury. Skills are considered a robot's ability to recognize, plug-in, or open an object. In 2021

Table 1. Proposed performance metrics that can be measured remotely by the *TB* performance circuits and the web dashboard.

Metric	Definition	Aspect Assessed	Units
Execution Time	Trial completion time	Speed	Seconds
Success Rate	Successful/total attempts	Robustness	Percentage

and 2022 where the *TB* featured tasks to remove AA batteries from a battery compartment, during a 30-min live jury presentation various teams showcased their robot removing batteries from toy pianos to wireless key remotes. In 2023 the *TB* emulated a physical electrical probing task with a multimeter, teams showcased their robots testing electrical circuits on power strips, computer motherboards and inspecting battery voltage for cylinder and coin shaped batteries. In the BYOD challenge, the jury used the *TB* as a common ability reference to compare and contrast the transferability of a robot skill to a new domain. Jury members assessed the robot solution's skill transferability based on the BYOD performance, awarding a percentage of a fixed number of points to the team's overall score. Evaluation of robot skill transferability is still outside of the scope of the *DR.J* but this is an area of future work.

3 Results

For each of the *RGC* events, the final team performance ranking was determined by a vote of the expert jury during a decision meeting. We found the *DR.J* correctly identified the top performing team based on the metrics of solution robustness and execution speed. However, the jury was still necessary to assess the quality of the robot skill transferability in the BYOD challenge. A summary of the *RGC* competition results and team contributions are available on *GitHub*[2].

4 Future Proposed Performance Metrics

In future work, we plan on augmenting our benchmark with the following additional performance metrics: (1) **Computational complexity:** the amount of necessary computational resources, e.g. the measured CPU and GPU memory and clock cycles required; (2) **Utilized sensing modalities:** a measure of perception requirements, such as the number of visual, force, position, etc. sensors that are utilized; (3) **Energy consumption:** the amount of energy consumed by the system to measure the efficiency of the solution.

In addition, proprioceptive data from the robot, such as joint positions, velocities, and accelerations, could be obtained from ROS bags that are recorded during task executions. These data can provide insights into motion trajectory properties, such as length and jerk, which are useful in evaluating the efficiency and other metrics concerning the robots' motions.

[2] https://www.github.com/peterso/robothon-grand-challenge/.

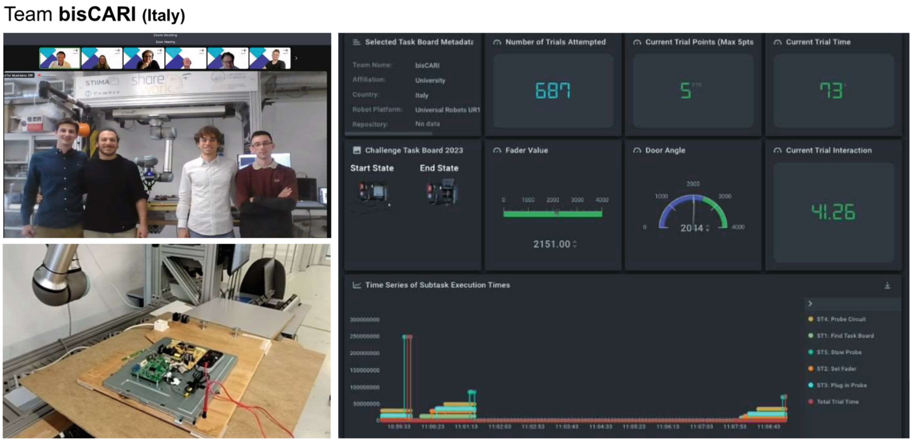


Fig. 2. Jury member perspective during team live presentation of benchmark results. Left: Video conference with two camera views (room view and close up workspace view). Right: Web dashboard of *TB* telemetry showing trial execution time.

5 Conclusion

The *RGC* annually benchmarked robot manipulation skills and demonstrated a measurable increase in performance with the electronic *TB*. The *DR.J* not only provided insight into the robot skill development but also provided transparency into benchmarking physical robot experiments remotely. The lesson learned was task performance circuits enabled both humans and robots to be evaluated and their performance gap directly measured. Leveraging sensorized physical *TB*s in future robot benchmarks and automatically publishing trial results can provide added visibility of progress in the benchmark across researchers around the globe.

References

1. Amigoni, F., et al.: Competitions for benchmarking: task and functionality scoring complete performance assessment. IEEE Robot. Autom. Mag. **22**(3), 53–61 (2015). https://doi.org/10.1109/MRA.2015.2448871
2. Bonsignorio, F., del Pobil, A.P.: Toward replicable and measurable robotics research [from the guest editors]. IEEE Robot. Autom. Mag. **22**, 32–35 (2015). https://doi.org/10.1109/MRA.2015.2452073
3. Falco, J., et al.: Benchmarking protocols for evaluating small parts robotic assembly systems. IEEE Robot. Autom. Lett. 1 (2020). https://doi.org/10.1109/LRA.2020.2965869
4. So, P., Sarabakha, A., Wu, F., Culha, U., Abu-Dakka, F.J., Haddadin, S.: Digital robot judge: building a task-centric performance database of real-world manipulation with electronic task boards. IEEE Robot. Autom. Mag. (2023). https://doi.org/10.1109/MRA.2023.3336473

Seamless Human-Robot Interaction Through a Distributed Zero-Trust Architecture and Advanced User Interfaces

Alessandro Peretti, Matteo Mazzola, Luca Capra, Marco Piazzola, and Cristiano Carlevaro(✉)

Spindox Labs S.R.L, 38123 Trento, TN, Italy
{alessandro.peretti,matteo.mazzola,luca.capra,
marco.piazzola,cristiano.carlevaro}@spindox.it

Abstract. The proposed work presents a novel interaction platform designed to address the shortage of skilled workers in the labor market, facilitating the seamless integration of robotics and advanced user interfaces such as eXtended Reality (XR) to optimize Human-Robot Collaboration (HRC) as well as Robot-Robot Collaboration (RRC) in an Industry 4.0 scenario. One of the most challenging situations is to optimize and simplify the collaborations of humans and robots to decrease or avoid system slowdowns, blocks, or dangerous situations for both users and robots. The advent of the LLMs (Large Language Model) have been breakthrough the whole IT environment because they perform well in different scenario from human text generation to autonomous systems management. Due to their malleability, LLMs have a primary role for Human-Robot collaboration processes. For this reason, the platform comprises three key technical components: a distributed zero-trust architecture, a virtual avatar, and digital twins of robots powered by the Robot Operating System 2 (ROS2) platform.

Keywords: XR · ZTA · LLM · Digital Twins · ROS2 · HRC · RRC · Industry 4.0

1 State of the Art

Historically, robots and their automation software were used to perform repetitive and rudimentary tasks, which were defined by a set of finite instructions on specific and well perimeter jobs. Nowadays, automation systems have been empowered by the advent of new technologies, especially AI-based that has led robots to resolve tasks more complex than ever. On the other hand, these methods are generating new challenges, not only in developing but also in deploying and usage of these new systems. For this reason, Human-Robot collaboration (HRC) as well as Robot-Robot Collaboration (RRC) processes have become a crucial point. Within industry 4.0 era, most of the laboratories have become

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 92–95, 2024.
https://doi.org/10.1007/978-3-031-76428-8_18

smart and they use robots or other autonomous systems to automatize different tasks to enhance internal productivity, workplace safety and product/service quality [1]. Recent studies have highlighted huge benefits in using well optimized processes for HRC and RRC [2]. HRC and RRC are delicate processes and a model to authenticate and authorize an action or task has a crucial role. Traditional security model approaches consider everything coming from internal system safe. To overcome this limitation Zero Trust Architecture or also called ZTA is used. It authenticates and authorizes every interaction between entities. The basic concept of this approach is "never trust, always verify" which means that entity inside the system should not be trusted by default. At every interaction request the system checks whether the action could be requested by the entity requester as well as if it can be performed by the entity receiver [4]. Over the past few years, this model has proven to be a remarkably effective security solution [3]. The driver of this recent robotics innovation can be detected in the use of LLM, defined as Large Language Model, which is a type of artificial intelligence model designed to understand and generate human language text at a large scale. LLMs are a so-called language model, which refers to a particular set of probabilistic models that can generate words exploiting the statistics learned from big training sets, mainly based on the transformers architecture, which relies on the attention mechanism [5]. LLMs are notable for their ability to achieve general-purpose language understanding and generation; their use can be applied in multiple IT domains, especially in the ones that are focused on task automatization. An autonomous agent system fueled by an LLM is able to unlock human-like peculiarities, such as memory used for learning new trajectories, recall previous stated knowledge, decision-making reasoning and use of external tools to interact with the ambient. Using these talents, the enhanced system can break down complex tasks that previous Artificial Intelligence (AI), Deep Learning (DL), Machine Learning (ML) approaches had difficulties with or were prone to error to complete, such as Human to Machine interaction, text generation or automatic software usage. The use of LLM can go beyond the generation of human language text. For example, an LLM can be abstracted in many ways to accomplish more complex tasks like managing autonomous systems. According to Insignia Ventures Partners, the market for autonomous AI agents is estimated to grow by 43 */percent* within 2028 [6]. LLM are mostly provided by big companies that train them using enormous datasets, resulting the model to be able to generalize well in multiple tasks, referring to data it's trained on. For specific domains, an LLM can be used after a process of finetuning, which is a small training phase that updates certain parts or layers of the model to be able to generalize well on new use cases. Finetuning process however can be complex, expensive and hard to evaluate, so another option is to apply Retrieval Augmented Generation (RAG), which refers to the ability to apply external knowledge at each inference, forcing the model to answer using a very specific resource [8].

2 Proposed Work

There are non-standard processes for HRC nor RRC and most of the time they can be used only by highly trained and qualified human workers. The proposed work is focused on the simplification of the interaction between humans and robots through the creation of a highly innovative IT interface. To achieve this, a virtual avatar has been developed. It serves as a bridge between human users and the company's knowledge base. The avatar is equipped with the capability to extract meaningful operations from the knowledge base and translate them into actionable instructions for the robots. Users can interact with the avatar using natural language and or through gestures captured by an RGBD camera, making the HRC accessible to non-skilled human workers. The avatar uses an LLM to extract the operations from an organized, ad hoc generated knowledge base, and translate them to a well-defined set of tasks that robots can perform in a specific environment. To simplify the interaction even more, the avatar looks like a person so that the workers can feel at ease as if they are speaking to one of their colleagues. Figure 1 shows the system architecture. A Graphical User Interface (GUI) is provided to have an overview of the system in real time. It represents the digital twins of the system robots. The proposed GUI is XR compliant, and it allows the users to have a 3D visualization over the real user environment of the whole robot's system. Edge Digital Twin [9,10] has been implemented through the Robot Operating System 2 (ROS2 [7]) platform, enhancing the platform's capabilities. The digital twin enables actuation and real-time sharing of sensory information, allowing robots to execute tasks with precision and adapt to dynamic environments. Furthermore, the digital twin component enables the orchestration of routines and tasks to support robots' cooperation based on the user directives. This innovative interaction platform significantly lowers the barriers for non-skilled workers to interface with complex robotic systems, providing a natural and accessible means of control. The integration of XR and natural language interfaces further enhances the user experience, making it easier for individuals with varying levels of technical expertise to interact with robots. As a result, this platform is poised to address the skilled worker shortage by making advanced automation technology more accessible and user-friendly in various industries. By combining a distributed zero-trust architecture, a virtual avatar, and digital twins of robots, this platform represents a groundbreaking advancement in human-robot interaction, with the potential to revolutionize the way we employ automation and robotics in the workforce.

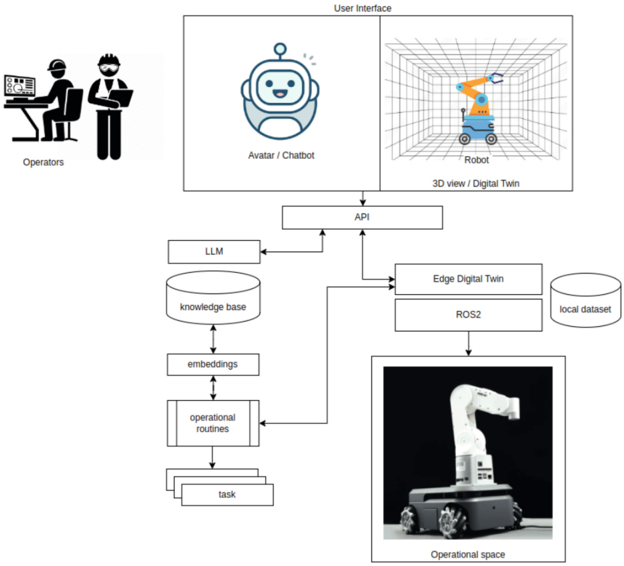


Fig. 1. System Architecture.

Acknowledgments. This work is based upon and motivated by the ongoing work on the Sermas project (https://sermasproject.eu/), Grant agreement ID: 101070351.

References

1. Nicora, M.L., Ambrosetti, R., Wiens, G.J., Fassi, I.: Human-robot collaboration in smart manufacturing: robot reactive behavior intelligence. J. Manuf. Sci. Eng. **143**(3), 031009 (2021)
2. Othman, F.U., Yang, E.: Human-robot collaboration in smart manufacturing environments: review and outlook. Sensors **23**(12), 5663 (2023)
3. Biplop, P., Muzzafar, R.: Zero-trust model for smart manufacturing industry. Appl. Sci. **13**(1), 221 (2023)
4. The Rise of Zero Trust: book, Jupiter Network - Engineering Simplicity
5. OpenAI: Better Language Models and their implications. https://openai.com/research/better-language-models
6. Zhang, J.: How LLM Agents are Unlocking new Possibilities. https://www.wiz.ai/how-llm-agents-are-unlocking-new-possibilities/
7. ROS2 Documentation. https://docs.ros.org/en/foxy/index.html
8. Chen, H.T., Xu, F., Arora, S.A., Choi, E.: Understanding retrieval augmentation for long-form question answering, computation and language. arXiv:2310.12150v1 [cs.CL] (2023)
9. Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., Calinescu, A.: Digital twins: state of the art theory and practice, challenges, and open research questions. J. Ind. Inf. Integr. **30**, 100383 (2022)
10. Picone, M., Mamei, M., Zambonelli, F.: A flexible and modular architecture for edge digital twin: implementation and evaluation. ACM Trans. Internet Things **4**(1), 1 (2023)

Follow Me: An Architecture for User Identification and Social Navigation with a Mobile Robot

Andrea Ruo(✉), Lorenzo Sabattini, and Valeria Villani

University of Modena and Reggio Emilia, 42122 Reggio Emilia, RE, Italy
{andrea.ruo,lorenzo.sabattini,valeria.villani}@unimore.it

Abstract. Over the past decade, a multitude of service robots have been developed to fulfill a wide range of practical purposes. Notably, roles such as reception and robotic guidance have garnered extensive popularity. In these positions, robots are progressively assuming the responsibilities traditionally held by human staff in assisting customers. Ensuring the safe and socially acceptable operation of robots in such environments poses a fundamental challenge within the context of Socially Responsible Navigation (SRN). This article presents an architecture for user identification and social navigation with a mobile robot that employs computer vision, machine learning, and artificial intelligence algorithms to identify and guide users in a social navigation context, thereby providing an intuitive and user-friendly experience with the robot.

Keywords: Robot guidance · Service robot · Socially-responsible navigation

1 Introduction

Over the past few years, the emerging applications of robotics require robots to perform tasks in social spaces, i.e., environments shared with people, making it crucial to enable robots to operate in a socially acceptable manner. Compared to robot navigation in non-social environments, such as underwater or warehouse environments, SRN considers both non-social obstacles and social agents, i.e., people, taking into account their comfort, naturalness, and social interactions. In a social navigation context, a mobile robot autonomously operates within an environment, guiding a user to a specific destination. This task involves the challenge of successfully navigating a path while avoiding collisions with obstacles in the environment [1]. Some of these social spaces, such as museums and shopping centers, can be large and crowded, making it crucial for robots to move around displaying appropriate social behaviors [2]. Human-robot interaction plays a crucial role in this research scenario. It is crucial for the robot to monitor the user's position in the environment to ensure reliable and safe guidance. The European Union-funded project SPENCER [3] developed a reception

This work was supported by Horizon Europe program under the Grant Agreement 101070351 (SERMAS).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 96–100, 2024.
https://doi.org/10.1007/978-3-031-76428-8_19

robot designed to assist, inform, and guide passengers in large and crowded airports. This robot combines map representation, laser-based people and group tracking, and activity and motion planning. Stricker et al. [4] proposed a robot-based information system for a university building. The reception robot provides information about offices, employees, and laboratories in the building and can guide visitors to their desired locations. Gross et al. [5] developed TOOMAS, an interactive reception robot for shopping. This robot can autonomously approach potential customers, navigate through the market, and guide clients to their chosen products, providing an accompanied shopping experience throughout. Research on robot guidance for visually impaired individuals [6] has led to the implementation of a guidance system on a robot specifically designed to assist this user category.

This work originates from the European project SERMAS, which aims to develop innovative, formal, and systematic methodologies and technologies for modeling, developing, analyzing, testing, and studying the use of socially acceptable advanced technology systems. In this paper, we propose the development of an architecture using ROS2 that, through the application of computer vision, machine learning, and artificial intelligence algorithms, is capable of identifying and guiding a person within a social navigation context. The main objective is the implementation of a system that allows the robot to guide a user while instantly verifying that the human is following, by checking that the distance between the robot and the human is less than the desired distance. To achieve this goal, the system is designed to undergo several operational phases. Firstly, the robot detects the presence of a human presence; subsequently, the person will perform an intention communication action, using gesture recognition techniques, to communicate their desire to be guided by the robot. Thirdly, the robot will identify the person using facial recognition methods and then move along a predefined path, monitoring the distance from the human and stopping if it exceeds the desired distance. This architecture has been implemented in an experimental validation test.

2 Hardware and Software Implemented

In order to implement the architecture of the proposed system, we utilized the following hardware and software components. The mobile robot used for this purpose is the MiR100, shown in Fig. 1. The robot's broad base allows us to mount an Intel® RealSense™ D435i on a tripod, which is used to capture video stream at the rear of the robot and measure the distance from the user. Within this system, we employed: i) OpenCV, which provides a range of pre-trained models and algorithms that can be used for common computer vision tasks, such as object detection, image recognition, and facial detection; ii) MediaPipe[1], used for hand and skeleton recognition, along with a dedicated model for hand gesture recognition; iii) TensorFlow, utilized for face recognition, leveraging a pre-trained

[1] Real-time Hand Gesture Recognition using TensorFlow & OpenCV: https://techvidvan.com/tutorials/hand-gesture-recognition-tensorflow-opencv/.

Fig. 1. Mobile robot MiR100 during experimental validation.

artificial intelligence model known as SENet, belonging to the VGGFace model family[2]; ROS2 was used to control the various hardware components and libraries mentioned in this architecture.

3 Architecture

The proposed architecture, as shown in Fig. 2, consists of three nodes: the *manager_node* is responsible for ensuring the correct execution order of the architecture. The first task to be executed is gesture recognition. To achieve this, the *manager_node* calls the *gesture_rec* service using the *TryGesture* interface. In case of a "False" response, the node will stop the process by calling the *home_base* service. In the event of a "True" response, the node calls the service *realsense_sub*, enabling facial recognition. The *realsense_sub* node is responsible for skeleton recognition of the user to ensure that the user is indeed following the robot. The skeleton recognition algorithm can be started by associating it with the identified person to be followed, who is identified using the *face_id* service. The use of facial recognition control is chosen to prevent the robot from losing the previously identified person, especially in urban environments, making the system more robust. To perform these operations in parallel, the *realsense2_camera_node* needs to be running, which provides */color/image_rect_raw* containing RGB video frames and */depth/image_raw* containing depth frames associated with them. The purpose of the final node, *cmd_node*, is to command the robot's velocity while monitoring the distance between the robot and the user at each instant. This process is repeated in a loop until the person reaches their destination or moves out from the robot's field of view. In such a case, the main *manager_node* calls the *home_base* service, allowing the robot to return to its starting position.

4 Experimental Validation

The aim of the experimental validation is to allow guiding a user to a specific destination considering a predefined path. Throughout the entire execution of

[2] Visual Geometry Group at the University of Oxford: https://www.robots.ox.ac.uk/~vgg/software/vgg_face/.

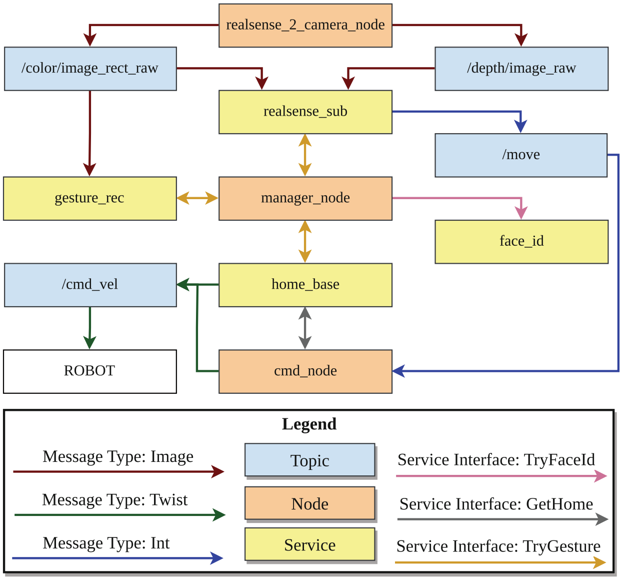


Fig. 2. Proposed architecture.

the experiments, the distance between the robot and the user is tracked and recorded. The test participant approached the robot and performed a gesture to communicate their presence to the robot and be recognized. After the successful identification of gestures, facial recognition was activated to keep the identified user's skeleton active for the purpose of calculating the distance from the user using RealSense. The robot began moving along a straight path while continuously tracking and recording the real-time distance between the robot and the user. The system was developed to ensure that the robot stops when the detected distance from the user exceeded a certain desired distance, arbitrarily set at 2 m.

At the beginning of the test, the user followed the robot until a certain point in time where the detected distance exceeded the allowed desired distance, leading to the robot coming to stop, as shown in Fig. 3. During the acquisitions interval [(324:602), (1048:1234), (1535:1712), (1907:1962), (2612:2682)] when the user was at a distance greater than the desired distance, the robot remained stationary. Subsequently, the user approached the robot again, allowing the robot to resume its movement. During the different tests we carried out, noise was observed in the data acquisition process by the RealSense, especially at the end of the experiments when the user moved out of the robot's field of view to conclude the test. In order to improve the data acquired by the RealSense camera, Exponential Moving Average (EMA) was implemented, as shown in Fig. 3. This is a first-order infinite impulse response filter that applies weighting factors that decrease exponentially. The weights for previous data for each older datum decreases exponentially, never reaching zero.

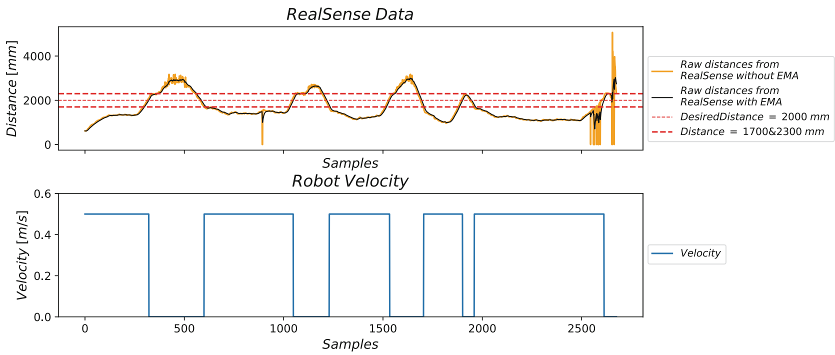


Fig. 3. Plot representing the measured distances from the RealSense with and without the use of EMA in parallel with the robot speed during experimental validation.

5 Conclusion

In this work we presented an architecture for identifying and guiding a person within a social navigation context using gesture recognition and facial recognition, ensuring that the robot can move along the path only when the user is within a desired distance. As a further development, we are integrating algorithms that allow the robot to move autonomously by performing collision avoidance and environment mapping.

References

1. Song, K., et al.: Navigation control design of a mobile robot by integrating obstacle avoidance and LiDAR SLAM. In: International Conference on Systems, Man, and Cybernetics. IEEE (2018)
2. Silva, S., Verdezoto, N., Paillacho, D., Millan-Norman, S., Hernández, J.D.: Online social robot navigation in indoor, large and crowded environments. In: International Conference on Robotics and Automation. IEEE (2023)
3. Triebel, R., et al.: Spencer: a socially aware service robot for passenger guidance and help in busy airports. In: Field and Service Robotics: Results of the 10th International Conference. Springer (2016)
4. Stricker, R., et al.:. Interactive mobile robots guiding visitors in a university building. In: IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. IEEE (2012)
5. Gross, H.-M., et al.: TOOMAS: interactive shopping guide robots in everyday use-final implementation and experiences from long-term field trials. In: IEEE International Conference on Intelligent Robots and Systems. IEEE (2009)
6. Hwang, H., et al.: System configuration and navigation of a guide dog robot: toward animal guide dog-level guiding work. In: International Conference on Robotics and Automation. IEEE (2023)

Detecting ErrPs Signals in HRI Tasks

Alessandra Fava[1(✉)], Adriana Lucchese[1], Roberto Meattini[2], Gianluca Palli[2], Valeria Villani[1], and Lorenzo Sabattini[1]

[1] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy
{alessandra.fava,adriana.lucchese, valeria.villani,lorenzo.sabattini}@unimore.it

[2] Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, 40126 Bologna, Italy
{roberto.meattini,gianluca.palli}@unibo.it

Abstract. In the last two decades, electroencephalography (EEG) signals have been used as a relevant source of information in human-robot interaction (HRI). In particular, in the last years Error Related Potentials (ErrPs) have been introduced. These potentials can be leveraged during interaction tasks to mark the mismatch between a robot's behavior and human expectations. These signals are used to better adapt the robot to human needs, through a control based on these signals. This work aims to investigate ErrPs to study their potential through an experiment, in order to use them as feedback for adapting and correcting a robot system. We present a setup and experimental protocol: the experiment is divided into five tasks with seven subjects. For every task, we have 120 events, with a 25%–35% probability of error. We used Matlab2023a and the toolbox EEGLAB2023.0 for EEG analysis. We performed this experiment with a Baxter robot and the interaction with the robot was done in two different ways, with a keyboard or in a teleoperation scheme. The tasks are designed to reproduce, for example, a problem teleoperated pick and place in the industry.

Keywords: HRI · ErrPs · EEG · Robot

1 Introduction

In recent years, Human-Robot Interaction (HRI) [1] has shown a particular interest in the use of EEG signals. In particular, this work focuses on the Error Related Potentials (ErrPs). The ErrPs are specific Event-Related Potentials (ERP), and these signals are particular EEG signals. The ErrPs are evoked involuntarily when the subject perceives an error from the robot or another human during an interaction task [1], while the ERPs are evoked after a specific

This work is supported by the "Lively Ageing" project, funded by the Italian Ministry of Health.

A. Fava and A. Lucchese—The two first authors contributed equally.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 101–106, 2024.
https://doi.org/10.1007/978-3-031-76428-8_20

event. In the literature, there are studies on the use of these signals as feedback after the movement of the robot to customize and improve the robot system: this is achieved using these signals for detecting errors during the interaction [2]. The ErrPs are used to control humanoid robots or drones [3], where the command from the EEG signals are converted in command to the robot, or to control neuroprosthetic devices [4]. These signals are also used as reward functions in approaches based on inverse reinforcement learning (IRL) [5]. During an HRI task, if the action performed by the robot is as expected by the user, the feedback is positive and the robot receives a positive reward. Instead, if the robot's movement is different from the user's expectation, the ErrP is evoked and the feedback is negative, so the robot adjusts its movement. The ErrPs are characterized by a specific waveform with determined latency, the duration being from 600 to 800 ms. Specifically, they are characterized by a positive peak after 50–100 ms, called P100, where P indicates positive amplitude and 100 is the peak latency. After, there is a negative peak, between 200–500 ms, called N200, and, at the end, a positive peak between 300–600 ms, called P300.

This work aims to explore the nature of ErrPs and understand how these signals can be used in a Human-Robot Interaction problem through an experiment. The experiment is divided into five tasks: all the subjects performed the tasks in the same order, with increasing complexity. During each task, the participants gave a command to the robot and they were asked to verify the correctness of the robot's movement.

2 Experiment

2.1 Setup

Seven subjects participated in the experiment: among them, we had 5 males and 2 females. The mean and standard deviation of their ages are respectively 28.143 and 3.848 years. All the subjects volunteered to participate in the experiment. The study protocol followed the Helsinki Declaration and compliance to participate in the study was obtained from written informed consent before starting the experiment. All the data were reported anonymously. We used a robot for all the experiment tasks, namely the Baxter from Rethink Robotics. This is a collaborative robot with two arms and a display on the face. We interacted with the robot in two different modalities. For the first three tasks, with the keyboard of a PC. Instead, for the remaining tasks, we used a motion capture device, namely the Vicon, to create a teleoperation system based on gesture detection. The markers of this technology were placed on the right wrist of the subject. We captured the movement of the right wrist as a command to be sent to the robot. For every task, we have implemented 120 events with a 25%–35% of probability of error: this is because, in order to evoke the ErrPs, it is important that the robot does not work correctly. As a consequence in our experiments, the total number of events representing an error is 30 to 40 for every task. For the recording of EEG signals, we used a neural helmet "Enobio 20" with 19 electrodes: Fp1, Fp2, P7, P4, Cz, Pz, P3, P8, F3, Oz, T8, F8, C4, F4, Fz, C3, Fpz, T7, F7. The electrodes were located following the international system 10/20 [6]. We used 17 dry electrodes and 2 gel electrodes. The reference electrode was the Fpz and, for the ground, the left ear lobe was used. The sample rate used was 500 Hz, and data were recorded with Neuroelectrics software NIC2 [7] and a PC (Intel Core i5 CPU@2.50GHz, Windows 11 a 64-bit).

2.2 Tasks

The experiment was developed to recreate a semi-realistic scenario of HRI, such as a pick and place task with teleoperation. Every subject performed the tasks in the same order with increasing complexity. The subject gave the command to the robot and then verified the correct execution of the command by the robot through observation. A wrong action of the robot was represented by the execution of any other command, excluding the command given by the participant; this applies to all tasks. The tasks are described as follows:

First Task: The implemented action is the movement of the right arm of the Baxter, which can rotate to its left or right. The command is given by the user by pressing "r" (right) or "l" (left) on the keyboard.

Second Task: The robot can open or close the gripper. Also in this case, the command is given by pressing "c" (close) or "o" (open) on the keyboard.

Third Task: In this task, the subjects move their right wrist, where there are markers of the motion capture device. The acquired movements from this device are commands to move the right arm of the robot. The movements implemented are a rotation to the right or the left. This task is the same as the first task but performed in teleoperation with the motion capture device.

Fourth Task: The subject uses the keyboard to do a pick and place task. The motions allowed by the robot are: right ("r"), left ("l"), up ("u"), down ("d"), open ("o"), and close ("c").

Fifth Task: This is a variation of the fourth task. In this case, the subjects move their arm to move the robot in directions: right, left, up and down. Opening and closing the gripper are executed only with the keyboard.

2.3 Preprocessing and Analysis of Data

We recorded the EEG signals as described in Sect. 2.1. After recording, we data was preprocessed with Matlab2023a and EEGLAB2023.0 toolbox, following the pipeline shown in Fig. 1.

The experiment was implemented with two different computers, one for EEG signal recording and the other for robot control. Using two different PCs is a problem for the start time of acquisitions because we obtain a gap of a few seconds despite the synchronization. The synchronization between the time of acquisition from the PCs is very important because it will facilitate the work of locating events. First of all, we obtained the timeline of the acquisition data

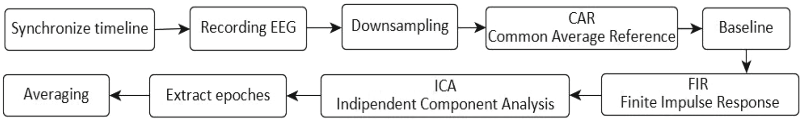


Fig. 1. Preprocessing architecture

coincident with that of the task. After that, data was downsampled from 500 Hz to a frequency of 256 [HZ] and we implemented a Common Average Reference (CAR) filter and baseline correction. Next, data was filtered with a FIR sinc filter with a Hamming window with a low cutoff frequency of 0.5 Hz and a high cutoff frequency of 10 Hz to remove low-frequency drift, power noise, and high-frequency muscle activity. Finally, an algorithm of EEGLAB was applied to reject bad channels by kurtosis calculation and we did an independent component analysis (ICA) for artifact elimination.

During the task recording, the movements of the robot were marked as "correct" or "error". The error movements are important because they evoke the ErrPs, and for this reason, we used a probability of error defined in Sect. 2.1. If the robot's movement corresponds to the command, the movement was marked as "correct". If the robot's movement did not correspond with the command, the movement was marked as "error", and in these signals, we expected to recognize the ErrPs. With these two marks of signals, we extracted the epochs from the recorded data. Finally, epochs from "correct" and "error" trials were averaged. The outcome of this stage is two different signals: one with the event "error" and the other with the "correct" commands for every channel.

To find the peaks we did a latency analysis with the following time window: [0–100] ms, [100–200] ms, [200–300] ms, [250–350] ms, [400–450] ms, [500–650] ms. For this first experiment, we conducted a simple qualitative analysis of the signals obtained by averaging. In particular, we visually inspected the temporal signal in order to identify the difference between the signal with "error" and the signal "correct" for every channel.

In Fig. 2 we can see, in red, the average of the epochs with error and, in black, the average of epochs without error, for one channel, selected as the most significant during results visualization, for one of the subjects. In the case of "error" signals, we can see a positive peak shortly after the zero, which corresponds to the instant when the robot starts its movement. Close to 100 ms, a negative peak can be identified, together with a positive peak between 200–300 ms. Such peaks are not present in the averaged "correct" epochs and, as a result, it can be assumed that they represent an ErrP.

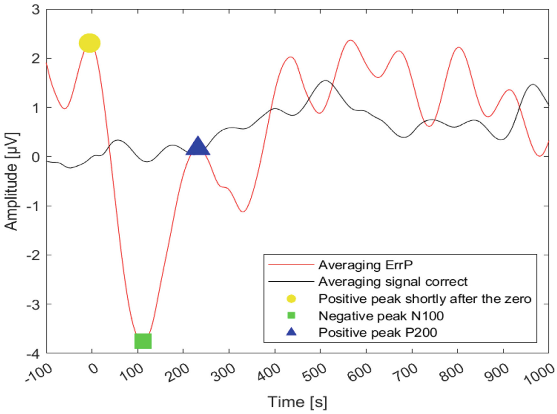


Fig. 2. Result for one channel of one subject.

3 Conclusion

This experiment aims to evoke the ErrPs in HRI. Our approach is preliminary since we carried out a qualitative analysis of EEG signals. During the analysis, we identified some limitations of our experiment. First of all, the number of trials with an error is too small to define well the waveform of ErrPs. To overcome this, it is needed to extend the experiments in order to increase the number of available trials. Other limitations are the use of two different computers to synchronize the timeline, and the use of Vicon for teleoperation, which is very sensitive to disturbance (e.g. experiment performed in a cluttered room).

Despite these limitations, it is still possible to observe a different behavior between signals with errors and signals without errors. In the signals with error, we can identify the ErrPs according to the literature, because we can identify the peaks that characterize the particular waveform of ErrPs.

References

1. Yasemin, M., Cruz, A., Nunes, U., Pires, G.: Single trial detection of error-related potentials in brain-machine interfaces: a survey and comparison of methods. J. Neural Eng. **20** (2022). https://doi.org/10.1088/1741-2552/acabe9
2. Wang, X., Chen, T., Wang, Y.-K., Lin, C.-T.: Implicit robot control using error-related potential-based brain-computer interface. IEEE Trans. Cogn. Dev. Syst. **15**(1), 198–209 (2023). https://doi.org/10.1109/TCDS.2022.3151860
3. Mao, X., et al.: Progress in EEG-based brain robot interaction systems. Comput. Intell. Neurosci. **2017**, 1–25 (2017). https://doi.org/10.1155/2017/1742862
4. Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J., Millan, J.D.R.: Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control. Sci. Rep. 13893 (2015). https://doi.org/10.1038/srep13893
5. Kim, S.-K., Kirchner, E., Stefes, A., Kirchner, F.: Intrinsic interactive reinforcement learning - using error-related potentials for real world human-robot interaction. Sci. Rep. **7** (2017). https://doi.org/10.1038/s41598-017-17682-7

6. Homan, R.W., Herman, J., Purdy, P.: Cerebral location of international 10–20 system electrode placement. Electroencephalogr. Clin. Neurophysiol. **66**(4) (1987). https://doi.org/10.1016/0013-4694(87)90206-9
7. NIC2. https://www.neuroelectrics.com/solution/software-integrations/nic2

Towards Mixed Reality Applications to Support Active and Lively Ageing

Marta Gabbi, Valeria Villani, and Lorenzo Sabattini(✉)

Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy
{marta.gabbi,valeria.villani,lorenzo.sabattini}@unimore.it

Abstract. The global population is ageing at a significant pace. The percentage of older adults is expected to increase to 24% by 2030, leading to a consequent growth in the number of people affected by dementia. This condition calls for special attention to the daily physical and psychological needs of the individuals involved. Simultaneously, aged care problems, such as elevated costs and the shortage of professional caregivers, are worsening the scenario. Robotic technologies are playing a pivotal role in assisting elderly individuals in retaining their autonomy and providing innovative, direct assistance to them. In the literature, there is evidence of the significant impact that these technologies have on elderly people, especially the ones with dementia. These positive results translate into decreased depression and anxiety and into improved overall wellbeing and quality of life. This paper introduces a mixed reality application created for older adults with dementia, with the aim of stimulating movement and cognitive functioning. The application is deployed to Microsoft HoloLens 2 and the person is asked to play the game that loads in front of them. This is intended as a first step that will pave the way towards effective and intuitive interaction with assistive robotic systems. The aim of this work is to help elderly people with dementia keeping their brain active and ageing in an environment filled with opportunities, rather than limitations.

Keywords: Mixed Reality · HoloLens · Cognitive Training

1 Introduction

The global population is ageing at a significant pace and the percentage of older adults is expected to increase to 24% by 2030. As a result of this growth, it is also increasing the number of individuals affected by dementia [13].

Dementia is a term that encompasses several diseases that affect the ability to perform daily activities. Dementia has physical, psychological, social and economic impacts not only on people diagnosed with it, but also on their families

This work is supported by the "Lively Ageing" project, funded by the Italian Ministry of Health.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 107–111, 2024.
https://doi.org/10.1007/978-3-031-76428-8_21

and their carers. Currently, more than 55 million people have been diagnosed with dementia worldwide, with Alzheimer disease being the most common form of dementia, contributing to 60–70% of cases. Every year, there are 10 million new cases, making this a crucial difficulty to face [4].

The growth in the ageing population is combined with a longer life expectancy, resulting in higher costs and challenges associated with providing care for the elderly. Among these challenges, we find the constant need for more professional caregivers and shrinking workforce.

Currently, there is no cure for dementia, but there are alternative forms of care that aim to improve the wellbeing of people with dementia by being physically active and involved in activities and social interactions that keep their mind engaged and maintain daily function.

Technology has proven to be an innovative resource to support patients affected by dementia, their families and their caregivers in many assisting tasks. Several studies have shown that interactions with human-like or animal-like robots can help improve the quality of life of elderly people with dementia. In the literature, there is also evidence of the effectiveness of virtual reality therapy for patients with dementia [6]. These positive results include a reduction in depressive symptoms, anxiety symptoms, and a decrease in negative physiological factors [7,9,12]. Simultaneously, an improvement in factors related to quality of life has been observed, along with an enhancement in interaction and communicative actions [10,11].

Another important aspect of technology as an aid for people with dementia is the feeling of independence. These patients may feel like they are loosing their autonomy, having to rely on caregivers and family members for daily tasks. Technology can assist them during these activities, giving them the feeling of being independent.

Among the technologies that can be used, mixed reality (MR) can offer people with dementia cognitive stimulation and memory aid. MR merges virtual reality and augmented reality, creating an environment in which real and virtual objects and subjects coexist and interact. MR allows more immersive and dynamic interactions between the user and the holograms, integrated as part of the real world.

Given the importance of supporting people with dementia and helping them improve their overall well-being, this paper introduces a MR application, with the aim of enhancing short-term memory and visuo-spatial skills. We present a therapeutic activity in the form of a game, in which the user needs to collect red and green cubes and place them on the same-colored surface. Our application is created using Unity as development platform and is deployed to Microsoft HoloLens 2, a wearable MR device that enables users to interact with holograms.

2 Mixed Reality Application

This proposed MR application encourages movement in a known environment, helping to counteract the onset or progression of visuo-spatial disorders. Furthermore, it also stimulates cognitive functioning by requesting users to categorize the objects they collect.

2.1 Interaction Task

The aim of this activity is to stimulate the user's cognitive functioning and visuo-spatial abilities. At the beginning of the activity, the scene is loaded using the user's head as origin of the spatial reference system. As will be explained later in Sect. 2.2, the first thing the person needs to do is to change the origin of the coordinate system by scanning a QR code positioned in the room. The objects displayed in the scene will then adjust their position with respect to the new origin. After this setup phase, the activity can start.

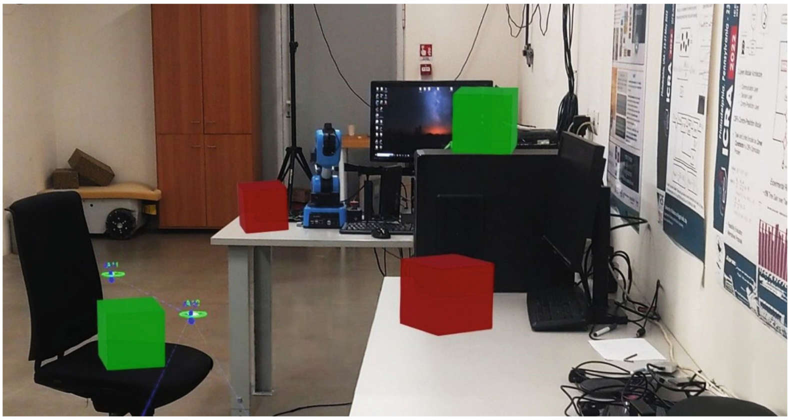


Fig. 1. Interaction task. The cubes are spread across the room, blending in with the physical environment.

Several red and green cubes are spread across the room, as shown in Fig. 1. The holograms are placed on the floor, on tables or other elements, blending in with the physical environment. The user is required to move within the space, collect the cubes and position them on the same-colored platform. Once the object is correctly placed, the counter located above the platform updates and a sound effect is emitted. The game ends when all the cubes are placed on the right platform.

2.2 Development of the MR Environment

The application was developed using Unity, a real-time 3D development platform for building 2D and 3D applications, like games and simulations [3]. Unity uses Microsoft C# programming language and allows developers to create custom scripts for a game or an app.

For the generation of this application we used the Mixed Reality Toolkit (MRTK), which is an open-source toolkit developed by Microsoft for Unity [2]. It is used to accelerate the development of immersive cross-platform MR applications in Unity by providing a wide range of components, scripts and predefined resources that assist the developers. Among its common features, we can find gesture recognition, interaction with the real world and manipulation of virtual objects.

For the development of the app we also used the World Locking Tool (WLT), which provides a stable and reliable world-locked coordinate system [5]. It anchors the virtual world to the physical one by placing holograms in specific positions relative to real physical elements and other holographic objects. The implementation of this tool, particularly its *Space Pins* feature, has thus enabled the creation of an application that aligns the Unity space with the real world and addresses the scale error in holographic size perception on HoloLens applications.

Once the application is created in the Unity development environment, it is then deployed to HoloLens 2.

Microsoft HoloLens 2 is a wearable MR headset that enables users to interact with holograms integrated in the visualization of the real world [1]. It allows high-quality, high-fidelity, hands-free interaction. The MR headset has see-through holographic lenses that use a sophisticated optical projection system to generate multi-dimensional holograms with extremely low latency. HoloLens 2 is equipped with cameras and sensors, such as depth sensors, accelerometer and gyroscope, which are designed to capture information concerning what the user is doing and the surrounding environment.

For HoloLens applications, the origin of the coordinate system is located at the center of rotation of the user's head, meaning that holograms are positioned in space relative to the user's head height. Although this ensures a more immersive experience, it also makes the application entirely tailored to the specific user. Thus, it is challenging to design an activity that is equally accessible to everyone.

To overcome this issue, we used the *Space Pins* feature of the WLT. We started with a sample project provided by Microsoft [8], which allows to recognize QR codes placed in the physical environment and to transform the coordinate system accordingly, setting the scanned point as the new origin. Building upon this project, we then developed the rest of our application, changing the structure of the scene and its components according to our requirements.

3 Conclusions and Future Developments

The ageing of the population and the consequent increase in the number of elderly people affected by dementia is a long-term demographic trend that entails several challenges and problems. Technology can be an ally in addressing these challenges, supporting patients, families, and caregivers with a broad range of capabilities used in aged care.

The focus of this work is investigating the effectiveness of MR applications as a support for elderly people who might be subject to dementia, by designing activities aimed at maintaining and preserving cognitive functioning and visuo-spatial skills.

The next step of our research is to conduct a test that both verifies the feasibility and the usability of our therapeutic MR application. This will be initially accomplished by introducing elderly subjects to Microsoft HoloLens 2 and seeing how they interact with the device. Afterwards, the user will test the application in a safe and known environment.

An important future step for our work would be to have elderly individuals use the application daily, aiming to observe if there are actual benefits resulting from the daily engagement in the activity.

As a further step, we aim at building upon this concept, to integrate the MR application with social robots: the main idea is that of leveraging on the ease of interaction and on the flexibility of application provided by MR technologies to pave the way towards extensive deployment of social robots to support elderly people in daily-life contexts.

References

1. Microsoft hololens 2. https://www.microsoft.com/en-us/hololens
2. Mixed reality toolkit. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/
3. Unity development platform. https://unity.com/
4. World health organization. https://www.who.int/news-room/fact-sheets/detail/dementia
5. World locking tool. https://learn.microsoft.com/en-us/mixed-reality/world-locking-tools/documentation/concepts
6. Appel, L., et al.: Virtual reality to promote wellbeing in persons with dementia: a scoping review. J. Rehabilitation Assistive Technol. Eng. **8**, 205566832110539 (2021)
7. Chen, S.C., Moyle, W., Jones, C., Petsky, H.: A social robot intervention on depression, loneliness, and quality of life for taiwanese older adults in long-term care. Int. Psychogeriatr. **32**, 981–991 (2020)
8. Finch, M.: QR space pins sample (2022). https://github.com/microsoft/MixedReality-WorldLockingTools-Samples/tree/master/Advanced/QRSpacePins
9. Gustafsson, C., Svanberg, C., Müllersdorf, M.: Using a robotic cat in dementia care: a pilot study. J. Gerontol. Nurs. **41**(10), 46–56 (2015)
10. Joranson, N., Pedersen, I., Rokstad, A.M., Ihlebæk, C.: Change in quality of life in older people with dementia participating in paro-activity: a cluster-randomized controlled trial. J. Advanced Nurs. **72** (2016)
11. Jøranson, N., Pedersen, I., Rokstad, A.M.M., Aamodt, G., Olsen, C., Ihlebæk, C.: Group activity with paro in nursing homes: systematic investigation of behaviors in participants. Int. Psychogeriatr. **28**(8), 1345–1354 (2016)
12. Petersen, S., Houston, S., Qin, H., Tague, C., Studley, J.: The utilization of robotic pets in dementia care. J. Alzheimer's Disease: JAD **55** (2016)
13. Shishehgar, M., Kerr, D., Blake, J.: A systematic review of research into how robotic technology can help older people. Smart Health **7–8**, 1–18 (2018)

A Mixed Reality Interface for Human-Swarm Interaction

Mattia Catellani(✉), Flavia Nironi, and Lorenzo Sabattini

Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Reggio Emilia, Italy
{mattia.catellani,lorenzo.sabattini}@unimore.it

Abstract. The increasing deployment of multi-robot systems underscores their potential across diverse research and applied domains. Despite advancements in robot autonomy, challenges persist in achieving full autonomy in certain scenarios. In this paper, we propose leveraging human-robot interaction to intelligently combine human and robot capabilities. By assigning a human operator the role of supervisor, robots focus on environmental data retrieval, enhancing safety and task execution efficiency. This paper presents a Mixed Reality solution utilizing the Microsoft HoloLens 2 headset, allowing a human operator to designate specific areas for a multi-robot team to reach and cover.

Keywords: Human-Robot Interaction · Mixed Reality · Multi-Robot Systems

1 Introduction

Recent years have seen a notable growth in the deployment of multi-robot systems (MRSs), indicating their growing potential in a variety of research and applied fields [2]. One of the most extensively researched applications where MRSs are successfully used is coverage control [1], whose goal is to optimize the deployment for monitoring an area. Even though research has been moving towards increasing robots autonomy, there are still many situations where robots are unable to operate fully autonomously. An interesting approach to overcome existing limitations could be to intelligently exploit human-robot interaction (HRI) in order to combine their capabilities, instead of aiming for a complete automation of the task. As pointed out in [3], a human operator can be assigned with the role of supervisor, exploiting their intelligence and their possibly external point of view, while robots can concentrate on retrieving information from the environment and fulfill the operation. In addition, this strategy enhances safety for the human being, demanding the possibly dangerous on-field execution of the task to the robots. Recently developed Augmented Reality (AR) or

This work was supported by Horizon Europe program under the Grant Agreement 101070351 (SERMAS).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 112–117, 2024.
https://doi.org/10.1007/978-3-031-76428-8_22

Mixed Reality (MR) interfaces offer a promising approach for HRI, thanks to the possibility to define a shared environment where humans and robots can interact and collaborate. MR devices such as Microsoft HoloLens, in particular, offer an immersive experience to the user, allowing different types of interactions, not only limited to a touch input on a screen.
In this paper, we develop a MR solution to interact with a MRS. In our implementation, a human operator is equipped with a Microsoft HoloLens 2 headset, enabling them to designate a specific area with a desired shape to be reached and covered by the team.

2 Preliminaries

Notation and Definitions. In the rest of the paper, we denote by $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ the set of natural, real, real non-negative, and real positive numbers, respectively. Given the matrix $\Sigma \in \mathbb{R}^{n \times m}$, we define $|\Sigma|$ as its determinant. The environment where the robots are supposed to operate will be denoted as $Q \subset \mathbb{R}^2$. An arbitrary point in Q is denoted by $q \in Q$.

Assumptions. Consider a MRS composed by n robots moving in two dimensions, controlled in order to reach a specific area of the environment. We assume each robot to be modeled as a single integrator system, whose position $p_i \in \mathbb{R}^2$ evolves according to $\dot{p}_i = u_i$, where $u_i \in \mathbb{R}^2$ is the control input, $\forall i = 1, \dots, n$.

We assume robots are able to localize themselves and other robots detected inside their circular sensing region, defined by a radius $R_s \in \mathbb{R}^2$. Communication among them is not allowed.

3 Background

In this section, we will briefly present the fundamentals of our control strategy. More in details, we will present Gaussian Mixture Models and coverage control, which will be used to define an environment's probability density and to drive robots toward the specified area, respectively.

3.1 Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is a probabilistic model that assumes that a sample set is generated from a combination of k Gaussian distributions, each characterized by a mean point $\mu_i \in \mathbb{R}^2$ and a covariance matrix $\Sigma_i \in \mathbb{R}^{2\times 2}$ (see [4]). The overall model is created by combining these components using specific weighting factors $w_i \in \mathbb{R}_{>0}$, with $\sum_{i=1}^{k} w_i = 1$. When a user draws a polygon $S \subset Q$ in the environment with a Mixed Reality (MR) device, a GMM fitting the desired shape can be estimated using *Maximum Likelihood* methods as detailed in [5]. This approach employs an *Expectation-Maximization* algorithm to iteratively determine the best parameters (μ, Σ, w) from the sample distribution.

Once the GMM has been defined, one can calculate the probability function for each individual component as follows:

$$\phi_i(q, \mu_i, \Sigma_i) = \frac{\exp\left(-\frac{1}{2}(q-\mu_i)\Sigma_i^{-1}(q-\mu_i)^T\right)}{\sqrt{|\Sigma_i|(2\pi)^d}}. \tag{1}$$

Finally, the overall probability function can be calculated as the weighted sum of each component:

$$\Phi(q, \mu, \Sigma) = \sum_{i=1}^{k} w_i \phi_i\left(q_i, \mu_i, \Sigma_i\right). \tag{2}$$

In our work, we utilize GMMs to specify a specific probability density so that robots are guided into the targeted areas, aiming at assigning a higher density to the region selected by the human operator. The samples set for the fitting operation is retrieved discretizing the region of interest.

3.2 Coverage Control

After explaining how the region defined by the human operator shapes the probability density, we'll now describe how this density guides robots to the target area. Our approach is based on Voronoi-based coverage control, where robots position themselves optimally to maximize coverage over the specified probability density. This technique uses Lloyd's algorithm to achieve an ideal configuration that covers as much area as possible. To make this work, we divide the entire area into cells using Voronoi partitioning, assigning specific regions to each robot. Since our robots have limited sensors and no global knowledge of their teammates' positions, we compute the Voronoi partitioning in a decentralized manner based on directly measurable data. This requires using a constrained Voronoi partitioning method as defined in [6]. Each robot, as a result of this limited Voronoi partitioning, obtains a cell $V_i \subset Q$ where it has full sensor coverage. Subsequently, every robot autonomously calculates its cell's centroid, considering the probability function (2).

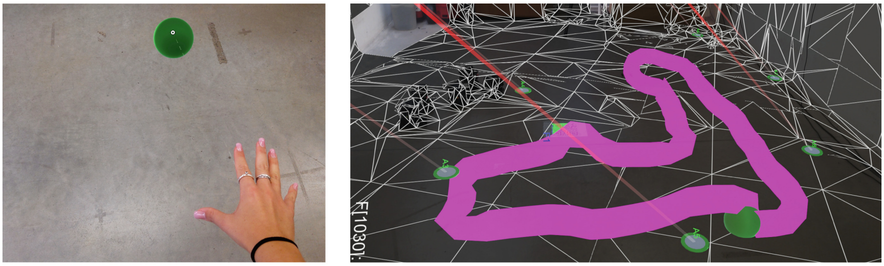
$$C_{V_i} = \frac{\int_{V_i} q\Phi(q, \mu, \Sigma)dq}{\int_{V_i} \Phi(q, \mu, \Sigma)dq}. \tag{3}$$

Finally, the control action can be calculated in order to move each robot towards the centroid of its cell with a proportional law $u_i = -k_p(p_i - C_{V_i})$, with $k_p \in \mathbb{R}_{>0}$.

This control strategy brings the team in an optimal configuration, maximizing the area coverage and consequently the information gathering from the environment.

4 Control Architecture

In this section, we will present the tools employed to build up the MR interface for HRI. As previously stated, the goal of our implementation is to create a



(a) (b)

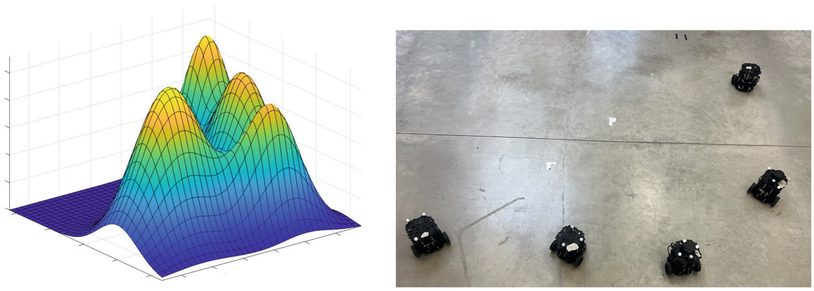
Fig. 1. (a) Hand gestures for cursor manipulation. (b) Region of interest drawn in the MR environment.

shared environment for a MRS and a human operator, in order to combine their capabilities and fulfill the task. In particular, the goal of the presented solution is to allow a human to define a desired region of interest inside the environment, and move the robots into this area in order to retrieve information, exploiting human intelligence for the high-level planning of the task while demanding the on-field execution to the robotic team.

In our work, we make use of Microsoft HoloLens 2 headset featuring a see-through display for Mixed Reality. A virtual scene has been created with Unity, featuring a virtual sphere to be used as a cursor by the operator, who can manipulate the cursor with specific hand gestures, as shown in Fig. 1a. QR codes have been placed in known locations in the real environment, to be detected by the HoloLens cameras in order to align the reference frame of the virtual scene with the real-world reference frame. While moving the sphere around, a trail is drawn to show the area that is being selected (see Fig. 1b), and the coordinates of the cursor are continuously sent to an external computer to keep track of the movement. Finally, once the elaboration system has fitted a GMM to the region that the human operator has defined, it communicates the GMM parameters to the robots, making them move in accordance with the coverage control algorithm outlined in Sect. 3.2. Robots and the external computer exploit ROS 2 to run the control software and to establish communication with each other.

5 Experimental Evaluation

Finally, we demonstrate the execution of an experiment where a human operator must identify a region in a real-world environment that has to be monitored by robots, with the aim of evaluating the effectiveness of the proposed architecture. In order to monitor the robots using a motion capture system, the user must wear the HoloLens 2 headset and scan a QR code to align the inertial reference frame with the global one, which is specified in the real world. Then, they can grab the virtual sphere and encircle the desired area to be reached, resulting in the final shape, as shown in Fig. 1b.



(a) (b)

Fig. 2. (a) Probability density associated to the drawn region of interest. (b) Final configuration of the MRS.

Then, the external computer fits a GMM to the selected region, resulting in the probability distribution shown in Fig. 2a, where the higher attractive effect of the area encircled by the human operator is clearly visible. Finally, robots are controlled as described in Sect. 3.2. A sensing range $R_s = 2$ m is considered for each robot, and a random starting position is chosen for each of them. As we can see from Fig. 2b, the final configuration assumed by the robotic team fits the desired area defined by the operator, and robots optimally spread in order to maximize the covered surface of the region of interest.

6 Conclusions

In this work we proposed a MR solution to interact with a MRS. Our solution makes use of Microsoft HoloLens 2 headset to encircle a region of the environment to be reached by the robots. We performed a qualitative analysis of results employing real mobile robots, showing that the MRS successfully reaches the desired area. In future works, we intend to gather quantitative data to examine the effectiveness of our solution. Additionally, we plan to carry out user studies to evaluate the architecture's usability and assess the degree to which people feel comfortable engaging with robots in a shared environment.

References

1. Cortes, J., Martinez, S., Karatas, T., Bullo, F.: Coverage control for mobile sensing networks. IEEE Trans. Robot. Autom. **20**(2), 243–255 (2004)
2. Dorigo, M., Theraulaz, G., Trianni, V.: Swarm robotics: past, present, and future [point of view]. Proc. IEEE **109**(7), 1152–1165 (2021)
3. Hussein, A., Abbass, H.: Mixed initiative systems for human-swarm interaction: opportunities and challenges. In: 2018 2nd Annual Systems Modelling Conference (SMC), pp. 1–8. IEEE (2018)
4. Kotz, S., Balakrishnan, N., Johnson, N.L.: Continuous Multivariate Distributions, Volume 1: Models and Applications, vol. 1. Wiley (2004)

5. McLachlan, G.J., Lee, S.X., Rathnayake, S.I.: Finite mixture models. Ann. Rev. Stat. Appl. **6**, 355–378 (2019)
6. Pratissoli, F., Capelli, B., Sabattini, L.: On coverage control for limited range multi-robot systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2022)

CBF-Based STL Motion Planning for Social Navigation in Crowded Environment

Andrea Ruo(✉), Lorenzo Sabattini, and Valeria Villani

University of Modena and Reggio Emilia, 42122 Reggio Emilia, RE, Italy
{andrea.ruo,lorenzo.sabattini,valeria.villani}@unimore.it

Abstract. A motion planning methodology based on the combination of Control Barrier Functions (CBF) and Signal Temporal Logic (STL) is employed in this paper. This methodology allows task completion at any point within a specified time interval, considering a dynamic system subject to velocity constraints. In this work, we apply this approach into the context of Socially Responsible Navigation (SRN), introducing a rotation constraint. This constraint is designed to maintain the user within the robot's field of view (FOV), enhancing human-robot interaction with the concept of side-by-side human-robot companion. This angular constraint offers the possibility to customize social navigation to specific needs, thereby enabling safe SRN. Its validation is carried out through simulations demonstrating the system's effectiveness in adhering to spatio-temporal constraints, including those related to robot velocity, rotation, and the presence of static and dynamic obstacles.

Keywords: Robot guidance · Socially-responsible navigation · Control barrier function · Signal temporal logic

1 Introduction

In the recent years, an increasing number of robots have entered human environments. To navigate in these places, the robot needs to be aware of the humans in the environment, and treating humans simply as obstacles may not be enough. Furthermore, the robot's motion should be safe, legible and acceptable to humans rather than being optimal from the sole point of view of the robot [1]. In SRN context, a mobile robot operates autonomously within an environment, facing the challenge of navigating a path successfully while avoiding collisions with obstacles in the environment [2]. Certain social spaces can be wide and crowded. In these scenarios, it becomes crucial for robots to move exhibiting appropriate social behaviors [3], such as the side-by-side human-robot companion. Specifically, when guiding the user to their destination, the robot can stand next to

This work was supported by Horizon Europe program under the Grant Agreement 101070351 (SERMAS).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 118–123, 2024.
https://doi.org/10.1007/978-3-031-76428-8_23

or in front of the human, while moving forward. This factor can significantly impact the acceptance of human-robot interaction.

Recent literature has explored similar concepts in the context of SRN, investigating various approaches and examples. The SPENCER project, funded by the European Union [4], has developed a reception robot tailored to assist, inform, and guide passengers in large and crowded airports. This robot combines map representation, laser-based people and group tracking, and activity and motion planning. Stricker et al. [5] proposed a robot-based information system for a university building. The reception robot provides information about offices, employees, and laboratories in the building and can guide visitors to their desired locations. In the future, we expect to see social robots sharing urban areas with people. To achieve this integration, robots have to develop several skills, including the ability to navigate alongside with humans [6]. Research on human-robot side-by-side navigation is relatively new compared to traditional robot navigation, in which robots navigate in a safe and human-like manner.

In the given application context, it is essential to include safety-related constraints, including obstacle avoidance, velocity limits, and speed reduction when the robot is in close proximity to people. Furthermore, space-time constraints might be relevant to guarantee the efficient and safe execution of activities, particularly social environments. Temporal constraints can appear in various manifestations, such as time limits to complete a specific task, time intervals to complete a sequence of tasks, or priorities assigned to different activities based on their importance. These constraints might be stipulated by environmental necessities or user preferences.

To address and implement the discussed constraints, tools such as STL and CBF can be valuable. Temporal logics, such as STL [7], enable the specification of spatio-temporal constraints, enhancing the expressiveness of Boolean logic by incorporating the temporal dimension. This allows the use of specific expressions as constraints, such as "*the robot must reach the goal pose within 10 s*". The relationship between the semantics of an STL task and time-varying CBFs enables the formal control of systems, ensuring compliance with spatio-temporal constraints and maintaining safety.

This paper focuses on advancing the CBF-based STL motion planning methodology, first presented in [8]. The methodology, originally designed for general motion planning with temporal constraints, is now applied in the domain of SRN. In doing so, we introduce and validate a angular constraint. This addition is aimed at enhancing human-robot interaction by ensuring the user remains within the robot's FOV. This contribution enables personalized social navigation, especially when the robot serves as the user's companion, ensuring visual contact to enhance engagement and safety in SRN.

2 Preliminaries and Problem Statement

Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{u} \in \mathcal{U} \subseteq \mathbb{R}^m$ be the state and input of a nonlinear input-affine control system:

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u} \tag{1}$$

Referring to the work in [8], the problem under consideration in this paper can be stated as follows.

Problem 1. Given the dynamical system in (1) *and an STL fragment* ϕ *[9], derive a control law* $\boldsymbol{u}(t)$ *so that the solution* $\boldsymbol{x} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n$ *to (1) is such that* $(\boldsymbol{x}, 0)$ *satisfies* ϕ *providing safety-critical guarantees regarding non-linear velocity constraints, angular constraint and obstacle avoidance.*

3 Approach

In this paper we focus on the need of keeping the user in a limited angular sector, with amplitude β, in order to be detected by the robot, as shown in Fig. 1.

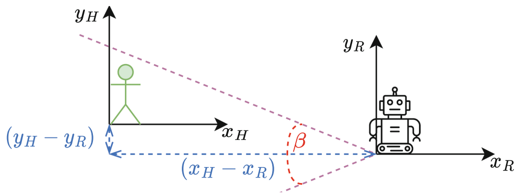


Fig. 1. Angular constraint.

A CBF is defined starting from the relative position of the detected user H respect to the robot R, expressed as ${}^R\boldsymbol{p}_H = [x_H - x_R, y_H - y_R]^T$, with two separates components, one from the left boundary of the FOV (i.e., the lower dashed line in Fig. 1) and one from the right boundary (i.e., the upper dashed line in Fig. 1) [10]. The two components are treated as different constraints inserted into an optimization solver. To this end, we introduce the following $h(\cdot)$ for the robot:

$$h({}^R\boldsymbol{p}_H) = \begin{bmatrix} h_1({}^R\boldsymbol{p}_H) \\ h_2({}^R\boldsymbol{p}_H) \end{bmatrix} = - \begin{bmatrix} tan(\frac{\beta}{2}) & 1 \\ tan(\frac{\beta}{2}) & -1 \end{bmatrix} {}^R\boldsymbol{p}_H. \tag{2}$$

Imposing the condition $h(\cdot) \geq \mathbf{0}$ is equivalent to constraining the robot's orientation such that the user is in the angular sector β, as shown in Fig. 1. The time derivative of this formulation is then expressed as

$$\dot{h}({}^R\boldsymbol{p}_H, \hat{\boldsymbol{u}}) = \frac{\partial h({}^R\boldsymbol{p}_H)}{\partial {}^R\boldsymbol{p}_H} {}^R\dot{\boldsymbol{p}}_H = - \begin{bmatrix} tan(\frac{\beta}{2}) & 1 \\ tan(\frac{\beta}{2}) & -1 \end{bmatrix} {}^R\dot{\boldsymbol{p}}_H. \tag{3}$$

The velocity of H with respect to R is then expressed through kinematic computations:

$${}^R\dot{\boldsymbol{p}}_H = {}^R R_W {}^W\dot{\boldsymbol{p}}_H - \begin{bmatrix} {}^R v_x \\ {}^R v_y \end{bmatrix} + \omega \begin{bmatrix} y_H - y_R \\ -(x_H - x_R) \end{bmatrix}. \tag{4}$$

To solve Problem 1, it is possible to integrate a dedicated CBF that enables the execution of a side-by-side human-robot companion in the quadratic problem described in [8], obtaining the formulation expressed in (5). In conclusion, we can obtain $\boldsymbol{u}(\boldsymbol{x}, t)$ as

$$
\begin{aligned}
&\min_{\hat{\boldsymbol{u}} \in \mathcal{U}} \hat{\boldsymbol{u}}^T Q \hat{\boldsymbol{u}}, \\
\text{s.t. } &\frac{\partial \mathfrak{b}(\boldsymbol{x}, t)}{\partial \boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{x})\hat{\boldsymbol{u}} + \frac{\partial \mathfrak{b}(\boldsymbol{x}, t)}{\partial t} \geq -\alpha(\mathfrak{b}(\boldsymbol{x}, t)), \\
&\|v_x + v_y\| \leq v_{max}, \\
&\frac{\partial h({}^R\boldsymbol{p}_H)}{\partial {}^R\boldsymbol{p}_H} {}^R R_W {}^W\dot{\boldsymbol{p}}_H - \frac{\partial h({}^R\boldsymbol{p}_H)}{\partial {}^R\boldsymbol{p}_H} \begin{bmatrix} {}^R v_x \\ {}^R v_y \end{bmatrix} + \frac{\partial h({}^R\boldsymbol{p}_H)}{\partial {}^R\boldsymbol{p}_H} \omega \begin{bmatrix} y_H - y_R \\ -x_H + x_R \end{bmatrix} \geq -\alpha(h({}^R\boldsymbol{p}_H)),
\end{aligned}
\tag{5}
$$

where the first constraint formulates the CBF-STL constraint to ensure the completion of a series of tasks within specified time intervals while ensuring obstacle avoidance. The second constraint imposes that the squared norm of the non-linear velocity is less than or equal to v_{max}. The third constraint guarantees that the user is always in the robot's FOV.

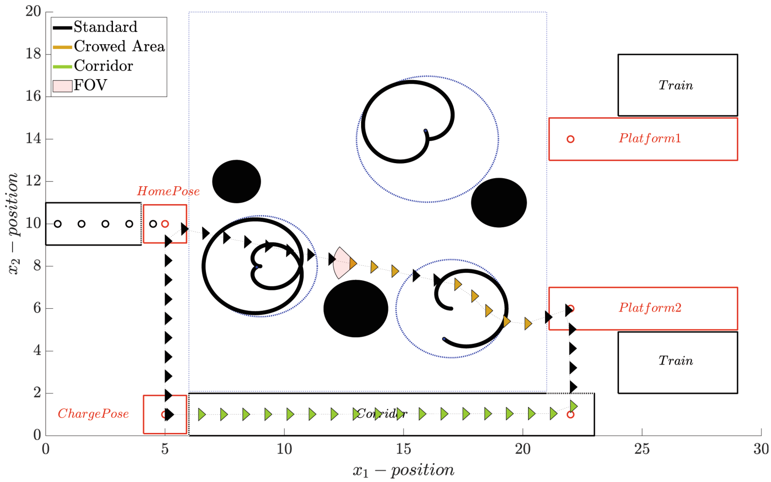


Fig. 2. The robot starts from *ChargePose* and, upon detecting the user, moves to *HomePose*. The robot guides the user to the *Platform2* while ensuring velocity constraints, angular constraints, and collision avoidance. Subsequently, the robot returns to the home pose through an obstacle-free corridor.

4 Simulation Results

We simulated an SRN scenario in Matlab, shown in Fig. 2 and in the accompanying video[1], using a similar simulation proposed in [8]. The input control $\boldsymbol{u}$ are the angular velocities of the robot, considering the three-wheeled omnidirectional robot model described in [11]. We assumed maintaining the user at an arbitrary distance with a certain level of noise from the robot.

Using the proposed approach, it is possible to observe the results of motion planning, which allows for the identification of a valid path for the robot and the satisfaction of STL specifications subject to non-linear velocity constraints, angular constraint, ensuring the compliance with safety guarantees.

5 Conclusion

In this work, we introduced an extension to our prior research, incorporating a new constraint to facilitate side-by-side interaction between humans and robots. We presented a simulation to validate the proposed approach. As a further development, we are going to integrate prediction methods to enable the robot to adjust its motion based on the behavior of the person being accompanied.

References

1. Singamaneni, P.T., Favier, A., Alami, R.: Human-aware navigation planner for diverse human-robot interaction contexts. In: International Conference on Intelligent Robots and Systems (IROS) (2021)
2. Song, K., et al.: Navigation control design of a mobile robot by integrating obstacle avoidance and lidar slam. In: IEEE International Conference on Systems, Man and Cybernetics (SMC) (2018)
3. Silva, S., Verdezoto, N., Paillacho, D., Millan-Norman, S., Hernández, J.D.: Online social robot navigation in indoor, large and crowded environments. In: International Conference on Robotics and Automation (ICRA) (2023)
4. Triebel, R., et al.: SPENCER: a socially aware service robot for passenger guidance and help in busy airports. In: Field and Service Robotics: Results of the 10th International Conference. Springer (2016)
5. Stricker, R., et al.: Interactive mobile robots guiding visitors in a university building. In: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (2012)
6. Repiso, E., Ferrer, G., Sanfeliu, A.: On-line adaptive side-by-side human robot companion in dynamic urban environments. In: International Conference on Intelligent Robots and Systems (IROS) (2017)
7. Maler, O., Nickovic, D.: Monitoring temporal properties of continuous signals. In: International Symposium Formal Techniques in Real-Time and Fault-Tolerant Systems. Springer (2004)

[1] DOI: https://doi.org/10.5281/zenodo.10255241.

8. Ruo, A., Sabattini, L., Villani, V.: CBF-based motion planning for socially responsible robot navigation guaranteeing STL specification. In: 2024 European Control Conference (ECC), pp. 122–127 (2024). https://doi.org/10.23919/ECC64448.2024.10591282
9. Lindemann, L., Dimarogonas, D.V.: Barrier function based collaborative control of multiple robots under signal temporal logic tasks. IEEE Trans. Control Netw. Syst. (2020)
10. Bertoncelli, F., Radhakrishnan, V., Catellani, M., Loianno, G., Sabattini, L.: Directed graph topology preservation in multi-robot systems with limited field of view using control barrier function. Submitted to IEEE Access (2023)
11. Liu, Y., Zhu, J.J., Williams, R.L., II., Wu, J.: Omni-directional mobile robot controller based on trajectory linearization. Robot. Auton. Syst. (2008)

Multimodal Analysis of User Engagement with a Recommender Robot in Cafe Settings

Yujin Li[1], Nguyen Tan Viet Tuyen[1,2], and Oya Celiktutan[1](✉)

[1] Department of Engineering, King's College London, London WC2R 2LS, UK
eugeneliuk@outlook.com, {tan_viet_tuyen.nguyen,oya.celiktutan}@kcl.ac.uk
[2] School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BH, UK

Abstract. Predicting user engagement is essential to ensure natural human-robot interactions (HRI) and to enhance social acceptance of robots. Motivated by this, this paper introduces a framework for predicting human engagement during HRI in retail settings. Specifically, we (1) introduced novel annotations for the LISI-HRI dataset, a new in-the-wild dataset comprising interactions between a recommender robot and customers of a cafeteria, to model levels of human engagement during HRI, (2) analysed the effectiveness of modelling engagement state with our selected multimodal features, and (3) designed and evaluated a Support Vector Machine (SVM) based user engagement prediction approach. The obtained results aim to serve as a baseline on this dataset for engagement prediction in real-world human-robot interaction scenarios.

Keywords: User Engagement Prediction · Human-Robot Interaction · Social Robots · Multimodal Learning

1 Introduction

Recognising user engagement and adjusting robots' communicative behaviours are essential for robots to increase their social acceptance when interacting with users [1]. However, it is challenging for robots to accurately predict user engagement levels using their non-verbal signals observed by robots in a crowded environments [2]. In this paper, we first addressed the definition of user engagement level in the cafeteria HRI context and then, selected a set of human nonverbal features for modeling their engagement state. Next, we utilised the LISI-HRI dataset [3] and annotated customer engagement levels, considering their body motion, head pose, and emotional features. The LISI-HRI dataset [3] was collected through interaction scenarios between customers and a humanoid robot (i.e., Pepper) in a cafeteria setting. Finally, this paper aims to clarify two questions: (1) Which nonverbal feature collected in the crowded environment works best? To verify that, we used both single and joint features as inputs for machine

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 124–129, 2024.
https://doi.org/10.1007/978-3-031-76428-8_24

learning classifiers. (2) What is the correlation between user nonverbal features and their engagement levels? To answer these questions, an analysis was conducted to unveil characteristics of engagement states.

2 Related Work

There is a significant amount of work on engagement prediction in human-machine interaction as summarised in a recent survey paper [2]. The most related work used non-verbal cues, such as head motions, facial expressions, and eye gaze, to model user engagement initiated with small talk in a public space [4]. They integrated Deep Neural Networks (DNNs) for real-time engagement prediction, with a primary focus on identifying disengagement rather than discerning specific engagement levels. Another work [1] introduced a novel bartending robot architecture, customising drink recommendations based on user engagement inferred from body motion. Our study extends prior research into a novel HRI setting-the cafeteria environment. To our knowledge, no prior investigations have addressed how to define user engagement specifically in this context. Additionally, our work explores potential features for modelling user engagement in this unique setting. Consequently, there are gaps in predicting user engagement in certain human-robot interaction settings, such as cafes, which requires clear guidelines to define different levels of engagement and quantitative analysis of visual feature contributions in engagement level prediction.

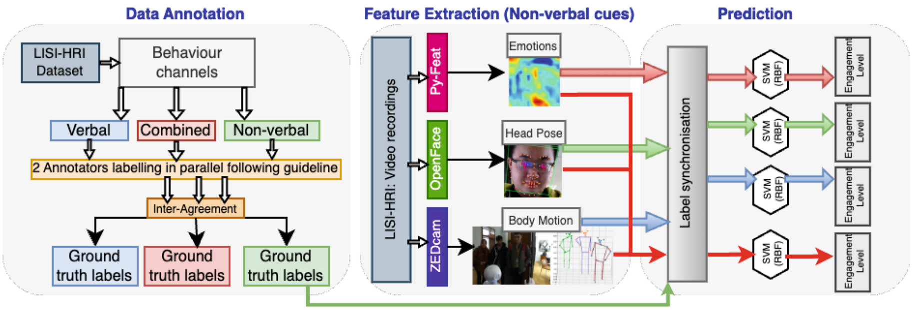


Fig. 1. The multimodal approach for engagement prediction.

3 A Multimodal Approach for Engagement Prediction

The overview of the proposed approach is illustrated in Fig. 1.

3.1 Dataset Annotation

This paper extends the LISI-HRI [3] dataset by annotating customer engagement and using it for training and testing the designed framework. The dataset

Table 1. Average Recall score of each model from cross validation.

Model	SVM(RBF)	RF	NB
Avg Recall Scores	0.58	0.46	0.48

was collected in a real-world cafeteria setting, where Pepper was equipped with human-like behaviors and served at the cafe shop as a marketing agent. Multiple modalities, including conversation scripts, touch-screen events, audio, and video, were collected from both external and the robot's onboard sensors. They were synchronised together to form a 7-hour multimodal HRI dataset.

Building upon annotation methods proposed by authors of [4,5], we designed a rule-based annotation scheme involving two annotators, who are postgraduate students from a university in the UK, to determine engagement levels through verbal and nonverbal cues collected from the dataset. Initially, two annotators independently label segments for each modality. We retain segments only if both annotators agree on the label after comparing against annotation guidelines and engaging in extensive discussions. Customer engagement levels, as defined in previous studies [5], encompass no, low, and high engagement. Each interaction session in the LISI-HRI dataset was segmented into smaller windows, ranging from 3 to 90 s, considering the interaction contexts (e.g., sports, food, weather, etc.). The annotation task was conducted using the ELAN annotation toolbox [6], with visual, audio, and a combination of visual and audio cues.

3.2 Feature Extraction and Data Processing

User engagement is complicated to model as it can be represented by different channels, including both verbal and nonverbal cues. This paper focuses on analysing nonverbal features as they are more reliable to deduce user's mental state [2]. Head pose and facial expressions have been widely used to collectively depict human engagement [4,7]. It has been shown that the combination of multiple non-verbal cues can better describe user attention and emotional states [7], and it can compensate for each other when one is missing [4]. Indeed, body posture [8] and proxemics are essential features for estimating engagement [7]. Taken all together, this paper utilises user body motion, head pose, and emotions as inputs for the HRI engagement prediction task.

Py-Feat [9] and OpenFace [10] are used to capture emotion scores and head pose information respectively. Body motion and the depth information are captured by Stereolabs ZED cameras. The extracted feature data are synchronised against the labels and merged into a multimodal dataset. After removing outliers and incomplete instances, the dataset is composed of 13 no-engagement, 107 low-engagement and 503 high-engagement labels. Due to the imbalanced class distribution, our preliminary investigation showed that machine learning algorithms produce poor results. To counteract this, an under-sampling technique, the Tomek-Link algorithm [11], was leveraged to remove ambiguous data points, providing clear decision boundaries for models. After sub-sampling, the class distribution was as follows: 13 samples of no-engagement class, 88 samples of low-engagement class and 483 samples of high-engagement class.

4 Experimental Results and Discussion

Initially, we selected optimal hyperparameters for each model with the grid search method within default range of values from Scikit-learn python toolbox [12]. After running predictions with Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB) algorithms, the results in Table 1 indicates that SVM algorithms demonstrate a performance on the classification task with the highest average recall score overall. Hence, we explored the contribution of different modalities to engagement prediction with SVM.

Figure 2 presents the prediction results from individual and joint feature inputs in terms of recall for classifying engagement levels in the cafeteria setting. Here, no-engagement, low-engagement, and high-engagement classes are denoted as 1, 2, and 3 respectively. It can be observed that body motion and head pose features contribute more to the recognition of high-engagement states, while emotion features are more reliable to identify non-engagement states. As shown in Fig. 3, the underlying reason could be that low dynamics characterise high-engagement class, implying that users were focused in the interactions, thus showing less movement. It should be emphasised that the dynamics of movements are extracted by computing standard deviations of body landmarks coordinates and head orientation angles in each data segment. Meanwhile, emotional information demonstrates a greater advantage in predicting no-engagement levels accurately. Overall, when combining all modalities, the SVM-based prediction approach yields a better result than others. This approach yields an average recall of 80% across 3 engagement levels, as depicted in Fig. 2. One of the reasonable explanations is that those features can compensate each other when one of them is missing or performing poorly on prediction, as revealed in [4].

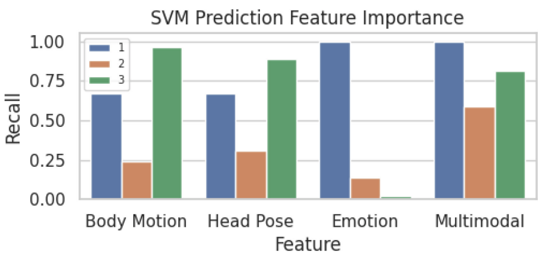
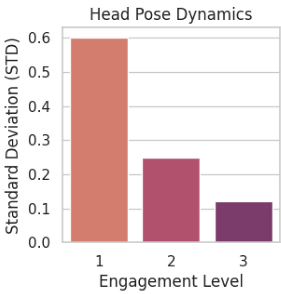
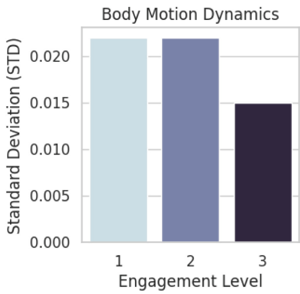


Fig. 2. Prediction results and contribution of nonverbal modalities.



(a) Head movement dynamics



(b) Body movement intensity

Fig. 3. Head and body motion dynamics among correctly predicted instances.

5 Conclusion and Future Work

Despite the inherent complexity in modelling engagement, we developed a tailored annotation scheme for the LISI-HRI dataset collected in a cafe, extracting visual features (head pose, body motion, and emotional states) for classifier training. Results reveal SVM's reliability, achieving an 80% average recall by leveraging multimodal information, suggesting a potential for an automatic, cost-effective, and accurate real-time engagement prediction system. To the best of our knowledge, our analysis is the fist study uncovering a correlation where low dynamics of body and head movement contribute to accurately depicting high-engagement states in an in-the-wild cafeteria setting. Simultaneously, emotional information excels in predicting the no-engagement level. These insights address gaps in understanding user engagement in the cafe setting and help us refine the predictive capabilities of our models.

As future work, we aim to expand the dataset and refine data quality, including addressing issues such as user faces were not adequately captured. On the other hand, due to notable imbalance between low- and high-engagement data points, a binary classification approach, focusing on engaged or not en- gaged states, will be considered. Finally, to bolster our model, we will integrate verbal and textual features, encompassing a wider range of multimodal features to describe user engagement.

Acknowledgements. This work was supported by the EPSRC LISI project (Grant Ref.: EP/V010875/1) and The Royal Society CL-HRI project (Grant Ref.: RGS\R2\212084).

References

1. John, N.E., Rossi, A., Rossi, S.: Personalized human-robot interaction with a robot bartender. In: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, pp. 155–159. Association for Computing Machinery, Barcelona, Spain (2022)

2. Salam, H., Celiktutan, O., Gunes, H., Chetouani, M.: Automatic Context Driven Inference of Engagement in HMI: A Survey (2022)
3. Tuyen, N.T.V., Adamski, M., Wang, Y., Okazaki, S., Celiktutan, O.: The impact of robot's body language on customer experience: an analysis in a cafe setting. In: Companion of the 2023 ACM/IEEE HRI, pp. 248–252 (2023)
4. Ben-Youssef, A., Varni, G., Essid, S., Clavel, C.: On-the-fly detection of user engagement decrease in spontaneous human-robot interaction using recurrent and deep neural networks. Soc. Robot. **11**, 815–828 (2019)
5. Oertel, C., Salvi, G.: A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. In: Proceedings of the 15th ICMI, pp. 99–106. ACM, Sydney, Australia (2013)
6. Brugman, H., Russel, A., Nijmegen, X.: Annotating multi-media/multimodal resources with ELAN. In: LREC, pp. 2065–2068 (2004)
7. Michalowski, M.P., Sabanovic, S., Simmons, R.: A spatial model of engagement for a social robot. In: 9th IEEE AMC, pp. 762–767 (2006)
8. Anzalone, S.M., Boucenna, S., Ivaldi, S., Chetouani, M.: Evaluating the engagement with social robots. Soc. Robot. **7**, 465–478 (2015)
9. Cheong, J.H., Xie, T., Byrne, S., Chang, L.J.: Py-Feat: Python Facial Expression Analysis Toolbox. CoRR (2021)
10. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.-P.: OpenFace 2.0: facial behavior analysis toolkit. In: FG 2018, pp. 59–66 (2018)
11. Elhassan, T., Aljurf, M.: Classification of imbalance data using tomek link (t-link) combined with random under-sampling (rus) as a data reduction method. Glob. J. Technol. Optim. S **1**, 2016 (2016)
12. Pedregosa, F., et al.: Scikit-learn: machine Learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

Key Factors for Social Acceptance of Robots in the Industrial and Service Oriented Human-Robot Interaction Domains

Silvia Proia[1](✉), Graziana Cavone[2], Raffaele Carli[1], and Mariagrazia Dotoli[1]

[1] Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy
{silvia.proia,raffaele.carli,mariagrazia.dotoli}@poliba.it

[2] Department of Civil, Computer Science, and Aeronautical Technologies Engineering, University Roma Tre, Rome, Italy
graziana.cavone@uniroma3.it

Abstract. A literature review is conducted, presenting an overview of the crucial factors that impact the robots acceptance in the human-robot interaction (HRI) settings. In particular, our goal is to differentiate between factors whose measurement can be performed through indicators already available in the literature and those requiring additional efforts for quantification and incorporation into the design and control of robots. The key elements are categorized according to two primary categories, i.e., manufacturing and service. Within the manufacturing domain, six factors are identified: (perceived) occupational safety, physical ergonomics, cognitive ergonomics, efficiency, design, and privacy. Instead, within the service domain, four factors are acknowledged: safety, design, cognitive and emotional comfort, and privacy. The results of our review indicate that consistent effort is still required to define additional indices to quantify social acceptance of robots and improve their design in this perspective.

Keywords: Social acceptance of robots · human-robot interaction · collaborative robots · social robots

1 Introduction

Collaborative robots, often referred to as *cobots*, represent a new generation of robots designed to work alongside and interact with humans beyond enclosed spaces. These robots have facilitated the realization of Industry 4.0, and are serving as the foundation for the upcoming Industry 5.0, whose focus prioritizes human needs and interests in the production process. Even if there is a steadily increasing use of cobots within industrial settings, robots directly delivering services to human users in shared workspaces, i.e., social robots or *sobots* [1], possess the highest expected growth rate. With humans and robots now

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 130–135, 2024.
https://doi.org/10.1007/978-3-031-76428-8_25

cohabiting the same space, a novel discipline has emerged within the field of robotics: human-robot interaction (HRI). In the last few years several scientists have tried to review and classify the research on collaboration and interaction between human and robot introducing new variables related to human factors and highlighting the most extensively studied concepts, i.e., safety, ergonomics and mental stress, and efficiency [2]. For instance, the progress and prospects in the collaboration between human and robot are suggested by Ajoudani *et al.* who present in their review [3] the latest advancements in bi-directional human-robot interfaces aimed at enhancing human-robot perception, specifically focusing on estimating variations in human physical or cognitive states. Physical and cognitive collaboration in the industrial environment is also the focus of the extensive review proposed by Villani *et al.* [4]. Physical and psychosocial factors that collectively pose a threat to the fundamental comfort and well-being of individuals, increasing the likelihood of diseases and injuries, and influencing their overall quality of life, are the focal points of the ergonomic review in [5]. From the above analysis of existing surveys, two crucial findings emerge. The focus of research in the realm of human factors lies in the exploration of the concepts of ergonomics and mental workload. Special attention is given to incorporating elements associated with safety and connectivity of robots. On the other hand, none of the above cited articles focuses on metrics for measuring social acceptance, which gauges the extent of comprehension and trust that humans place in the adoption of new technological solutions. In addition, to the best of the authors' knowledge, the exploration of social acceptance in the literature is limited, being only related to preliminary studies or questionnaires pertaining just to social robots functioning in the service oriented HRI domain [6,7], or only to collaborative robots working in the industrial oriented HRI domain [8–10], with few studies [11] delving into social acceptance applied to diverse fields of use. This is instead a matter of considerable significance in advancing human-centric interaction, necessitating additional exploration, especially concerning the mobility and coherence of robots. Specifically, delving further into the design of a set of movements and responses attributed to machines is a critical and indispensable aspect for enhancing operators' inclination and willingness to collaborate with them. Therefore, the evident gap in the related literature is that there is no survey exploring the key factors influencing social acceptance of both cobots and sobots. For the sake of coping with this gap, the current paper identifies the key factors to consider in making robots socially accepted, both in the industrial and service oriented HRI domains, building on insights from the related literature.

2 Factors for Robots Social Acceptance

The social acceptance concept, defined as "an individual's psychological state with regard to his or her voluntary or intended use of a particular technology" [12], is examined in the literature through the identification of key factors that make robots socially accepted by humans. Specifically, the aim is to pinpoint

Table 1. Categorization of the key factors according to two primary categories.

Category	Key Factors	Ref. No.
Manufacturing	(Perceived) Occupational Safety	[2]
	Physical Ergonomics	[5]
	Cognitive Ergonomics	[8,9]
	Efficiency	[2]
	Design	[8,9]
	Privacy	[9]
Service	Safety	[7]
	Design	[6]
	Cognitive and Emotional Comfort	[7]
	Privacy	[10]

factors that have already been measured by utilizing metrics present in the literature and those requiring additional efforts for quantification and incorporation into the design and control of robots. Currently, robots are employed in both industrial and service oriented HRI domains. Therefore, the present review on acceptance of robots results in the primary classification of the works according to two categories, i.e., manufacturing and service (see Table 1).

In the realm of manufacturing, six essential acceptance factors are recognized, specifically, (perceived) occupational safety, physical ergonomics, cognitive ergonomics, efficiency, design, and privacy [2,5,8,9]. It is not surprising that *(perceived) occupational safety* is the primary issue associated with the introduction of robots in workplaces. In addition to ensuring a safe design for cobots to prevent injuries to the working staff, it is imperative that employees feel safe while working alongside robots, and efforts should be made to prevent accidents and collisions [2]. Despite the implementation of safety features in cobots, i.e., the speed and separation monitoring and power and force limiting ISO safety requirements, which are paramount when enabling close HRI, workers may remain unaware of these mechanisms. Therefore, it is important for employees to have the chance to interact with the cobot beforehand, experimenting with it to cultivate trust in its integrated safety functions. *Physical ergonomics* contributes to compromising the fundamental comfort and well-being, decreasing the likelihood of diseases and injuries, and impacting the overall quality of work. Physical burden on workers' bodies, resulting in the so-called musculoskeletal disorders, is measured in the literature through different criteria, including the Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), postural Loading on the Upper Body Assessment (LUBA), and Occupational Repetitive Action (OCRA) [5]. In addition to physical ergonomics, *cognitive ergonomics* hold paramount importance for individuals during the implementation of disruptive technologies like cobots. Specifically, psychosocial factors [9] encompass aspects such as participation, corporate culture, feedback, motivation, stress, teamwork, fairness, job satisfaction, role clarity, over- and under-demand, social influence, and fear of job loss [8]. The *efficiency* factor [2] is defined as the enhancement of the overall industrial process, whether achieved through streamlining the operator's actions in task completion, scheduling activities, or planning optimal actions performed by both the worker and the robot. Hence, certainly, the concept of HRI acceptance is influenced by efficiency and, consequently, by performance expectancy [9] that can be shaped by anticipated productivity gains, enhancements in quality, decreased implementation and setup durations. It is widely acknowledged that the *design*, i.e., customizing the cobot configuration to meet the employees' requirements, is essential. Compact size for the ability to operate even in open layouts without fencing, flexibility of use, i.e., easy programmability, quick adaptation, and relocation, especially for companies producing small batches with a large number of variants, can be some of the motivating factors to invest in a cobot. Obviously, since there is frequent turnover among employees working with cobots, it is sensible to design the workplace in a manner that allows adaptation to the needs and anthropometric data of different individuals [8,9]. *Privacy* issues [9] could impede operators from utilizing robots, especially those related to data protection. These concerns, often intertwined with legal regulations, must be integrated in the lifecycle of the robot. In addition legal concerns, defined as rules of conduct and actions prescribed by a governing body or formally recognized as binding, should be considered.

On the other hand, in the realm of service, four essential acceptance factors are recognized, specifically, safety, design, cognitive and emotional comfort, and privacy [6,7,10]. The concept of personal *safety* [7] holds particular significance in fostering the acceptance and seamless integration of sobots into individuals' lives. It is characterized by users feeling comfortable in readily sharing their personal space with the robot. The presence of feelings of uncertainty may emerge in HRI, especially when interactions occur within users' domestic or medical environments. The robot's physical *design* is a crucial element as it offers the initial chance for communication and interaction with human users. All the design features of robots [6], i.e., perceived usefulness, user-friendliness, facilitating conditions, physical appearance and visual attractiveness, size, and anthropomorphism, that in our case refers to the attribution of a human shape, characteristics, or behavior to robots, must be considered for the social acceptance of robots. *Cognitive and emotional comfort* of operators is of fundamental importance in making robots socially accepted. Important factors to take into account undoubtedly include enjoyment, characterized by the user experiencing feelings of joy or pleasure in association with the robot's use, companionship, which refers to the user's perceived ability to form a relationship with the robot, perceived behavioral control and decreased anxiety toward robots, and social influence [7]. Regardless of gender, age, or culture, individuals commonly view a robot with a lifelike appearance as a friendly companion. When sobots are deployed in homes or hospitals, users place significant weight on the opinions of their family or colleagues. As sobots become more integrated into public spaces, understanding the security implications, particularly in settings like hospitals, where safeguarding patients' and families' privacy, as well as the privacy of medical staff, becomes crucial for robot acceptance. Therefore, akin to the industrial category, *privacy* emerges as the final but equally significant factor [10].

3 Discussion and Conclusions

This research investigates the key factors identified from the literature that expedite or hinder the acceptance process, focusing on people's willingness to use robots in the industrial and service HRI domains. Measuring the acceptance factors and identifying appropriate indicators or metrics is essential for integrating them into the control architecture and design of robots. The current study demonstrates how some factors are already quantified while others are only partially or not quantified at all. In particular, in the manufacturing domain, it is observed that safety, physical ergonomics, and efficiency can be measured, while progress is still needed for proper design. Specific psychosocial aspects of cognitive ergonomics, as well as certain legal and privacy factors, remain unexplored and unquantified. In the service realm, further advancements are required in all four classified factors. Therefore, although the concept of acceptance is a key factor for successful technology adoption, the results of this work show that there are still few studies and, consequently, few well-established tools and well proven strategies in place to measure acceptance. In future works, to enhance robots' acceptance by users, it is necessary to more extensively quantify the above-mentioned key factors, identify and integrate indicators and metrics in the control architectures of robots, e.g., optimizing the trajectory planning and tracking of the robot, as well as collision avoidance and detection, and scheduling of activities between robot and user, and in the design.

Acknowledgements. This work was supported by the AI-DROW Project through the Italian Ministry for University and Research under the PRIN 2022 program, funded by the European Union – Next Generation EU.

References

1. Cusano, N.: Cobot and sobot: for a new ontology of collaborative and social robots. Found. Sci. **28**(4), 1143–1155 (2023)
2. Proia, S., et al.: Control techniques for safe, ergonomic, and efficient human-robot collaboration in the digital industry: a survey. IEEE Trans. Autom. Sci. Eng. **19**(3), 1798–1819 (2021)
3. Ajoudani, A., et al.: Progress and prospects of the human-robot collaboration. Auton. Robots **42**, 957–975 (2018)
4. Villani, V., et al.: Survey on human-robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatron. **55**, 248–266 (2018)
5. Lorenzini, M., et al.: Ergonomic human-robot collaboration in industry: a review. Front. Robot. AI **9**, 262 (2023)
6. Ma De Graaf, M., Allouch, S.B.: Exploring influencing variables for the acceptance of social robots. Robot. Auton. Syst. **61**(12), 1476–1486 (2013)

7. Chatzoglou, P.D., et al.: Factors affecting acceptance of social robots among prospective users. Int. J. Soc. Robot. 1–20 (2023)
8. Kopp, T., et al.: Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. Int. J. Adv. Manuf. **112**, 685–704 (2021)
9. Jacob, F., et al.: Picking with a robot colleague: a systematic literature review and evaluation of technology acceptance in human–robot collaborative warehouses. Comput. Ind. Eng. 109262 (2023)
10. Whelan, S., et al.: Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int. J. Soc. Robot. **10**(5), 643–668 (2018)
11. Savela, N., et al.: Social acceptance of robots in different occupational fields: a systematic literature review. Int. J. Soc. Robot. **10**(4), 493–502 (2018)
12. Hendrick, H.W., Brown, O.: Human Factors in Organizational Design and Management. Elsevier Science Ltd. (1984)

On the Development of Programming by Demonstration Environment for Human-Robot Collaboration in a Furniture Painting Cell

Joan Lario[1](✉), Francisco Fraile[1], Emima Ioana[1], and Francisco Blanes[2]

[1] Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València, Camino de Vera S/N, 46022 València, Spain
jlario@cigip.upv.es

[2] Automatics and Industrial Informatics Research Institute (ai2), Universitat Politècnica de València, Camino de Vera S/N, 46022 València, Spain

Abstract. The advanced manufacturing capabilities such as robot arms, sensors, and software can create a safe, dynamic, and real-time decision-making for a collaborative working environment between humans and robots. The current paper presents a human–robot collaboration system to be applied in a furniture painting cell. The aim is to increase the level of automatization of the painting process and minimize the repetitive operations performed in the conventional manual painting process. The use of autonomous robot arms, programmed by demonstration, will reduce the worker exposure time to chemicals employed in the painting process and minimize the repetitive movements performed during manual painting, which will improve human safety. Current research aims to define a framework where humans and robots can share the working area simultaneously thanks to integrating several systems to control the working environment, the robot arm, and the operation, summarizing current technologies' main challenges to achieve a symbiotic collaboration painting cell.

Keywords: Collaborative robotics · Programming by Demonstration · Industry 4.0 · Smart manufacturing

1 Introduction

In conventional furniture painting processes, the operator performs repetitive motions with the painting gun, which can develop an accumulation of local muscle fatigue in a mid or long time, which may end in work-related musculoskeletal disorders (Lorenzini et al., 2019). The traditional application of robots in manufacturing and assembly processes has been carried out in isolated environments, with physical barriers that prevent the operation if a worker accesses their working area (Pérez et al., 2020). In this conventional robot vision applied in industry, robots are programmed to perform repetitive tasks, such as welding, painting, or handling products. The development of new

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 136–141, 2024.
https://doi.org/10.1007/978-3-031-76428-8_26

sensors, algorithms, and communication protocols for advanced manufacturing systems allows robots to cooperate with operators in safe conditions (Wang et al., 2020). In this collaborative environment, the robot performs the repetitive tasks of spraying the sealant, painting, or glossing coating, and the operator visually inspects the furniture. The Human-Robot Collaboration (HRC) proposed in the current use case aims to create a working environment where humans and robots share a task to accomplish, which is a furniture part coating or painting. The integration of advanced control solutions based on machine vision, LIDAR or laser integrated in the robot cell, allows humans and robots to share a physical space, performing a coordinated or synchronous activity with high accuracy and safety.

2 Case Study Description

Conventional furniture production relies upon a high volume of human labour to perform tasks of low added value, such as handling, coating, polishing or assembly. A hazardous atmosphere can be produced in some specific process (volatile compounds or particles), which requires special personal protective equipment (PPE) or an air exhaust system to control the dust or air flow conditions. Also, the operations are composed of repetitive movements with light or medium-weight tools (industrial paint spray gun) that may generate some kind of occupational disease in the long term. Developing and integrating autonomous collaborative robot cells for coating or painting processes can reduce the operator's workload and improve working conditions and quality of life. Currently in the manual painting cell the workers perform the total spraying task for each part painted. This requires operators with high levels of training, experience, and skills, who are applying repetitive movements that could be automated. The manual operations with low added value put a burden on the European countries that cannot compete with the salaries of the third countries. Also, due to the nature of the painting process, many companies cannot balance the workload of the qualified personnel required to perform the painting process. In most cases, the lack of trained personnel requires an elevated, time-consuming training phase that increases lead time and production costs (Arrais et al., 2021). Additionally, performing repetitive tasks during operators shift reduce the cognitive engagement of the worker. A collaborative human-robot environment will be created by introducing several AI-PRISM solutions, translating this movement to the robot arm, and reducing operator mental fatigue.

3 Proposed Painting Process Based on Human-Robot Collaboration

The furniture manufacturing industry has conventionally required highly skilled operators to perform assembly or repetitive tasks. The development of a new collaborative robotic painting system, which can work side by side with humans, due to their safety features incorporated in their design, allows to improve the level of an automatization of the furniture production lines. The safety features, hardware and algorithms, are one of the main points to consider for adopting collaborative robots cell, which are necessary to assess the operator's body areas exposed to risks so that the speed of the collaborative robot is adjusted accordingly.

3.1 Operator-Robot Relationship

Combining human and robotic systems in complementing competencies can improve ergonomics, working conditions and process performance. Thanks to programming by demonstration technology, the subject matter expert can train the robots to perform the painting task. In this collaborative environment, the operator's main task will be the painting inspection and quality assurance once the painting program has been created and validated to be employed in production. Compared to robotic systems, the main capabilities that humans possess are the skills and ability of perception, processing, reasoning, and decision-making, which will be complex to achieve in automatic robotic painting cells due to constraints related to available technology or the available economic budget. Additionally, the worker can supervise and adjust the parameters of the painting process through the user interface to maximize reliability and performance.

3.2 Conceptual System Architecture

The first technical task the painting robotic system should perform is recognising and locating the furniture part, the human and itself in the painting area. Ambient digitalisation will be achieved by integrating 3D scanners (LIDAR) or RGBD cameras into a Robot Operating System (ROS), which will be able to create a real-time digital model of the collaboration environment. The LIDAR technology will supply the capability to identify objects and obtain the distance map. This capability and agent level reasoning will allow to modify the robot arm speed if a human is nearby. Additionally, to improve safety, a set of RGBD cameras with human recognition will be installed to track operators and avoid collisions or unsafe interactions (Fig. 1). The error detection of furniture reference or position at the beginning of the painting will be carried out by an AI-based perception module, based on the information of real-time digital model.

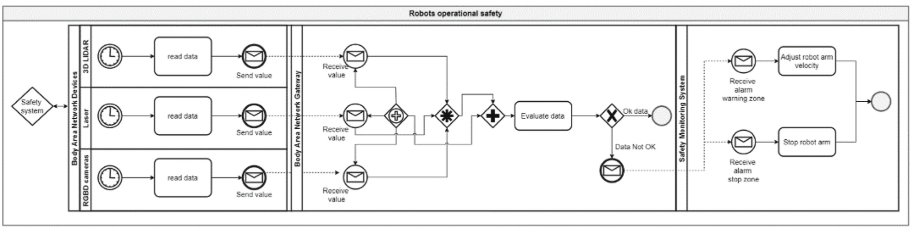


Fig. 1. Business Process Model and Notation (BPMN) for Safety system for Ambient digitalization solution.

To ensure fast and reliable communication between different sensors that conform to the Ambient Digitalization module, a network that employs Software-Defined Networking (SDN) technology will be implemented. The SDN technology aims to reduce the latency of the critical data transmitted, improving the computing solutions' efficiency. The agile and informed decision-making will be supported by a network that emphasises Quality of Service (QoS) to ensure and prioritise transmission of essential data. The proposed communication infrastructure prioritises data flows, ensuring uninterrupted service in a centralised operating environment. The safety functionalities will be directly connected to the ROS and the control system to achieve a human-robot collaborative working environment. These are critical requirements for the robot arm to execute the painting program autonomously.

3.3 Programming by Demonstration

A general tendency in manufacturing is the move towards more flexible automation methodologies that can deal with smaller batch sizes of more customized products. To this end, robotic applications must be rapidly reconfigurable to deal with various tasks while speeding up their deployment on the factory floor. When deploying robotic applications, traditional methods such as manual programming and teleoperation can be time-consuming and require a high level of expertise from the operator. These drawbacks have been addressed by two distinct communities in the robotic field (Pagter, 2016; Schaal, 1999). First, the learning community argues that learning methodologies can facilitate task programming while speeding up deployment (Schaal, 1999). Particularly, in the imitation learning approach, a motion model is generalized based on movement primitives learned from observation of humans performing a task (Pagter, 2016). The automation of painting furniture presents several difficulties since products or parts change almost every year, thus requiring a high volume of programming. Also, many references or parts are to be processed with a different geometry, which requires a specific program for each part.

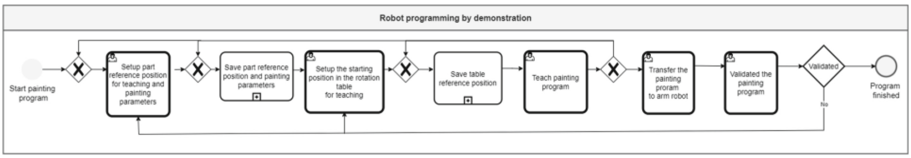


Fig. 2. BPMN for Robot programming by demonstration.

The furniture industrial case proposes that a demonstration solution teaches the robot trajectory with intuitive programming (Fig. 2). The main advantage of this programming technology is that highly trained operators can create painting or finishing programs with complete abstraction of complex programming languages or concepts (Wang et al., 2020). An experienced operator demonstrates how to paint the chair using programming via demonstration hardware: Oculus controller. With the sensing infrastructure, programming via demonstration hardware captures the details of the painting tool and furniture trajectories, movements and positions. The acquired data is synchronized to ensure accurate robot action mapping. Advanced algorithms then process this data to learn the action plan for the robot. The action plan includes the individual arm movements involved in painting and the rotation of the frame when maneuverings the furniture. Once validated, the program is stored in a file system and referenced in a database, allowing it to be later invoked to paint that reference to the automatic painting cell. If the painting program requires adjustment, a skilled operator can manually adjust the learned robotic painting parameters through a user interface tool. For example, the human can make subtle changes on the robot arm trajectory to improve painting results. Thanks to this supervision capability, the collaborative robotics system not only allows the expert worker to ensure program reliability but also promotes continuous improvement of the robot's performance, leading to greater productivity and overall product quality.

4 Conclusion and Future Work

The proposed conceptual frameworks aim to achieve a symbiotic collaboration environment, where humans and robots share a working area simultaneously without physical separation, reducing the cycle time, risks and cost compared with manual painting operations. Under this proposed scenario, an agent-level reasoning, acting and control solution will be developed for collision avoidance to implement reliable and safe working environments. To achieve this purpose, a multimodal communication for sensor integration and control will be addressed for adaptive control for robot execution. In this new HRC scenario, if the operator accesses the safety area to inspect the part after the robot has painted agent-level reasoning, acting, and control from the AI-PRISM solution, it will be triggered to avoid collision between the human and robot arm.

Acknowledgements. AI-PRISM has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101058589.

References

Arrais, R., Costa, C.M., Ribeiro, P., Rocha, L.F., Silva, M., Veiga, G.: On the development of a collaborative robotic system for industrial coating cells. Int. J. Adv. Manuf. Technol. **115**(3), 853–871 (2021). https://doi.org/10.1007/s00170-020-06167-z

Liu, S., Wang, L., Wang, X.V.: Symbiotic human-robot collaboration: Multimodal control using function blocks. Procedia CIRP **93**(March), 1188–1193 (2020). https://doi.org/10.1016/j.procir.2020.03.022

Lorenzini, M., Kim, W., Momi, E. De, Ajoudani, A.: A new overloading fatigue model for ergonomic risk assessment with application to human-robot collaboration. Proceedings - IEEE International Conference on Robotics and Automation, 2019-May, pp. 1962–1968 (2019). https://doi.org/10.1109/ICRA.2019.8794044

Pagter, J. de: Building trust in robots: A narrative approach. In: IEEE Spectrum (2016)

Pérez, L., Rodríguez-Jiménez, S., Rodríguez, N., Usamentiaga, R., García, D.F., Wang, L.: Symbiotic human–robot collaborative approach for increased productivity and enhanced safety in the aerospace manufacturing industry. Int. J. Adv. Manuf. Technol. **106**(3–4), 851–863 (2020). https://doi.org/10.1007/s00170-019-04638-6

Schaal, S.: Is imitation learning the route to humanoid robots? Trends Cogn. Sci. **3**(6), 233–242 (1999). https://doi.org/10.1016/S1364-6613(99)01327-3

Shen, W., Wang, L., Hao, Q.: Agent-based distributed manufacturing process planning and scheduling: A state-of-the-art survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 36(4), 563–577 (2006). https://doi.org/10.1109/TSMCC.2006.874022

Wang, L., Balasubramanian, S., Norrie, D.H., Brennan, R.W.: Agent-based control system for next generation manufacturing. In: Proceedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference Gaithersburg, MD 0 September 14–17 (1998)
Wang, L., Liu, S., Liu, H., Wang, X.V.: Overview of human-robot collaboration in manufacturing. In: Lecture Notes in Mechanical Engineering. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-46212-3_2

Ensuring Trustworthiness of Hybrid AI-Based Robotics Systems

Alexander Eguia[1(✉)], Nuria Quintano[1], Irina Marsh[2], Michel Barreteau[3], Jakub Główka[4], and Agnieszka Sprońska[4]

[1] Tecnalia, Basque Research and Technology Alliance (BRTA), Derio, Spain
{alexander.eguia,nuria.quintano}@tecnalia.com
[2] CBRNE Ltd., London, UK
irina.marsh@cbrneltd.com
[3] Thales Research and Technology, Palaiseau, France
michel.barreteau@thalesgroup.com
[4] Łukasiewicz Research Network - Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland
{jakub.glowka,agnieszka.spronska}@piap.lukasiewicz.gov.pl

Abstract. Hybrid Artificial Intelligence (HAI) algorithms are well adapted to the industrial environment since they require significantly less data than data-driven Artificial Intelligence (AI) algorithms, matching the industry constraints. However, HAI does not fully address the issue of trust (validity, transparency, explainability, and ethics) which must be tackled to achieve world-class HAI beneficial to humans individually, organisationally, and societally. This paper focuses on describing the methods used to collect and elicit trustworthiness requirements in the European ULTIMATE project. It includes requirements register and a trustworthiness glossary of terms.

Keywords: hybrid AI · robotics · trustworthiness · ethics

1 Introduction

AI technologies are anticipated to bring a broad range of economic and societal benefits across multiple sectors.

The integration of AI systems into products and services has elicited growing concerns regarding their potential impact on fundamental rights and safety risks posed to users. Notably, apprehensions have been raised regarding the potential infringement on key rights such as non-discrimination, freedom of expression, human dignity, protection of personal data, and privacy [1].

The challenge to bring these values and principles to the AI based systems engineering process is addressed by the European project ULTIMATE [2].

The principle of Hybrid Artificial Intelligence (HAI) is the amalgamation of data-driven AI algorithms, finding patterns in data, and model-driven AI algorithms, relying on physical models and constraints [3]. This fusion of methods

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 142–146, 2024.
https://doi.org/10.1007/978-3-031-76428-8_27

gives the correct context to the data, essentially improving the behaviour of AI algorithms [4]. Physical models and constraints enable HAI to provide better results with limited training data. Because of this, HAI algorithms are well suited for the industrial environment. Also most problems in industry are too complex to be accurately solved by a model-driven approach [5].

Every critical AI system requires a complete tool chain ensuring trust at all stages, and this remains true under this hybrid approach [6].

The ULTIMATE project is investigating novel learning approaches to design HAI-based algorithms with increased explainability and interpretability. Then, these HAI-based solutions will be evaluated using statistical, experimental, formal methods, and tools to consolidate their trustworthiness. The project will assess the behaviour of those HAI-based solutions under operational conditions of the selected use cases. In this paper we concentrate on the use case that is exploring collaborative robotics that handle parts in a workshop shared with humans. In the workstations, the manipulators pick components from the containers and place them on the assembly table. Human workers assemble the components and place them back in the containers, which are then transported by the mobile robot to the warehouse.

The main objective for this use case is to achieve reliability and efficiency of the robotic support. The HAI functionalities are expected to provide the following improvements for the manufacturing process:

– Increased safety for humans by detection and precise tracking of humans allowing to adjust the robots' movements;
– Better mutual understanding of human and robotic actions for example during the handover of parts;
– Better overall efficiency of the workshop.

This article presents how we approached the identification of trustworthiness requirements. To elicit trustworthiness requirements for the use case we used the Value Sensitive Design (VSD) principles. This working approach is explained in the next section "Materials and Methods".

2 Materials and Methods

VSD approach used for identifying and prioritising trustworthiness requirements was initially introduced to integrate public values into human-computer interactions. Then the methodology was expanded to encompass the inclusion of values in various domains of technological design [7].

We have created a methodology based on the VSD principles for Ultimate project. For AI trustworthiness, we have considered two main international guidelines: (1) the *Ethics Guidelines for Trustworthy AI* proposed by the EU Independent High-Level Expert Group on Artificial Intelligence (AI-HLEG) and (2) the *Ethically Aligned Design* (EAD) proposed by the Institute of Electrical and Electronic Engineers (IEEE) Standard Association.

In 2019, The AI-HLEG group identified seven key values an AI system should fulfil to be considered "trustworthy" [9]. The General Principles of EAD, published in 2018, articulates nine high-level ethical principles which apply to all types of autonomous and intelligent systems [10], which are similar in scope with the ethical and social values proposed by the AI-HLEG.

Those two references helped to create the *AI Trustworthiness User Requirement Template* to support the elicitation and prioritisation for the trustworthiness requirements for the use case. An extract of the template is shown on Table 1 for the purpose of illustration, and the full version is available as Supplementary Material S1.

Table 1. Extract of the AI Trustworthiness User Requirements Template.

VSD ethical and social values	Component	Hybrid AI Trustworthiness General Requirements Priority
Human autonomy	Human agency	Is the hybrid AI implemented in the work and labour process? Consider task allocation between the hybrid AI and humans for appropriate human oversight and control. Does the Hybrid AI enhance or augment human capabilities? Consider safeguards to prevent overconfidence/over-reliance in the results offered by the hybrid AI.
	Human oversights	What is the level of human control or involvement? Who is the human in control and what are the actions or tools for intervention? Consider mechanisms to ensure human control and oversight.
Technical robustness and safety	Resilience to attack and security	Consider to put measures in place to ensure the integrity and resilience of hybrid AI against potential attacks. Consider how you system behave in unexpected situations and environments and what would be needed to mitigate a potential negative outcome.

In the first column "VSD ethical and social values" we have listed the nine values and principles: the seven ethical values proposed by the AI-HLEG supplemented by the two IEEE EAD principles, *Awareness of misuse* and *Competence*, which do not have a direct correspondence in the AI-HLEG. In the second column "Component" the building block substantiating the values and the principles are detailed, and are used as a benchmark for grouping in the third column "Hybrid AI Trustworthiness General Requirements" as detailed by *Assessment List for Trustworthy Artificial Intelligence* (ALTAI) document [11]. The last column "Use Case Specifications and Priority Category (Priority)" supports the further specification of the trustworthiness requirements for the use case and allows for prioritisation. This information is further analysed by HAI developers and evaluators to determine which trustworthiness requirements to work on.

3 Results

Using established methodology system trustworthiness requirements register has been created: 27 of the 52 trustworthiness characteristics have been marked as high priority, 9 have been marked as medium, while only 4 dimensions were classified as low priority. The remaining 12 characteristics did not apply due to the characteristics of the use case.

3.1 Trustworthiness Characteristics Mind Map

We created a mind map (available as Supplementary Material S2) that considers the majority of AI criteria to be assessed for trustworthiness purpose. We labelled the terms identified as requirements or European values to be respected by the referenced documents. The technical-related group stands above the rest with great emphasis on quality-based criteria, however it is the category focused on ethical issues that concentrates most of the attention.

3.2 Glossary of Terms

A glossary of terms has been compiled (available as Supplementary Material S3) to define all the trustworthiness concepts as each reference provided a slightly different set of definitions. Information, including ISO/IEC standards, IBM's dedicated programs for trustworthy AI [12], and contributions from the National Institute of Standards and Technology (NIST) [13], along with academic publications, supplements the resources from AI-HLEG and IEEE.

4 Discussion and Supporting Materials

The robotic system trustworthiness requirements register has started to be analysed to identify main trustworthiness goals for the HAI-based solution in the context of the robotic system. An initial analysis establishes two goals at system level: safety and efficiency. Safety in terms of reducing the probability of expected harms and the possibility of unexpected harms which may derive from the robot-human interaction. Efficiency in terms of the elapsed time the assembly process in which the robot participates is required to complete.

Next step of the work is to identify quantitative indicators which are meaningful for representing the measurement of variables.

Principal conclusions obtained from this preliminary work are: (a) trustworthiness is a broad concept and needs to be analysed in detail; (b) achieving a common understanding among different stakeholders needs to be encouraged since it is not easily obtained; (c) AI trustworthiness evaluation requires understanding the business context and priorities.

The supporting information can be downloaded at: https://zenodo.org/communities/ultimate-horizoneu/ including the full VSD template (Supplementary Material S1), detailed trustworthiness characteristics mind map (Supplementary Material S2) and complete glossary of terms (Supplementary Material S3).

Acknowledgements. This research was funded by European Commission, grant number 101070162 (ULTIMATE).

References

1. Artificial intelligence act Briefing. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf. Accessed 28 Sept 2023
2. ULTIMATE: MUlti-Level Trustworthiness To IMprove The Adoption Of Hybrid ArTificial Intelligence. https://ultimate-project.eu/. Accessed 28 Sept 2023
3. Garcez, A.D.A., Gori, M., Lamb, L.C., Serafini, L., Spranger, M., Tran, S.N.: Neural-symbolic computing: an effective methodology for principled integration of machine learning and reasoning. arXiv preprint arXiv:1905.06088 (2019)
4. Ostheimer, J., Chowdhury, S., Iqba, S.: An alliance of humans and machines for machine learning: hybrid intelligent systems and their design principles. Technol. Soc. **66**, 101647 (2021)
5. Bruneliere, H., et al.: Model-driven engineering for design-runtime interaction in complex systems: scientific challenges and roadmap. In: Software Technologies: Applications and Foundations, pp. 536–543. Springer, Cham (2018)
6. Mattioli, J., Pedroza, G., Khalfaoui, S., Lero, B.: Combining data-driven and knowledge-based AI paradigms for engineering AI-based safety-critical systems. In: Workshop on Artificial Intelligence Safety (SafeAI) (2022)
7. Taebi, B., Correljé, A., Cuppen, E., Dignum, M., Pesch, U.: Responsible innovation as an endorsement of public values: the need for interdisciplinary research. J. Responsible Innov. **1**, 118–124 (2014)
8. Van de Poel, I.: Investigating ethical issues in engineering design. Sci. Eng. Ethics **2001**(7), 429–446 (2001)
9. Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 28 Sept 2023
10. Ethically Aligned Design. A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v1.pdf. Accessed 28 Sept 2023
11. Assessment List for Trustworthy Artificial Intelligence (ALTAI). https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal. Accessed 28 Sept 2023
12. Trustworthy AI | IBM Research. https://research.ibm.com/topics/trustworthy-ai. Accessed 28 Sept 2023
13. NIST Trustworthy & Responsible Artificial Intelligence Resource Center. https://airc.nist.gov/Home. Accessed 28 Sept 2023

AI-Based Positioning on a Micro Scale

Tomasz Kołcon[1](✉), Iveta Eimontaite[2], Piotr Gemza[1], Krystian Goławski[3], Miron Kołodziejczyk[1], and Adam Wołoszczuk[1]

[1] Łukasiewicz Research Network – Industrial Research Institute for Automation and Measurements PIAP, Al. Jerozolimskie 202, 02-486 Warsaw, Poland
{tomasz.kolcon,piotr.gemza,miron.kolodziejczyk,
adam.woloszczuk}@piap.lukasiewicz.gov.pl
[2] Cranfield University, Cranfield MK43 0AL, UK
iveta.eimontaite@cranfield.ac.uk
[3] VIGO Photonics, Poznańska Street 129/133, 05-850 Ożarów Mazowiecki, Poland
kgolawski@vigo.com.pl

Abstract. The presentation aims to show an example of the use of AI in the positioning of semiconductor elements with accuracy at the micrometer level. The stand will be used in the production process of infrared detectors. It was shown how AI will support humans in decision-making and improve the production process, which was previously fully manual. The impact on the individuals in terms of physical and psychological wellbeing are also described.

Keywords: Artificial Intelligence · Microelectronics · Collaborative robotics

1 Introduction

VIGO Photonics is a European manufacturer of semiconducting materials and instruments for photonic and microelectronic, specialized in MWIR and LWIR detectors and modules. During many years of research and development, VIGO evaluate own manufacturing technology to improve detectors' parameters almost 10 times. Thanks to improvement VIGO provides unique product in the world. The manufacturing process is largely based on manual activities and the operator's experience. VIGO participate in AI-PRISM project to develop innovative solution to ensure the expected production volumes and take care of workforce wellbeing. In the case of this use-case for human robot collaboration, a classic robotic arm is replaced by a precise XY stage and equipped with a microscope camera. The use of such a set is necessary due to the size of semiconductor structures, which usually have dimensions of several dozen micrometers (Fig. 1).

2 Current State

2.1 Manual Gluing Stick to the Chip

This process consists of several stages: station calibration, inserting the chip, finding the center of the structure, stick insertion and gluing the stick to the chip (Fig. 1).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 147–151, 2024.
https://doi.org/10.1007/978-3-031-76428-8_28

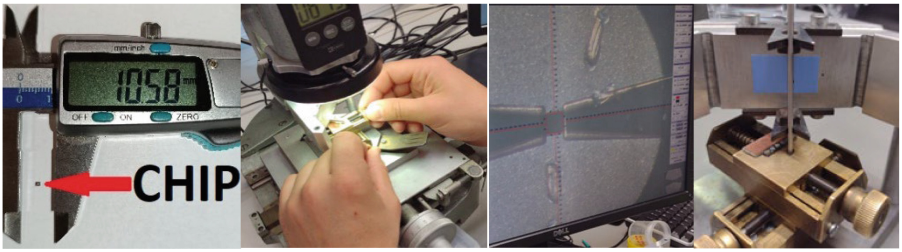


Fig. 1. Positioning the chip to the stick, stick insertion and gluing to the chip.

The most challenging and time-consuming moment is finding the center of the chip. The process is done manually by the operator by turning knobs on a mechanical XY stage (Fig. 1) (Fig. 2).

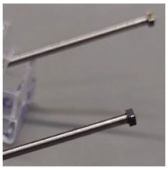


Fig. 2. Assembly prepared for next manufacturing steps.

2.2 Human Factor

The current procedure is currently managed manually by operators, impacting both their physical and psychological well-being. Firstly, the increased mental workload and concentration required for positioning the center of the chip result in heightened stress. This aspect of the process, in addition to inducing eye strain and demanding precise finger motor skills, leads to physical fatigue that accumulates over the course of a shift. Furthermore, the fast turnaround and precision level needed for task completion contribute to increased mental strain.

While operators develop their own routines, such as stacking component steps, to enhance their sense of independence and comfort, these individual adjustments do not fully mitigate the overall impact of the work on their well-being. Additionally, the reliance on short production runs and the dependence on the tacit knowledge operators acquire over time [1], collectively make the process challenging to automate.

2.3 Identified Issues

During the analysis of this use-case, the following problems were identified: high degree of production rejects, long procedure time, requires long-term concentration by the operator.

3 AI-Based Human-Robot Collaboration Solution

3.1 Human-Robot Collaboration

Human-robot collaboration, in this case, involves relieving the operator of the most stressful and time-consuming elements of the entire process. The full automation of the cycle in this scenario lacks economic or practical justification. Human-robot collaboration proves most effective by leveraging the strengths of both partners: human higher cognition and flexibility, and robot precision and repeatability [3]. Based on an observations of operators' behavior and interviews about the challenges from their perspective [1], the most challenging task for the operator is the positioning of the stick relative to the center of the chip, and this aspect of the process will be automated. The operator will continue to supervise this activity and, in case of issues, will either approve or reject the solution proposed by the AI. Additionally, in the production process, recognizing defects is crucial, and in this case, AI will assist in identifying potential imperfections, but the final decision will rest with the human. Retaining decision-making power with the operator has been shown to enhance technology acceptance and engagement [4, 5]. The proposed process is indicative of cognitive as opposed to physical collaboration between human and robot. The technology and process still depend on human cognitive flexibility and expertise. This emphasizes the removal of strenuous and physically challenging aspects from human to technology, while the human performs teaching technology of the new assembly components and final quality control. In the proposed human-robot collaboration within the current process, it is expected to improve the physical well-being of the operator while preserving self-efficacy and independence.

3.2 Mechanical Concept

The main idea of building the new station was to use the experience from the current manual version and expand it with elements enabling automatic positioning, including a motorized XY table. A new microscope camera has also been developed, which enables the visibility of micrometer-sized details (Fig. 3).

3.3 AI-Driven Positioning

Due to short production runs and not always precisely defined features that enable finding the center of the chip, AI was used (Fig. 4).

Key points are marked in the first few images annotated on training sample and the AI is trained to recognize detect them and calculate the center of the chip. An additional option is the ability to automatically recognize structure defects. Thanks to real-time tracking of the chip center using AI and motorized XY tables, positioning is performed automatically.

The key point detection method (often called pose estimation because a popular use-case is human keypoints detection to estimate pose) has been chosen to detect predefined points on chip. These are first of all corners of ROI, but also other helper points, e.g. good candidates are distinctive points at the border of chip. Those additional points help to estimate pose of the chip and thus the position of ROI. Selection of additional keypoints

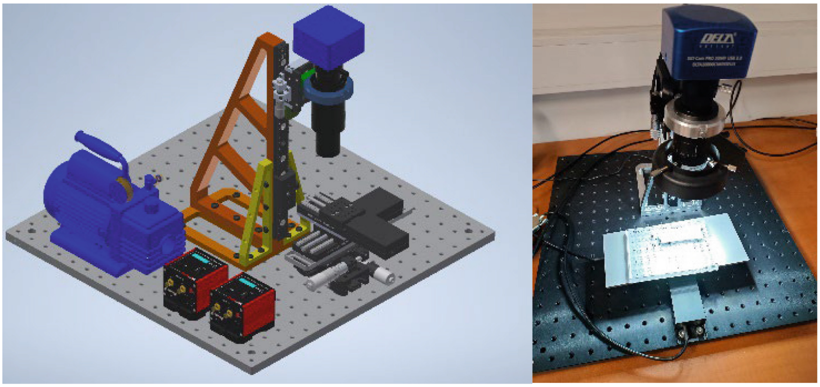


Fig. 3. Mechanical concept and prototype.

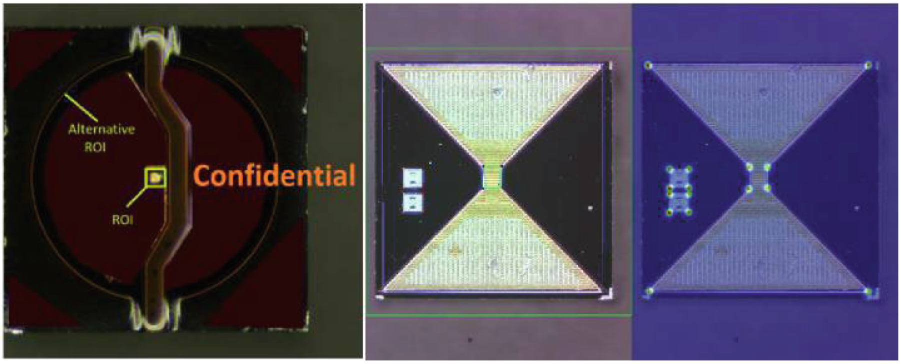


Fig. 4. Picture of the chip from microscope camera (confidential area is hidden) and AI-based ROI (Region Of Interest) detection.

is always based on expert judgement, but their number should be between 4 and 8, so that they do not dominate ROI points too much (the required annotation effort also matters here). Having defined keypoints one has to prepare ground truth template of their positions. Adopted scale and origin do not matter, only relative distances between key points should be preserved. For example, the coordinates could be on real metric scale with the origin at the center of ROI. MMP*ose* [6] from *openmmlab* is a repository with many state-of-the-art implementations available in modular fashion. In the proposed work well established heatmap-base [8] is used.

Assuming one has a template of ground truth keypoints on reference chip $p_1, \ldots, p_n$, and there is a special point p_c on the template - a center point - which is not explicitly detected by the detection algorithm then the detection algorithm returns:

$$\hat{p}_{i_1}, \ldots, \hat{p}_{i_k}, k \leq n.$$

If at least 3 keypoints ($k \geq 3$) are detected and whose keypoint score is greater than some assumed threshold, this allows to estimate affine transformation (using OpenCV's *estimateAffinePartial2D* [7]) between an image and the template.

$$M = \begin{bmatrix} \cos(\theta) \cdot s & -\sin(\theta) \cdot s & t_x \\ \sin(\theta) \cdot s & \cos(\theta) \cdot s & t_y \end{bmatrix}$$

In particular $\hat{p}_c = M \cdot p_c$ is estimated center point in the image coordinates.

Then knowing the distance from the camera to chip and having the model of the optics, the calculation of the center position is possible in a real scale.

4 Conclusions

The presented use case shows that human robot collaboration does not have to be only on a large scale using classic robotic arms. By replacing the robot with precise elements enabling positioning and adding a perception module as a microscopic camera, it is also possible to operate on a micro scale. Adding the tools created in the AI-PRISM project, we get a powerful solution.

Acknowledgment. The work has been conducted on AI-PRISM project (HORIZON-CL4-2021-TWIN-TRANSITION-01 - 101058589) [2] and the authors would like to thank all consortium members for collaboration.

References

1. Eimontaite, I., Fletcher, S., Goławski, K., Kołcon, T.: Breaking the barriers: multilingual user engagement to increase process engagement and technology acceptance in manufacturing (2023). https://doi.org/10.54941/ahfe1003503
2. AI-PRISM Homepage: https:// https://aiprism.eu. Last accessed 18 December 2023
3. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics **55**, 248–266 (2018). https://doi.org/10.1016/j.mechatronics.2018.02.009
4. Bröhl, C., Nelles, J., Brandl, C., Mertens, A., Nitsch, V.: Human-Robot Collaboration Acceptance Model: Development and Comparison for Germany, Japan, China and the USA. Int. J. Soc. Robot. **11**(5), 709–726 (2019). https://doi.org/10.1007/s12369-019-00593-0
5. Hancock, P.A., et al.: A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. Human Factors: The Journal of the Human Factors and Ergonomics Society **53**(5), Article 5 (2011). https://doi.org/10.1177/0018720811417254
6. MMPose homepage: https://github.com/open-mmlab/mmpose. Last accessed 18 December 2023
7. OpenCV documentation webpage: https://docs.opencv.org/4.x/d9/d0c/group__calib3d.htm. Last accessed 18 December 2023
8. Wang, J., et al.: Deep high-resolution representation learning for visual recognition. In: IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(10), 33493364 (2021). https://doi.org/10.1109/TPAMI.2020.2983686

Human Motion Prediction Metrics: From Time to Frequency

Michael Vanuzzo(✉), Marco Casarin, Mattia Guidolin, Stefano Michieletto, and Monica Reggiani

Department of Management and Engineering (DTG), University of Padova, Stradella S. Nicola, 3, 36100 Vicenza, Italy
{michael.vanuzzo,marco.casarin.4}@phd.unipd.it,
{mattia.guidolin,stefano.michieletto,monica.reggiani}@unipd.it

Abstract. Collaborative robotics has the potential to revolutionize industrial applications by integrating human and robot capabilities. However, for efficient and seamless collaboration, predicting human motion is essential. This allows robots to dynamically adjust their behavior and avoid potential collisions. Despite significant progress in this field in recent years, there is still uncertainty surrounding the metrics needed for a complete and accurate evaluation of algorithm performance. Currently, the evaluation of Human Motion Prediction (HMP) techniques is based on metrics focusing exclusively on geometric aspects. This work proposes a HMP metric to evaluate the realism and naturalness of predicted human motion sequences based on their frequency spectra. Using the Human 3.6M dataset, several experiments were conducted to demonstrate the effectiveness of the proposed metric. The results showed the ability of this metric to capture insights related to the realism of the predicted motion sequences, making it a valuable complementary tool alongside existing metrics for evaluating HMP algorithms.

Keywords: Human Motion Prediction · Metrics · Collaborative Robotics

1 Introduction

Collaborative robotics focuses on combining the precision and repeatability of robots with the adaptability and problem-solving skills of humans. This is particularly effective in industrial settings, where robots can provide valuable support to human operators in dangerous and physically demanding tasks. To achieve seamless Human-Robot Collaboration, these systems need to accurately anticipate human movements and dynamically self-adapt based on the operator's behavior. Several Human Motion Prediction (HMP) algorithms have been proposed in the literature based on modeling the human body through a skeletal representation [4–6,8].

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 152–156, 2024.
https://doi.org/10.1007/978-3-031-76428-8_29

The current metrics used for evaluating human movements only analyze geometric aspects of the predicted movements. However, these metrics fail to consider the realism of the movements being predicted as they focus only on the rotation angles or 3D position of the body skeleton. It has been observed [2] that the frequency spectra of motion sequences are strongly correlated with the realism of human movements.

In this work, we propose a novel metric for HMP based on the analysis of the motion frequency spectra of the predicted human movements. The effectiveness of the proposed metric has been addressed through several experiments conducted using the Human 3.6M H36M dataset [3], comparing three state-of-the-art HMP models [4,5,8], as well as the Zero-Velocity (ZeroVel) baseline [6].

2 Methodology

This section describes the metrics commonly used for HMP and proposes a novel metric based on the frequency spectrum of the motion sequences. In the context of HMP, a sequence is a time series of human poses, each defined by a set of features that fully describe the skeleton's configuration.

2.1 Geometric Accuracy Metrics

These metrics aim to measure the difference between each predicted frame and the ground truth within the K sequences of the test set, each spanning T frames.

Mean Angle Error (MAE). The MAE, also known as Euler Error, represents the standard metric to evaluate HMP algorithms [1,4–6], and its definition is:

$$\mathrm{MAE} = \frac{1}{K \cdot T} \sum_{k=1}^{K} \sum_{t=1}^{T} \|\hat{x}_{k,t} - x_{k,t}\|_2 \tag{1}$$

Here, $\hat{x}_{k,t}$ and $x_{k,t}$ denote the predicted pose and the ground truth, respectively, for frame t in sequence k. Each pose $x_{k,t}$ is represented by a vector containing $3 \cdot J$ elements, corresponding to the 3 Euler angles that describe the relative rotation of each of the J joints with respect to its parent joint.

Mean Per Joint Position Error (MPJPE). The MPJPE is a widely used metric for both pose estimation and prediction [5,7]. It inherently considers both the distance between joints, i.e., the length of the links defined in the skeleton, and the accumulation of errors along the kinematic chain. This metric is mathematically defined as follows:

$$\mathrm{MPJPE} = \frac{1}{K \cdot T \cdot J} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{j=1}^{J} \|\hat{p}_{k,t,j} - p_{k,t,j}\|_2 \tag{2}$$

Here, $\hat{p}_{k,t,j}$ and $p_{k,t,j}$ represent the predicted 3D position of joint j and its corresponding ground truth for frame t in sequence k.

2.2 Frequency Spectrum Similarity Metric

When evaluating a HMP algorithm, it is crucial to consider the realism of the generated sequences. However, a prediction yielding high accuracy in geometric metrics may fail to account for unnatural movements that present sharp discontinuities. Therefore, this study focuses on evaluating the realism of the generated motion sequences based on the analysis of their frequency spectrum.

Given a motion sequence $p_k(t)$ described by joint positions over time, its power spectral density $P_{k,norm}(f)$, normalised across all joints, can be computed using the Fourier transform. Then, different power spectral densities can be compared using the Wasserstein Distance (WD), a distance function that can be used between probability distributions. The proposed Power Spectral Densities Similarity (PSDS) metric, defined as the WD of order 1, can be computed as follows:

$$\mathrm{PSDS}\left(P_{k_1,norm}(f), P_{k_2,norm}(f)\right) = \int \left|F_{k_1}(x) - F_{k_2}(x)\right| \mathrm{d}x \qquad (3)$$

Here, $F_{k_i}(x)$ is the cumulative distribution function of $P_{k,norm}(f)$, similar to a probability distribution describing the probability that human joint movements will excite specific frequencies. A higher value is achieved when the power distribution shifts towards either low or high frequencies, a behavior in contrast to typical human movement.

3 Experiments and Results

To assess the effectiveness of the proposed PSDS metric, we conducted multiple experiments evaluating different HMP algorithms on the metrics described in Sect. 2. The study was based on the H36M dataset [3] and on three state-of-the-art Deep Learning (DL) models: History Repeats Itself (HRI) [5], Dynamic Multiscale Graph Neural Networks (DMGNN) [4], and Position-Velocity Recurrent Encoder-Decoder (PVRED) [8]. HRI employs an attention mechanism to identify similarities in past and current action sequences. DMGNN implements Graph Convolutional Network to discern relationships among skeleton joints at various abstraction levels. PVRED is based on a Recurrent Neural Network with Gated Recurrent Units to capture temporal relationships by considering both positional and velocity information. Additionally, the results include scores from the ZeroVel model proposed in [6], in which all the prediction frames are identical to the last input frame. Despite its simplicity, this model is commonly used as a valuable baseline. Notably, many algorithms performed worse than this model [6]. This highlights the challenges of accurately predicting future human poses, but also emphasizes the limits of current evaluation metrics.

Figure 1 presents the results of the three algorithms in terms of MAE and MPJPE metrics. All models perform better than the ZeroVel baseline, with the HRI model showing superior accuracy. However, it is important to note that the improvement relative to the ZeroVel model is not very pronounced.

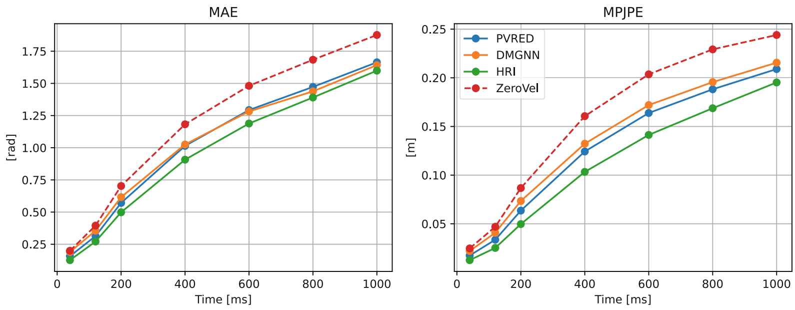


Fig. 1. MAE and MPJPE values during the first second of anticipation for the three models and the baseline, represented by the ZeroVel model.

Alongside the previous results, the outcomes obtained with the proposed novel metric are presented. Each point in Fig. 2 shows the PSDS metric computed on a 1 s sliding window, thus achieving a frequency resolution of 1 Hz. Values are reported in logarithmic scale over a 30 s span.

The results show how the three models outperform the ZeroVel model, highlighting a greater capability of generating natural movements. Furthermore, it is observable that the HRI model, despite providing the best accuracy with MAE and MPJPE, turns out to be the least effective in generating movements with frequencies that resemble natural human motion. The ZeroVel model, generating a constant pose throughout the prediction timeframe, leads to high PSDS values as the only spectral component generated is at 0 Hz. Therefore, this result highlights that ZeroVel predictions are highly implausible. Furthermore, the effectiveness of the PSDS metric is confirmed by the *train-test* value, which is computed using motion sequences from the test set. Given that the latter consists of real human motion recordings, the *train-test* represents the lowest achievable error.

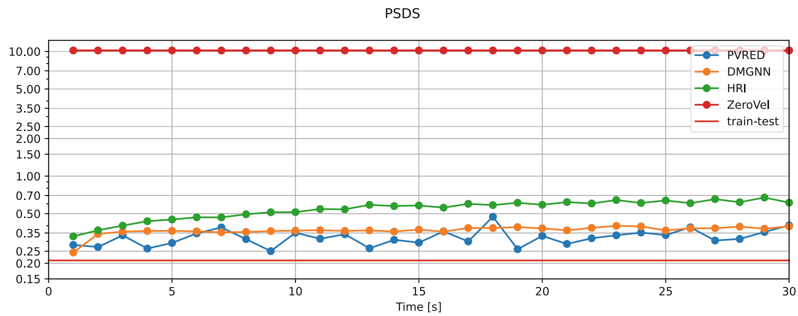


Fig. 2. PSDS values during 30 s of anticipation for the three models, the baseline (ZeroVel model), and the lowest achievable error (train-test).

4 Conclusions

This paper presents PSDS, a novel metric that emphasizes the significance of frequency analysis in determining the quality of generated movement sequences. While commonly used metrics focus on assessing geometric accuracy, PSDS uniquely evaluates the realism of predicted movements. Consequently, it introduces key information complementary to the existing metrics, providing essential insights for developing innovative prediction algorithms. This is particularly crucial in Human-Robot Collaboration field, where predictions must be geometrically accurate and also ensure realism and naturalness. By introducing a novel metric that evaluates these aspects, this paper contributes to a more comprehensive evaluation framework, paving the way for the development of enhanced predictive models.

Acknowledgement. This study was funded by Next-GenerationEU (Italian PNRR - M4 C2, Invest 1.3 - D.D. 1551.11-10-2022, PE00000004) within the MICS (Made in Italy - Circular and Sustainable) Extended Partnership, CUP: C93C22005280001 and by the PRESENCE (anticiPatoRy bEhaviors for Safe and Effective humaN-robot CoopEration) project (BIRD221598).

References

1. Aksan, E., Kaufmann, M., Cao, P., Hilliges, O.: A spatio-temporal transformer for 3D human motion prediction. In: 2021 International Conference on 3D Vision (3DV), pp. 565–574. IEEE (2021)
2. Gopalakrishnan, A., Mali, A., Kifer, D., Giles, L., Ororbia, A.G.: A neural temporal model for human motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12116–12125 (2019)
3. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6m: large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE Trans. Pattern Anal. Mach. Intell. **36**(7), 1325–1339 (2014)
4. Li, M., Chen, S., Zhao, Y., Zhang, Y., Wang, Y., Tian, Q.: Dynamic multiscale graph neural networks for 3D skeleton based human motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 214–223 (2020)
5. Mao, W., Liu, M., Salzmann, M.: History repeats itself: Human motion prediction via motion attention. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision - ECCV 2020, pp. 474–489. Springer, Cham (2020)
6. Martinez, J., Black, M.J., Romero, J.: On human motion prediction using recurrent neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2891–2900 (2017)
7. Véges, M., Lőrincz, A.: Absolute human pose estimation with depth prediction network. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2019)
8. Wang, H., Dong, J., Cheng, B., Feng, J.: PVRED: a position-velocity recurrent encoder-decoder for human motion prediction. IEEE Trans. Image Process. **30**, 6096–6106 (2021)

Towards More Effective Human-Robot Collaboration via Accurate Pose Estimation

Miguel Á. Mateo-Casalí[1], Laura Moya-Ruiz[1], Andrea Caraffa[2], Davide Boscaini[2], Amir Hamza[2], Paul Chippendale[2], Fabio Poiesi[2], and Francisco Fraile[1](✉)

[1] Universitat Politècnica de València, Valencia, Spain
ffraile@cigip.upv.es

[2] Technologies of Vision, Fondazione Bruno Kessler, Trento, Italy

Abstract. This paper explores the pivotal role of accurate pose estimation in enhancing Human-Robot Collaboration (HRC) effectiveness. It delves into the integration of novel AI-based zero-shot estimation algorithms within collaborative robotic platforms and discusses candidate algorithms suitable for industrial scenarios.

Keywords: Object 6D Pose Estimation · Human Pose Estimation · Human-Robot Collaboration

1 Introduction

Human-Robot Collaboration (HRC) is a prominent feature of modern Industry 5.0 applications. HRC applications maintain the involvement of humans in workplaces, leveraging robotics to alleviate mental and physical stress, which results in a high socio-economic impact [2]. Accurate detection and pose estimation plays a crucial role in HRC, as it makes robots aware of the collaboration context. Object 6D pose estimation (estimation of the six degree-of-freedom position of an object in a three dimensional space) algorithms use computer vision to provide a reliable estimation of the pose of any object in space, allowing robotic agents to perceive the status of the collaborative environment, including the contributions of human counterparts to the overall system's response. This perception capability is instrumental to articulate an appropriate reactive behaviour to complement and improve human actions [1].

As opposed to traditional models for object detection, zero-shot object 6D pose estimation models can detect objects in the collaboration environment without prior training and therefore provide higher levels of agility and flexibility to adapt to highly dynamic manufacturing environments, dealing with a large variety of products and tools. However, the integration of cutting edge zero-shot object 6D pose estimation models into HRC applications presents several challenges and aspects like sensing, model deployment, or seamless integration with

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 157–161, 2024.
https://doi.org/10.1007/978-3-031-76428-8_30

robot control, that need to be considered. This paper presents a comprehensive description of AI-based zero-shot pose estimation for 6D object detection, using an example application to describe how AI models can be deployed and integrated with robotic platforms. With this objective, next section describes the state-of-the-art in the field and presents potential candidate solutions. Then, the paper uses a simple pick and place application to illustrate how these models can be integrated into modern robotic platforms and presents a qualitative benchmark of different state-of-the-art models for pose estimation.

2 AI-Based Zero-Shot Accurate Pose Estimation for HRC

Pose estimation aims to determine the position and orientation of a 3D shape within a 3D scene. Different approaches have been developed for scenarios involving a 3D shape representing an *object* (rigid shape) or a *person* (non-rigid shape). Object pose estimation involves finding the rotation and translation of the object within a 3D environment, and typically requires as input a 3D representation of the object (e.g., a CAD model) [4,8,11,15,16]. On the other hand, human pose estimation focuses on identifying specific keypoints like hands, arms, legs, and head on a person [5,10,13].

Both objects and human pose estimation are crucial in HRC because of the coexistence of objects (e.g. tools, manufacturing items) and human workers in collaborative environments [7]. Object pose estimation is essential for enabling robots to interact efficiently with nearby objects, enabling tasks like manipulation, grasping, and assembly [12]. On the other hand, human pose estimation aids robots to assist and avoid collisions with human workers cooperating with them in shared workspaces [3].

Traditional deep learning methods used for pose estimation typically operate within a supervised setting and necessitate strong prior knowledge as input. In this supervised learning paradigm, these models learn from labeled training data, where inputs (such as images or sensor data) are paired with corresponding labeled outputs (poses or keypoints). The model learns to map inputs to outputs by minimizing the difference between predicted and actual poses. Moreover, these approaches often rely on strong prior information about the structure of the human body in human pose estimation (easy to define) or precise 3D representations of objects in object pose estimation (more challenging to obtain). These priors constrain the learning process and guide the model toward more accurate predictions.

However, these requirements restrict the practical applicability of pose estimation methods in industrial scenarios. Acquiring large annotated datasets is arduous due to the absence of suitable 3D annotation tools. Additionally, the excessively time-consuming annotation process is not scalable for the vast amount of data necessary to achieve satisfactory results. Moreover, supervised approaches tend to generalize poorly to data that are different from the ones used to train the model. On the other hand, accurate 3D representations of specific objects, like detailed CAD models, might not be available in some cases.

Current research addresses these limitations by introducing methods designed for zero-shot learning scenarios or by mitigating the necessity for strong priors [6, 9]. The zero-shot learning paradigm aims to improve the generalization ability of the model and make it more suitable for real case scenarios, where specific training data not always are available. This approach reduces its reliance on extensive training data.

3 Example Application

The deployment of the different AI models to leverage available computational resources is one of the technical challenges of the integration of AI models in collaborative robotics platforms [14]. For HRC applications, managing deployed AI workloads involves meticulous oversight of model life cycle, configuration and policy reinforcement within a distributed architecture [17]. Let us use a simple HRC application to illustrate and motivate the main architectural design considerations involved in the implementation and deployment of AI solutions in HRC applications. In this sample application, 6D pose estimation is used to detect both the position of a product instance and the human counterpart. This information is then used to guide the actions of a robot arm that performs a pick and place operation (Fig. 1).

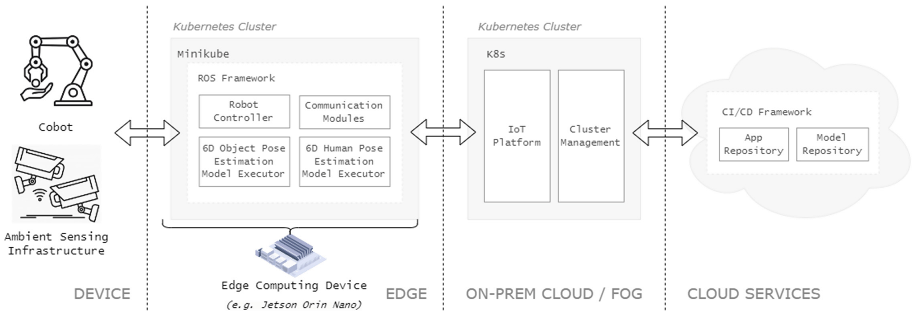


Fig. 1. Proposed architecture for a Collaborative Pick and Place app.

The device level involves all physical elements installed in the plant interacting directly with the environment. These include the cameras that provide the source images to perform 6D object and human detection and the cobot itself. While there may be a certain processing capacity at this level, it is limited, and it is usually unfeasible to run the AI detection models with its resources.

Hence, it is necessary to expand its computing capabilities using added resources while maintaining minimal response times and meeting real-time demands. These additional resources can be placed in either edge computing devices (physically alongside the elements) or within the fog (anywhere between the edge and the cloud), leveraging low latency deterministic networks for hard real-time requirements. Recent advances in cloud computing allow to manage distributed workloads at the edge, using distributions like Minikube, which are optimized for embedded devices. It is the alternative showed in the diagram, as it reduces complexity and aligns with the workplace's networking infrastructure. In this instance, the edge cluster oversees a ROS-based framework with various modules: exchanging data with physical devices, executing 6D pose estimation models, managing robot control, and communication modules to share information with other components outside the ROS framework (Fig. 2).

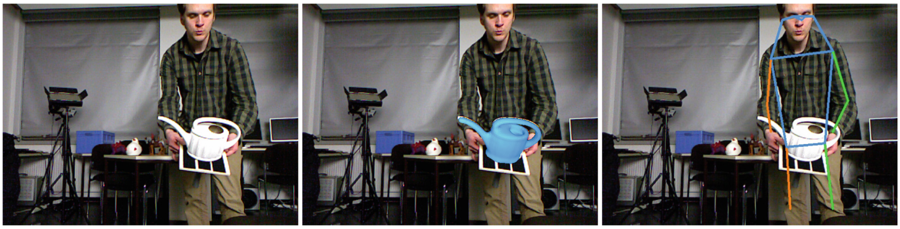


Fig. 2. From left to right: sample input image, object pose estimation result, human pose estimation result.

In both alternatives (edge or fog), managing and orchestrating the deployment of AI-driven workflows in distributed, independent clusters is necessary, handled through a"cluster management" component in the diagram. This component allows administrators to discover new model versions from cloud repositories, and perform centralized updates for one or multiple robotic agents within the same environment. Additionally, this component facilitates managing the deployment of numerous models across multiple robotic agents, requiring it to be placed in an intermediate fog cluster or on-premise servers. This ensures connectivity to various robots within an organization, encompassing all different units. The "IIoT Platform", which collects information from different devices and manages communications between them and other components, along with the cluster manager, is also deployed in this intermediate cluster.

Lastly, the Cloud Services encompass the highest level of information processing obtained from the plant, managing larger-scale functions like model training, data management, or storage, which do not require an immediate response and may demand higher processing capacity. Applications and tools providing such services work at this level, with deployment managed through integration. In the diagram, the main functions highlighted refer to the continuous integration and deployment (CI/CD) of workloads and models.

References

1. Adams, J., et al.: Cooperative material handling by human and robotic agents: module development and system synthesis. Expert Syst. Appl. **11**(2), 89–97 (1996)

2. Ajoudani, A., Zanchettin, A.M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., Khatib, O.: Progress and prospects of the human-robot collaboration **42**(5), 957–975. https://doi.org/10.1007/s10514-017-9677-2
3. Amorim, A., Guimares, D., Mendona, T., Neto, P., Costa, P., Moreira, A.P.: Robust human position estimation in cooperative robotic cells. Robot. Comput.-Integr. Manuf. **67**, 102035 (2021)
4. Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C.: Learning 6D object pose estimation using 3D object coordinates. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014, Proceedings, Part II 13, pp. 536–551. Springer (2014)
5. Chen, C.H., Ramanan, D.: 3D human pose estimation=2D pose estimation+matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7035–7043 (2017)
6. Corsetti, J., Boscaini, D., Poiesi, F.: Revisiting fully convolutional geometric features for object 6D pose estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2103–2112 (2023)
7. Fan, J., Zheng, P., Li, S.: Vision-based holistic scene understanding towards proactive human-robot collaboration. Robot. Comput.-Integr. Manuf. **75**, 102304 (2022)
8. He, Y., Huang, H., Fan, H., Chen, Q., Sun, J.: Ffb6d: A full flow bidirectional fusion network for 6d pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3003–3013 (2021)
9. Labbé, Y., et al.: Megapose: 6D pose estimation of novel objects via render & compare. arXiv preprint arXiv:2212.06870 (2022)
10. Li, S., Chan, A.B.: 3D human pose estimation from monocular images with deep convolutional neural network. In: Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, 1–5 November 2014, Revised Selected Papers, Part II 12, pp. 332–347. Springer (2015)
11. Li, Z., Wang, G., Ji, X.: CDPN: coordinates-based disentangled pose network for real-time RGB-based 6-DoF object pose estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7678–7687 (2019)
12. Liu, J., Sun, W., Liu, C., Zhang, X., Fu, Q.: Robotic continuous grasping system by shape transformer-guided multi-object category-level 6D pose estimation. IEEE Trans. Ind. Inform. (2023)
13. Martinez, J., Hossain, R., Romero, J., Little, J.J.: A simple yet effective baseline for 3D human pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2640–2649 (2017)
14. Rodriguez-Guerra, D., Sorrosal, G., Cabanes, I., Calleja, C.: Human-robot interaction review: challenges and solutions for modern industrial environments. IEEE Access **9**, 108557–108578 (2021)
15. Wang, C., et al.: Densefusion: 6D object pose estimation by iterative dense fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3343–3352 (2019)
16. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes. arXiv preprint arXiv:1711.00199 (2017)
17. Zhang, J., Keramat, F., Yu, X., Hernandez, D.M., Queralta, J.P., Westerlund, T.: Distributed robotic systems in the edge-cloud continuum with ROS 2: a review on novel architectures and technology readiness. Institute of Electrical and Electronics Engineers Inc. (2022)

A Concise Taxonomy of Human-Robot Interactions and Soft Skills Synergy

Wael M Mohammed(✉), Angela Lago Alvarez, and Jose L. Martinez Lastra

FAST-Lab, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 600, 33014 Tampere, Finland
wael.mohammed@tuni.fi

Abstract. With the substantial interest in human-centered manufacturing, human interactions with robots and embodied Artificial Intelligence (AI) have become more significant. Major topics related to the Human-Robot Interactions (HRI) include the mode of the interaction and the aspect of interaction. The mode of interaction addresses the method of the interaction and the nature of the information that the human and the robot exchange. The aspects of the interactions represent the dimensions of the interaction like physical and psychological among others. Therefore, this article provides a brief taxonomy to shape the landscape of the HRI. This taxonomy will help in determining the human skills and their relationship with robot interaction.

Keywords: Human-Robot Interaction · Soft Skills · Taxonomy

1 Introduction

HRI concerns the exchange of information between the human and the robot, facilitated by advances in sensors, communication, and actuation technologies, to enable an adequate level interaction. However, this interaction is still far from the level of the Human-Human interaction [1] due to its complexity. In fact, human interaction, which relies on soft skills like interpersonal skills, communication, listening, time managements and problem-solving among others, makes HRI challenging. Thus, developing solutions for HRI may benefit from understanding how humans interact with each other's and the impact of the soft skills.

Therefore, this article presents an HRI taxonomy in relation to the soft skills that the human will need in the interaction. Furthermore, this research work is part of the common work of the European projects AI Powered human-centered Robot Interactions for Smart Manufacturing (AI-PRISM). AI-PRISM aims to provide an ecosystem of collaboration and cooperation between Humans and Robots in manufacturing applications.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 162–167, 2024.
https://doi.org/10.1007/978-3-031-76428-8_31

2 Related Research Review

Interpersonal skills (IPS), also referred to as human skills, are a crucial aspect of human-human interactions [2], encompassing abilities to interact effectively and establish social relationships. They are necessary, not only for personal achievements but also for professional success. The European Skills Agenda, an European Commission initiative, prioritizes the development of IPS within the workforce, aiming to optimize their utilization in professional environments [3].

In the professional context, job-related skills determine the individual´s ability to fulfill job or task requirement. These skills, often referred to as technical skills, are field-specific or industry-specific and they can be acquired through education, training, or work experience. Alongside these, 'soft' skills, transferable across jobs, are gaining importance. These professional skills are dynamic and evolve with changes in technology, societal needs, and industry trends, making continuous learning and upskilling integral to professional development.

Human-Machine Interaction (HMI) is defined as an exchange of information between the human and the machine (e.g., computer, AI-application) [4]. Generally, this exchange allows humans and machines to complete tasks. The nature of these machines dictates the nature of the interactions. HRI is any exchange of information between humans and the robots. More precisely, a robot in this context represents an embodied AI [5]. When a human interacts with a robot, the interaction is recognised by an AI-based system. Within this interaction, the AI system behaves physically through the robot´s physical system or non-physically through other methods like visual or audio-based methods. Adding the physical aspect to the interaction results in what is known as Human-Robot Interaction. HRI may include different aspects, such as Human-Robot Collaboration (the physical aspect of the HRI), the emotional aspect (the feeling of the human while interacting with the robot), and the social aspect (the acceptability of the robot to communicate with the human). Other aspects may include cultural and psychological as well. There is no existing commercially available solution that encompasses all the possible aspects in one solution.

Taxonomy refers to the systematic classification of items or topics within a domain to help the readers categorize topics and build a logic toward understanding the structure of topics within its domain [6]. In HRI, Yanco and Duri present in [7] a taxonomy that classifies HRI based on task type, task criticality, robot morphology, ratio of humans to robots and robots to humans, compositions of robots team, level of shared interaction among teams, interaction roles, physical proximity between human and robots, decision support for operators, time/space, and autonomy level. This taxonomy provides clear and various categorizations of HRI. In a related taxonomy, Ferber [8] presents a taxonomy that describes the interactions between agents in a multi-agent system, classifying the interaction based on the goals, number of resources, skills, and type of situation. Another simple, yet accurate taxonomy of HRI is presented in [9], where HRI is categorized based on time, space, aim, and contact.

3 The Taxonomy

Embodied-AI refers to AI-based applications controlling physical systems like robots with autonomous independence. Build the taxonomy for HRI (Embodied-AI) requires an interaction model with entities owning sensors and effectors for signal exchange as presented in Fig. 1. There, the sensors, i.e. ears, receives signals and the effectors, i.e. hands, generate signals. Then, each entity includes skill sets that contribute to decoding the incoming signals, making decisions, and encoding the outcoming signals.

In HRI, the interaction depends mainly on two pillars: interaction mode and interaction aspect. The mode of interaction allow communication between humans and robots. From the human viewpoint, three systems engage during the interaction: the visual, auditory and the somatosensory systems. From robot's side, Loizaga et al. [10], describe human activities that can be digitalised using electronic sensors, like brain, eyes, respiration, electrodermal, cardiovascular, body temperature, and muscular activities. Besides them, robots also are capable to use visual, auditory and haptic modes to interact with the human.

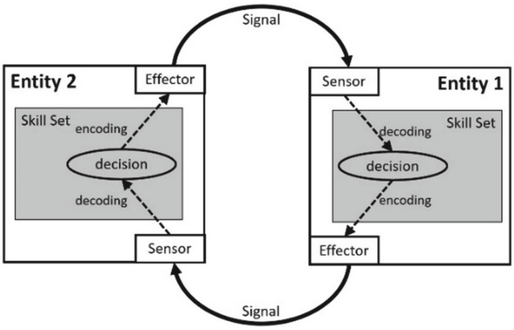


Fig. 1. Generic interaction model.

The aspect of the interaction defines the scope and forms the boundaries of the interaction to achieve the common aim. In this research, four aspects has been identified that can occur simultaneously between humans and robots. The first one is the physical aspect involving contact or non-contact interactions. As an example, human robot collaboration is a contact interaction and, cooperation or coexistence are non-contact physical interactions, as illustrated in [9]. This aspect exists as a bi-directional interaction where robot and the human take the role as signal receiver and signal generator.

The second aspect is the psychological aspect, which considers the human emotional state and human factors in the interaction. While robots lack emotions, humans can build empathy towards them [11], making the evolving emotional cognition in HRI noteworthy. Several research presents approaches for improving human experience by understanding their signals. Authors in [12, 13] highlight the significance of soft skills such as empathy, understanding, and effective communication in HRI. This way, any HRI must consider the psychological aspect to prevent possible problems like safety.

Thirdly, the social aspect of the interaction, that focuses on the adaptability of the interaction when it may include several humans and robots. In this regard, the humans and the robots must develop specific understanding of every possible counterpart. Moreover, robots must behave to suit human's social comfort.

The fourth aspect is the intellectual aspect. It involves long-term, multi-session interactions where humans and robots exchange knowledge being able to learn from and teach each other. This aspect is important to reach seamless, or close to, human-human interaction. To complement HRI taxonomy, Table 1 presents a matching of the needed soft skills for each aspect, soft skills have been adopted from Polakova et al. [14]. The selected resource categorises the soft skill into two categories: interpersonal skills, which enable the individual to work with others, and the intrapersonal skills that address the cognitive and self-management. Developing the soft skills for HRI involves complex challenges such as understanding human behaviour, guaranteeing long-term interactions, scalability, safety, privacy ethics, and measuring effectiveness [15]. Additionally, ongoing research in this field need to ensure they do not negatively impact human social and psychological wellbeing.

Table 1. Relations between the HRI aspects with the soft skills and the possible interaction modes.

Aspect	Soft skills	Possible Interaction Mode
Physical	Communication Skills, Organisational/Managerial Skills, Teamwork, Leadership Skills, Analytical and Critical Thinking	Visual, auditory, somatosensory, or haptic
Psychological	Initiative and Engagement, Emotional Intelligence, Persistence, Communication Skills	Brain, eyes, respiration, electrodermal, cardiovascular, body temperature and muscular activities
Social	Cope with ambiguity/uncertainty, Flexibility, Initiative and Engagement, Taking Responsibility, Emotional Intelligence	Brain, eyes and respiratory activities
Intellectual	Creativity, Analytical and Critical Thinking, Initiative and Engagement, Value Orientation	Brain activities, visual, auditory, somatosensory, or haptic

4 Conclusion

Besides the industrial applications, HRI must address the interactions in civil applications where the human is not trained and prepared to interact with the robots. The presented concise taxonomy provides preliminary view on the relations between soft skills and the interaction aspects. During the research, it was noticed that the human might be required to develop new skills or reskill themselves during the HRI. For the future work, this research will benefit from building systematic method for categorising the soft skills with respects to the relation to the interaction aspects. Additionally, an empirical classification of the contribution of the soft skill on the HRI can be valuable.

Acknowledgement. The research leading to these results has received funding from the European Union's Horizon Europe research and innovation program under grant agreement n° 101058589 which corresponds to the AI-PRISM project.

References

1. Celiktutan, O., Skordos, E., Gunes, H.: Multimodal Human-Human-Robot Interactions (MHHRI) Dataset for Studying Personality and Engagement. IEEE Trans. Affect. Comput. **10**, 484–497 (2019). https://doi.org/10.1109/TAFFC.2017.2737019
2. Beenen, G., Fiori, M., Pichler, S., Riggio, R.: Editorial: Interpersonal skills: individual, social, and technological implications. Front. Psychol. **14**, 1209508 (2023). https://doi.org/10.3389/fpsyg.2023.1209508
3. European Skills Agenda - Employment, Social Affairs & Inclusion - European Commission (2023). https://ec.europa.eu/social/main.jsp?catId=1223&langId=en. Accessed December 17
4. Eyam, A.T., Mohammed, W.M., Martinez Lastra, J.L.: Emotion-driven analysis and control of Human-Robot interactions in collaborative applications. Sensors (2021). https://doi.org/10.3390/s21144626
5. Akkaladevi, S.C., Heindl, C.: Action recognition for human robot interaction in industrial applications. In: 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), pp. 94–99 (2015). https://doi.org/10.1109/CGVIS.2015.7449900
6. Raven, P.H., Berlin, B., Breedlove, D.E.: The origins of taxonomy: a review of its historical development shows why taxonomy is unable to do what we expect of it. Science **174**, 1210–1213 (1971). https://doi.org/10.1126/science.174.4015.1210
7. Yanco, H.A., Drury, J.: Classifying human-robot interaction: an updated taxonomy. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583) **3**, 2841–2846 (2004). The Hague, Netherlands: IEEE. https://doi.org/10.1109/ICSMC.2004.1400763
8. Ferber, J.: Multi-agent systems: an introduction to distributed artificial intelligence. 1. publ. Harlow Bonn: Addison-Wesley (1999)
9. Schmidtler, J., Knott, V., Hölzel, C., Bengler, K.: Human centered assistance applications for the working environment of the future. Edited by Christopher M. Schlick and Jennifer Bützler. Occupational Ergonomics **12**, 83–95 (2015). https://doi.org/10.3233/OER-150226
10. Loizaga, E., Eyam, A.T., Bastida, L., Martínez Lastra, J.l.: A Comprehensive Study of Human Factors, Sensory Principles, and Commercial Solutions for Future Human-Centered Working Operations in Industry 5.0. IEEE Access **11**, 53806–53829 (2023). https://doi.org/10.1109/ACCESS.2023.3280071
11. Chin, J.H., Haring, K.S., Kim, P.: Understanding the neural mechanisms of empathy toward robots to shape future applications. Frontiers in Neurorobotics 17 (2023)
12. Apraiz, A., Lasa, G., Mazmela, M.: Evaluation of user experience in human-robot interaction: a systematic literature review. Int. J. Soc. Robot. **15**, 187–210 (2023). https://doi.org/10.1007/s12369-022-00957-z
13. Bonarini, A.: Communication in human-robot interaction. Current Robotics Reports **1**, 279–285 (2020). https://doi.org/10.1007/s43154-020-00026-1

14. Poláková, M., et al.: Soft skills and their importance in the labour market under the conditions of Industry 5.0. Heliyon **9**, e18670 (2023). https://doi.org/10.1016/j.heliyon.2023.e18670
15. Natarajan, M., et al.: Human-robot teaming: grand challenges. Current Robotics Reports **4**, 81–100 (2023). https://doi.org/10.1007/s43154-023-00103-1

Advancing Human-Robot Collaboration by Robust Speech Recognition in Smart Manufacturing

Oliver Avram[1](✉), Corrado Fasana[2], Stefano Baraldo[1], and Anna Valente[1]

[1] University of Applied Sciences of Southern Switzerland (SUPSI), 6962 Lugano, Viganello, Switzerland
oliver.avram@supsi.ch

[2] MCH-Tronics, Centro Galleria 3D, Via Cantonale, 6928 Manno, Switzerland

Abstract. In the dynamic context of human–robot collaboration (HRC) for smart manufacturing, AI's evolving capabilities present a diversified spectrum of opportunities. This paper emphasizes the crucial role of Automatic Speech Recognition (ASR) as a key enabler in collaborative robotics, addressing challenges posed by industry-specific jargon and communication styles through a prompting approach. Furthermore, it promotes the transformative potential of speech recognition on resource-limited devices as a fast-track towards unlocking the full collaborative capabilities of robots across diverse industrial applications.

Keywords: human-robot collaboration · artificial intelligence · ASR error detection · ASR error correction · prompt engineering · spelling correction

1 Introduction

As the number of robots used in industry increases, it is crucial to find fast but robust ways in which humans and robots can collaborate. In contrast to conventional programming methods and interfaces, nowadays more intuitive approaches can be employed for human-robot interaction [1]. In this regard, speech is probably the most intuitive means of communication, and its integration into HRC scenarios [2, 3] stands as a transformative step towards a more productive and synergistic working relationship between humans and robots. However, speech recognition errors stemming from factors like background noise, varied pronunciations, and industry-specific term misinterpretations, can induce frustration, especially when users must persistently reiterate sentences that the ASR system consistently fails to understand. Furthermore, in multimodal systems, inaccurate speech recognition disrupts the integration of different input modalities, jeopardizing the coherent interpretation of user intent. Ultimately, this frustration undermines the user experience, eroding confidence in the ASR system and adversely affecting the utility and reliability of the entire HRC interaction.

Starting from these premises, a robust speech recognition system might prove pivotal towards boosting the accessibility and productivity of collaborative robotic systems, and

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 168–173, 2024.
https://doi.org/10.1007/978-3-031-76428-8_32

towards enabling the robots to learn new tasks without explicit coding, by the joint use of intent recognition and imitation learning [4, 5]. For example, the exploitation of ChatGPT to command robots has been recently proposed [6]. Although the obtained results are promising, the users must consistently verify that generated commands align precisely with their intentions.

This research focuses on optimizing speech recognition for accuracy, relevance, efficiency, and adaptability to the unique language and terminology of specific industrial scenarios. Furthermore, full consideration is given to speech recognition for portable edge devices, which provides benefits like offline functionality, faster response times and security, which are highly valuable in industrial setups.

2 Methodology

In recent years, several promising Deep Learning-based ASR models (e.g., Wav2Vec 2.0 [7], and Fast Conformer [8]) have been developed. Many of these models are trained on noise-free, native-English audio datasets, potentially leading to suboptimal performance in industrial scenarios with high noise levels and diverse accents. OpenAI has addressed this with Whisper [9], an ASR model trained on a diverse dataset to handle different accents and noise levels (SNR below 10 dB). However, for optimal performance in industrial applications, recognizing domain-specific vocabulary is crucial. This motivated the development of an approach to easily bias token probabilities toward specific terminology that the model may not have been extensively trained on.

To enhance technical term recognition, the commonly used method is fine-tuning. However, this involves re-training the model and gathering extensive annotated data, which is time-consuming and sometimes impractical. In this work, we propose a simpler, faster solution using prompting [10] as a spelling guide to guide Whisper in handling industry-specific terms without requiring model re-training or extensive data collection. The proposed framework, shown in Fig. 1, also considers the portability of an optimized ASR model to mobile devices. The initial step involves a use case description, facilitated through a discussion with a domain expert, to pinpoint industry-specific terms

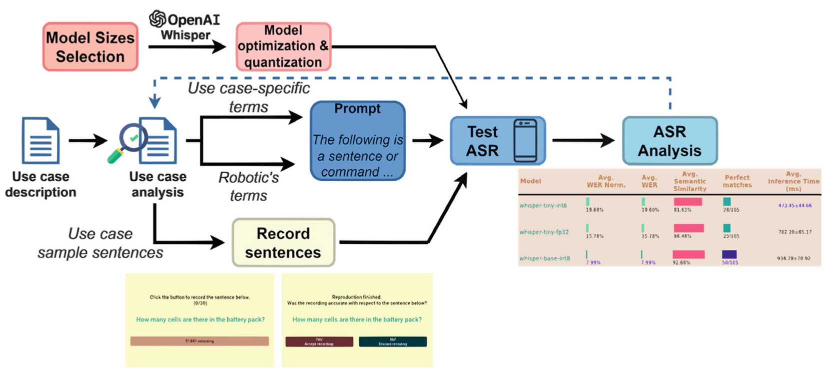


Fig. 1. Framework overview

challenging for ASR. These terms are then added to the prompt to boost their recognition probability. The discussion also identifies sentences used by human operators as potential commands during HRC. Once identified, a limited number of speech samples based on these sentences are recorded to optimize the specialized terms list.

For industrial applications, having a compact ASR model on a mobile device is advantageous for operator privacy, eliminating the need for large machines or cloud usage. However, such setting poses challenges due to the limited memory and processing resources, leading to potential performance issues. For this reason, in this work the Whisper model was optimized and quantized to allow its execution on a mobile platform while thoroughly evaluating the size-accuracy trade-off. The obtained model is tested on the previously recorded sentences, both with and without the prompting functionality. The analysis of the results offers valuable insights into adapting the prompt for better understanding use-case-specific terminology. Terms easily recognized without prompting can be removed, while less recognized terms are added to the prompt.

3 Use Case Scenarios and Preliminary Results

The proposed framework was employed to record English speech samples from 11 speakers (6 males and 5 females, from 4 EU countries) with different accents and English levels. Each subject was asked to record a total of 30 sentences containing terms specific to three manufacturing Use Cases (UCs) highly relevant for the implementation of HRC dynamics: battery disassembly (UC1), scanning and repairing of metal parts (UC2) and airplane engine nacelle assembly (UC3). Speakers were allowed to redo the recording to correct errors. After eliminating mispronounced sentences, the dataset contained 116, 109 and 105 recordings for the three UCs respectively. Some examples of recorded sentences can be seen in Table 1.

Based on the UCs' specific terminology, adapted prompts were defined. A prompt is exemplified in the following for UC3: "*The following is a sentence or command pronounced by a human operator while working with robots or collaborative robots named cobots. The phrase includes words from the robotic context such as robot, cobot, arm, tool, flange, joint, gripper, and screwdriver. These words can also appear in plural form. More importantly, the command contains terms specific to the use case such as aircraft, engine, nacelle, inner barrel, lip, D-DUCT, inlet cowl, aft bulkhead, forward bulkhead, fastener, thrust reverser, anti-ice duct, anti-ice exhaust, fan cowl, and acoustic panel. Each of them can also be in the plural form*". Apart from the listing of exact terms, the body of the three prompts is identical for all the UCs.

Implementing robust ASR models on edge devices may result in a non-negligible performance drop during quantization. Nevertheless, lots of work have been done recently to reduce the gap between unquantized and quantized models. In this work, the Whisper model has been quantized to INT8 precision using the Microsoft Olive framework, achieving performances comparable to that of the original FLOAT32 model. Due to the limitations of the used inference device, only the Tiny and Base versions of the model could be exploited.

In Fig. 2, the results of the tests carried out on a Snapdragon® 8 Gen 2 Mobile HDK, both with and without prompting, are presented in terms of Word Error Rate (WER). The WER decreases as model size increases in all three cases. For UC1, the Base model without prompting has over 6% WER, dropping by 3% with prompting. This trend continues across the other UCs with a more significant difference (7% for UC2 and 9% for UC3). Larger models generally exhibit better recognition capabilities, aligning with expectations. The performance gap between prompting and non-prompting decreases with increasing model size but remains significant. The prompting shows its value especially for smaller models, which becomes even more effective if the model is quantized. Examining the Whisper Base model without prompting, we can argue that differences in transcript performance across UCs relate to industry-specific terminology in the training data. Common and well-represented terms in publicly available data result in lower WER, while highly specialized or niche terms lead to lower performance, as seen in the battery disassembly UC versus the assembly of the engine nacelle UC. Overall, the number of accurately transcribed sentences significantly increases thanks to prompting, which allows improved recognition of specific terms, as shown in Table 1. Using prompting does not have a major impact on performance (max. ~ 140 ms across the conducted experiments).

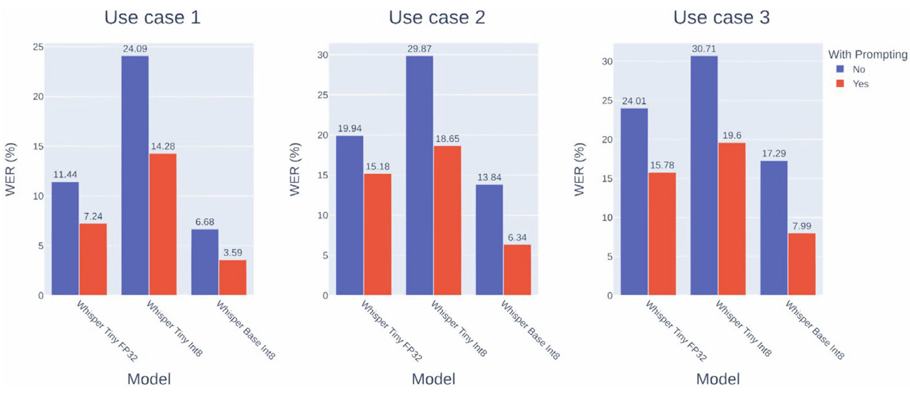


Fig. 2. Preliminary results

Table 1. Example of transcriptions with and without prompting.

Ground Truth	W/O Prompting	W/ Prompting
UC1 - How many cells are there in the battery pack?	Our ministers are there in the battery pack. *(Tiny Int8 – **WER**: 33.33%)*	How many cells are there in the battery pack? *(Tiny Int8 – WER: 0.00%)*
UC1 - Cobot, mount the suction cup gripper and remove the cells.	Cobbott, mount the suction cup, repair and remove the cells. *(Base Int8 – WER: 20.00%)*	Cobot, mount the suction cup, gripper, and remove the cells. *(Base Int8 – WER: 0.00%)*
UC2 – Robot, mount the Artec scanner.	Private mount D-ArtXCAMER. *(Tiny Int8 – WER: 80.00%)*	Robot, mount, the artec scanner. *(Tiny Int8 – WER: 0.00%)*
UC3 – Robot, mount the gripper and secure the acoustic panels using fasteners.	Robert, Mount the Reaper and Secure the acoustic panace using Fosters. *(Tiny Int8 – **WER**: 36.36%)*	Robot, mount the gripper and secure the acoustic panel~~s~~ using fasteners. *(Tiny Int8 – **WER**: 9.09%)*
UC3 - Some of the nacelle components include the aft bulkhead, forward bulkhead, inlet cowl and fan cowl.	Some of the Nacel components include the F-Dball-CAD, Forward-Ball-CAD, Inlet-Cowl and Fan-Cowl. *(Base Int8 – **WER**: 56.25%)*	Some of the nacelle components include the aft bulkhead, forward bulkhead, inlet cowl and fan cowl. *(Base Int8 – **WER**: 0.00%)*

4 Conclusion

This work proposes a framework to enhance the performance of an ASR model with respect to domain-specific vocabulary and discusses its portability to a mobile platform. The preliminary results obtained on a limited but relevant speech dataset confirm the prompting as a high-potential technique to enhance speech recognition with a focus on accuracy, efficiency and user experience improvement. Furthermore, it promotes a fast-track scalability to industrial applications which consider speech recognition on edge devices as a catalyst for a harmonious human-robot collaboration, ultimately enhancing productivity, well-being and the overall efficiency of manufacturing processes.

References

1. Villani, V., Pini, F., Leali, F., Secchi, C., Fantuzzi, C.: Survey on human-robot interaction for robot programming in industrial applications. Ifac-PapersOnline **51**(11), 66–71 (2018)
2. Saren, S., Mukhopadhyay, A., Ghose, D., et al.: Comparing alternative modalities in the context of multimodal human–robot interaction. J. Multimodal User Interf. (2023). https://doi.org/10.1007/s12193-023-00421-w
3. Chen, H., Leu, M.C., Yin, Z.: Real-time multi-modal human–robot collaboration using gestures and speech. J. Manuf. Sci. Eng. **144**(10), 101007 (2022)
4. Stepputtis, S., et al.: Language-conditioned imitation learning for robot manipulation tasks. Advances in Neural Information Processing Systems **33**, 13139–13150 (2020)
5. Bucker, A., et al.: Latte: Language trajectory transformer. In: 2023 IEEE International Conference on Robotics and Automation, ICRA, pp. 7287–7294. IEEE (2023)

6. Vemprala, S., Bonatti, R., Bucker, A., Kapoor, A.: Chatgpt for robotics: Design principles and model abilities. Microsoft Auton. Syst. Robot. Res **2**, 20 (2023)
7. Baevski, A., Zhou, Y., Mohamed, A., Auli, M.: wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Info. Proc. Sys. **33**, 12449–12460 (2020)
8. Rekesh, D., et al.: Fast Conformer with Linearly Scalable Attention for Efficient Speech Recognition. arXiv preprint arXiv:2305.05084 (2023)
9. Radford, A., et al.: Robust speech recognition via large-scale weak supervision. In: International Conference on Machine Learning, pp. 28492–28518. PMLR (2023)
10. Qiao, S., et al.: Reasoning with language model prompting: a survey. arXiv preprint arXiv: 2212.09597 (2022)

Check for updates

Human-Robot Collaboration in the Industry 5.0 Era: The Aerospace Perspective

José Ramón Vilanova Sánchez(✉), Rafael Luque(✉), Paloma Vega(✉), and Eduardo Ferrera(✉)

Center for Advanced Aerospace Technologies, La Rinconada 41309, Seville, Spain
{jrvilanova,rluque,pvega,eferrera}@catec.aero

Abstract. Industry 5.0 is meant to achieve a new level in terms of innovation and productivity. To do so, Human-Robot Collaboration (HRC) is going to play a crucial role: the combination of human's flexibility, with the repeatability and tirelessness of robots offers numerous opportunities for improvement in this new era. This paper explores HRC from the human perspective, showing a suite of the innovative technologies that, driven by AI and DL, and supported by robots, will raise the next level of industrial efficiency. Voice Assistance systems, Computer Vision powered benches, smart workstations or Extended Reality (XR) are meant to empower individuals on the manufacturing front. The paper also presents several solutions that have been implemented in a pilot aerospace factory to empower futuristic industrial operators, and how those solutions should be combined with cognitive robots to shape the roadmap and future of aerospace with precision, efficiency, and a touch of humanity.

Keywords: Industry 5.0 · Human-Robot Collaboration · Deep Learning · Artificial Intelligence · Aerospace

1 Introduction

In the contemporary industrial landscape, the emergence of Industry 5.0 stands as a pivotal revolution in manufacturing. The term Industry 5.0, introduced by Raba et al. [1], shifts away from Industry 4.0's virtual focus, embracing a human-centric paradigm. In such paradigm, both physical and digital tools, encompassing hands-on instruments and advanced technologies, collaborate with human capabilities to enhance them.

Industry 5.0 is characterized by the seamless fusion of technologies such as AI, IoT or XR. In general terms, Industry 5.0 converts the technology-oriented Industry 4.0 to a human-centric approach. In it, the technology is developed to improve the well-being of operators while increasing their productivity [2]. This marks a new era where both humans and machines are in the loop, transforming industrial paradigms for holistic efficiency and flexibility, completing the transition to a human-centric, sustainable, and resilient industry.

Aerospace industry represents a cutting-edge sector in such revolution, in which innovations remain essential. Yet, the level of automation falls short of optimal solutions,

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 174–179, 2024.
https://doi.org/10.1007/978-3-031-76428-8_33

due to critical factors such as low-cadence production pieces, flexibility needs and high-standard requirements [3]. The high costs related to automation systems are mostly unaffordable for SMEs and only occasionally assumed in big companies, delegating a large majority of work to human labours.

Hence, implementing the concept of Operator 5.0 [4] becomes a need for the sector, providing a low-cost, high-efficiency HRC. Currently these research lines are getting special attention, assigning the operator the most flexible tasks while robots take on the most repetitive, tedious and non-ergonomic ones [5].

This work describes some of the most promising technologies for implementing Operator 5.0 in the aerospace sector. The presented solutions integrate Voice-Assisted and Computer Vision systems, smart workstations, and XR frameworks. These technologies aim to empower and support human roles in manufacturing, enhancing worker's productivity through Human-Robot Interaction.

More precisely, the contributions of this paper are:

- A brief review of AI in Human-Machine Interactions, describing some cutting-edge developments.
- An overview of cognitive and smart systems implemented in the aerospace sector which include voice and visual guidance, and support systems for operators.
- The implementation of a Human-Robot interaction through anthropomorphic and mobile smart robots.

The structure followed in this publication is: Sect. 2 describes some AI-based systems that enable Human-Machine Interaction, Sect. 3 describes robot assistance based on cognitive systems, finally, Sect. 4 describes the conclusions and a roadmap of the solutions developed.

2 Artificial Intelligence in Human-Machine Interactions

The work presented in this paper entails some of the research done by CATEC in the project 5R Cervera Network, where the RTO is developing a Pilot Factory for interactive monitoring and support of manual processes, devoted to enhancing aerospace Industry. The following subsections presents some of the tools developed, to demonstrate the advantages of the Operator 5.0 approach.

2.1 SIMO: Speech Intelligence System to Manage Industrial Operations

In modern industry, the rise of Natural Language Processing (NLP) models and Speech Recognition Systems (SRS), fueled by the advent of Large Language Models (LLM), has become a prominent trend. Particularly, for industrial settings, characterised by paper-based processes, there exists a vast potential for innovation in terms of interactive speech recognition.

To address this need, CATEC has developed SIMO [6], a Voice Assistant system meticulously crafted for the challenges of noisy factory environments and collaborative robotics. Armed with state-of-the-art Deep Learning models, SIMO offers processing modules like NLP and TTS, to enhance human-machine interactions. SIMO has been designed as a ROS package, meant to be integrated with robotic applications that works in the industrial facilities.

Moreover, SIMO operates on a local server, ensuring continuous functionality even in areas where connectivity is restricted due to data security and privacy, such as many aerospace factories. Beyond mere recognition, SIMO elevates operations with hands-free functionality, real-time performance feedback, and a transformative shift towards paperless industrial processes.

2.2 Interactive, Cognitive and Multipurpose Smart Workstations

Workstations are an essential tool in aerospace assembly factories. Workers spend up to 8 h repeating manual tasks, before moving those manufactured pieces to the aircraft. Therefore, a workstation that adapts and understands the operators' needs, while interacting with them and the factory is a significant upgrade in the sector.

Industry 5.0 workstations needs to combine different AI systems, including Computer Vision, AR, DL and Speech Recognition techniques. Employing such capabilities, novel workstations are capable to provide personalized guidance to operators. Aerospace wire harnessing assembly provide a perfect use case, as it is one of the most complex and tedious tasks in manufacturing.

In the context of the 5R Network, CATEC has developed an AI based workstation [7], also designed as a ROS Package, where DL-based segmentation, in combination with laser-based projections offer a perfect framework for human-machine collaboration in manual tasks. This system is able to recognize, understand and guide human actions during harnessing tasks. The work also was furtherly extended by integrating SIMO, providing voice interaction to the loop, to enhance the worker's productivity and hands-free control of the tool (Fig. 1).

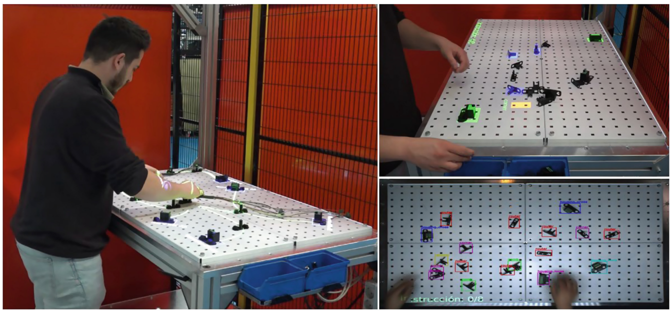


Fig. 1. Smart Workstation combining Deep Learning pieces recognition, SIMO and Augmented Reality systems.

Furthermore, CATEC is currently dedicated to extending such workstation over larger aerospace structures, particularly those that currently pose significant challenges to the industry, like fan cowl or Horizontal Tail Planes (HTP). In those endeavors, the integration of DL-powered AR serves not only as an instruction for equipping or assembling structures, but also as a guiding tool for visual quality verification tasks, which are essential for the rigorous standards prevalent in this sector.

2.3 Human Ergonomics: An Approach Based on a DL Human Recognition System

In the complex and demanding landscape of aerospace manufacturing, where the tasks lead to awkward and unnatural postures [8], prioritizing the well-being and ergonomics of human operators is essential to the industry's success. To guarantee this, CATEC is developing a DL Computer Vision safety management system, able to detect accidental falls and issuing calls for assistance. The design of this system uses YOLOv7-pose algorithm for keypoint detection and transmits help requests for triggering factory alarms or seeking support as needed.

Moreover, the real-time monitoring of operator's pose and gesture detection will empower a future work of CATEC, allowing aerospace manufacturers to proactively monitor and address ergonomic concerns. By attending this ergonomic concerns, operators will run the complex and tiring aerospace processes smoother and with a reduced likelihood of costly accidents.

3 Robot Assistance in Human-Machine Interactions

In the era of Industry 5.0, cutting-edge technology applications find their pinnacle in the collaboration between humans and robots. This collaborative paradigm envisions a harmonious interaction where cognitive machines, work alongside human operators, making their work less tedious, safer and more productive. The pillars of this collaboration are AI and DL: systems like the ones presented in Sect. 2, that gives the smart factory the capabilities to interpret human requests and to command robots thanks to human actions or gestures.

3.1 Autonomous Mobile Robots for Smart Logistics

The integration of Autonomous Mobile Robots (AMRs) heralds a new era of internal logistics efficiency in aerospace factories. In Industry 5.0, AMRs emerge as dynamic collaborators, navigating factory floors to fulfill a vast spectrum of supply tasks. Not only these AMRs autonomously transport materials and tools in the dynamic environments, but the combination with smart industries and workstations will also give them the capability to comprehend human requests.

This added layer of adaptability and responsiveness allows them to act as supporters in many scenarios, solving requests of operators, making them save traveling times and avoiding unnecessary work interruptions. CATEC is currently integrating an AMR, a RB-Theron from Robotnik, to solve simple requests from users in the aerospace factory pilot. Simple but impactful examples of such requests that add value to industrial operations can be "*hand me a battery for this electric screwdriver*" or "*I have no M8 screws left*".

3.2 Cobots for Manual Task Assistance

Collaborative robots have demonstrated their industrial utility in recent years but, nowadays, only their learning-by-demonstration and human-safety capabilities are used in current factories. For its implementation in Industry 5.0, cobots should improve and serve as a helper of the Operator 5.0 in tasks where a third hand is required. To support "third hand tasks", robots should be able to reinforce and complement what humans are doing, without interrupting their tasks.

In this context, the integration of voice recognition systems like SIMO allows a fluent communication between human and robot, avoiding the use of screens or keypads to command the robot actions. Moreover, the capabilities of smart workstations to recognize tools and pieces, combined with the track of human actions introduces a new level of collaboration, where cobots can predict next steps and support proactively the Operator 5.0, anticipating human requests.

4 Conclusions

Robots will play a key role in Industry 5.0, elevating industry to a new performance level. This paper describes some cutting-edge technologies based on Artificial Intelligence and DL-based systems, that not only set the ground for enhancing human operators, but also present the path for a real and efficient Human-Robot Collaboration. The integration of these systems, meant to impact directly in aerospace factories, with robotic collaborations would significantly improve the productivity of future factories, setting again the human in centre of the production, and providing assistance and guidance when needed.

Acknowledgments. This study has been partially funded by the project "5R, Cervera Network in robotic technologies for intelligent manufacturing", contract number CER-20211007 under "Centros Tecnológicos de Excelencia Cervera"; funded by "The Centre for the Development of Industrial Technology (CDTI)".

References

1. Rada, M.: INDUSTRY 5.0 - from virtual to physical, https://www.linkedin.com/pulse/industry-50-from-virtual-physical-michael-rada/. Last accessed 28 December 2023
2. Barata, J., Kayser, I.: Industry 5.0–Past, Present, and Near Future. Procedia Computer Science **219**, 778–788 (2023)
3. Judt, D., Lawson, C., Lockett, H.: Experimental investigation into aircraft system manual assembly performance under varying structural component orientations. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture **234**(4), 840–855 (2020)
4. Mourtzis, D., Angelopoulos, J., Panopoulos, N.: Operator 5.0: A survey on enabling technologies and a framework for digital manufacturing based on extended reality. Journal of Machine Engineering **22**, 5–31 (2022)
5. Chowdhury, S., et al.: Study of Robotics and Automation in the Aerospace Industry. In: 2nd International Conference on Mechanical and Applied Sciences (2022)

6. Luque, R., Galisteo, A.R., Vega, P., Ferrera, E.: SIMO: an automatic speech recognition system for paperless manufactures. Advances in science and Technology **132**, 129–139 (2023)
7. Luque, R., Blanco, E., Galisteo, A.R., Ferrera, E.: From augmented reality to deep learning-based cognitive assistance: an overview for industrial wire harnesses assemblies. In: Iberian Robotics conference 2022, pp. 113–124. Springer International Publishing (2022)
8. Seri, R.K., Norhidayah, H., Mohd, S.O.: A study on muscle fatigue associated with awkward posture among workers in aerospace industry. Advanced Engineering Forum **10**, 287–292 (2013)

Hybrid Robotic Control for Flexible Element Disassembly

Benjamín Tapia Sal Paz[1(✉)], Gorka Sorrosal[1], and Aitziber Mancisidor[2]

[1] Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), 20500 Arrasate, Spain
btapia@ikerlan.es

[2] Faculty of Engineering in Bilbao, University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain

Abstract. This work proposes a robotic arm control for the disassembly of flexible elements. With this, a control solution is sought for handling paths of flexible elements where the application of an adequate force is necessary to ensure the physical integrity of the element. The proposal consists of an hybrid scheme to control the interaction between the robot end-effector and its environment, where the use of a global and a local planner helps to adapt a theoretical trajectory with real-time information during the execution of the task. The proposed control is implemented in a simulated environment on the robotic arms Franka Emika and KUKA LBR iiwa, resulting in trajectories that reduce the force in the flexible element.

Keywords: hybrid control · robot manipulation · robotic disassembly

1 Introduction

Robotics present a flexible and affordable solution for many industrial cases. In some of them is important the control of the interaction between the robot end-effector and its environment. To meet the requirements set by that interaction, it is necessary to control both force and position [1]. However, it is not straightforward since position and force are dependent control spaces. For that reason, the use of classic position and force controllers is not suitable for these tasks.

To accomplish with that interaction requirements, compliant control methodologies are used in robotic arms [2]. The differences between them lie mainly in the determination of the interaction forces, and in how different types of feedback signals are incorporated to achieve both force and position control to guarantee the desired interaction. Impedance controllers [3] and hybrid controllers [4] are the most commonly used methodologies for the control of robotic interaction tasks. In the former, the establishment of a relationship between position and force generates a mass-spring-damper behavior. On the other hand, in hybrid controllers, position and force control work in parallel, thus generating orthogonal control spaces.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 180–185, 2024.
https://doi.org/10.1007/978-3-031-76428-8_34

Since these controllers are usually used in unstructured and dynamic tasks, main challenges are related to the significant work of design and adjustment of the parameters that guarantee the desired interaction. To Face them, this work proposes an hybrid robotic control for flexible elements disassembly. This control generates manipulation trajectories guaranteeing the application of an adequate force on the manipulated element.

2 Flexible Element Disassembly

Disassembly of flexible elements expose many of the challenges discussed in the previous section and is present in several industrial fields, in which we can highlight the disassembly and reuse of flexible elements used for sealing doors of household appliances, automobiles, among other devices. In them is necessary the control of the interaction between the robot end-effector and the flexible element in order to guarantee both the execution of the task and to safeguard the physical integrity of the element.

An important aspect in these are the interaction forces that appear during the execution of the task, which are dynamic and vary according to the state and characteristics of the specific flexible element being manipulated. These forces influence the development of the task, so it is necessary to be able to identify them to achieve the correct interaction control. This characteristic, where the robot must be able to identify the forces and act according to them, causes that nowadays these types of tasks are performed manually. Nonetheless, the repetitive and in some cases non-ergonomic characteristics of these tasks can be harmful to people. For that reason this work proposes a robotic system solution capable of carrying out this type of tasks.

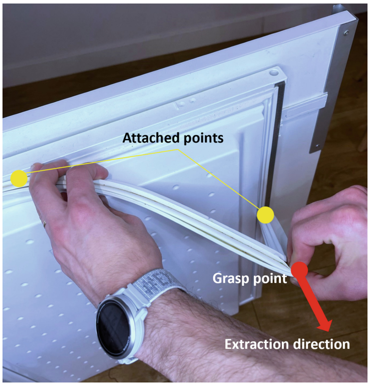


Fig. 1. Fridge flexible element disassembly task.

In Fig. 1, it can be observed the task of removing one of these elements in the door of a refrigerator. In this task, the flexible element must be detached from the door frame ensuring its physical integrity for future reuse. Thus, the objective of the task is to execute a removal trajectory that ensures the application of adequate force to the flexible element. Both trajectory planning and force exertion present a challenge due to the different conditions that arise between different cases, where different extraction paths and force exertions are suitable.

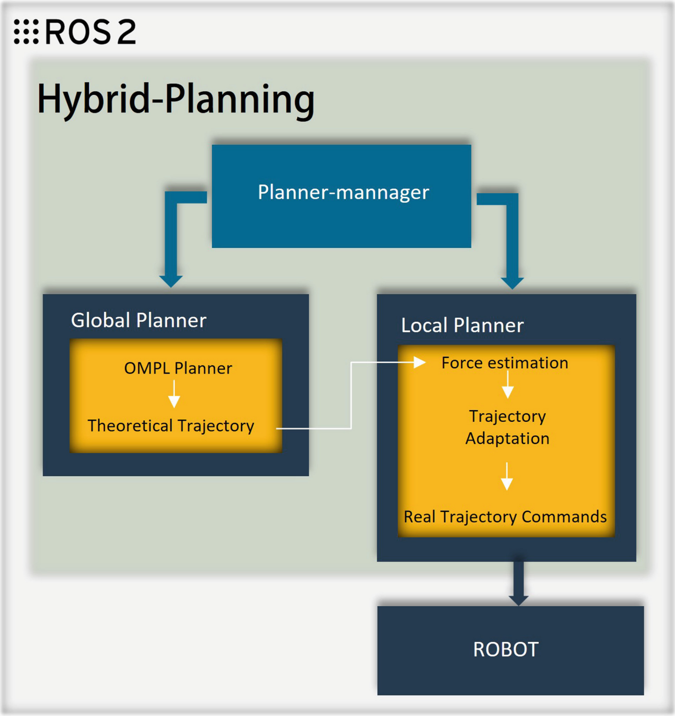


Fig. 2. Hybrid control scheme proposal with the use of hybrid-planning framework

2.1 Hybrid Force Control Strategy

For the challenges of flexible element manipulation, an hybrid force control is proposed based on the establishment of a reference trajectory that is modified by a force controller taking into account the interaction information of the system at runtime (green and blue trajectories in Fig. 3.)

The proposed control is developed on the ROS2 framework. This framework aims to improve the performance offered by other frameworks based on the "sense-plan-act" methodology, that are inadequate for tasks where the environment and requirements change dynamically during execution. This is because in interaction tasks, the environment may have changed from planning to execution. For this reason, it is necessary to migrate to methodologies where the action is linked to the sensing in real-time.

This framework addresses this problem through the use of global and local planners. These components have different functions and operate at different frequencies. In the case of the global planner, it is responsible for solving the global problem where there are no real-time requirements (theoretical trajectory). On the other hand, the local planner is responsible for adapting the solution of the global planner taking into account real-time information (interaction forces) about the environment.

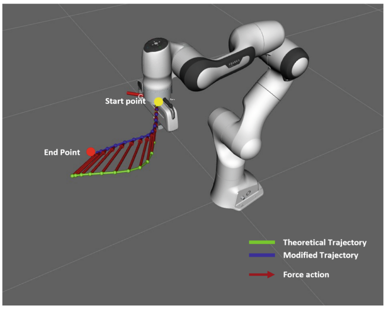
The implementation of the control proposal using the hybrid planning framework is shown in Fig. 2. Where the theoretical trajectory is generated by the global planner using the Open Motion Planning Library (OMPL) [5]. This trajectory is made up by consecutive waypoints where the local planner works. It looks for the following waypoint and calculates the interaction force in that point, and with a proportional force action adapts it. Finally, the local planner executes the movement to the deviated position and repeats this action through the entire trajectory.

To coordinate the calls to the global and local planners, the manager logic in this work consists of calling the global planner at the beginning of the task to generate the theoretical reference trajectory. The local planner is then responsible for determining the forces acting on the task to modify this trajectory and send it to the robot for execution.

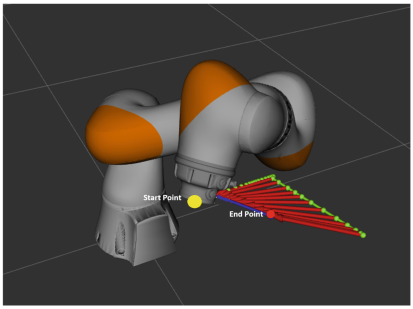
3 Results

The proposed control is implemented on the robotic arms Franka Emika and KUKA LBR iiwa, resulting in the trajectories shown in Fig. 3. To generate the reference trajectories, an arbitrary starting point was established (yellow point), which corresponds to the point where the flexible element is held by the support. The end position of the trajectory (red point) was chosen to replicate real-world applications. For the Franka Emika robot (see Fig. 3(a)), no planner requirements are specified. However, for the KUKA iiwa robot (see Fig. 3(b)), the planner is required to generate Cartesian trajectories. This approach allows for the evaluation of the proposal with different types of trajectories.

The green trajectories generated by the global planner represent theoretical extraction trajectories, but they do not consider the interaction forces present during execution. As the robot motion begins, the local planner calculate the force present and adjusts the trajectory accordingly. At this stage, the forces are generated in simulation assuming that the element follows a Hooke's law model (considering grasping point as the equilibrium position and a elasticity constant of 100 N/m). The red arrows represent the proportional action of the force control that modifies the trajectory. Finally the robot send the modified trajectory (blue trajectory) for execution. This trajectory removes the flexible element while safeguarding its integrity.



(a) Generic trajectories generation in Franka Emika robot



(b) cartesian trajectories generation in KUKA iiwa robot

Fig. 3. Simulation results of the proposed control strategy

4 Conclusions

This paper presents the implementation of a robotic control system for the disassembly of flexible elements. An hybrid control approach is proposed to adapt trajectories according to the interaction forces present during execution, with the use of the ROS2 hybrid-planning framework.

The resulting trajectories are smooth and continuous, making them suitable for the manipulation of flexible elements. The use of the global planner enables the specification of reference trajectories at the beginning of the task, and then with the local planner real-time information about the forces during the execution are incorporated to take force control actions and modified the trajectory to one in which less force is exerted on the flexible element.

The proposal was implemented in a simulation environment to evaluate the use of the hybrid-planning framework. The next step of this proposal is to implement it in the real world and work with the KUKA lbr iiwa robot. This will allow for the use of real-time force data obtained through the robot sensors.

Acknowledgements. This project is supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 955681, and by members of the VirtualSensorization Research Group from the University of Basque Country (Basque Goverment Ref. IT1726-22).

References

1. Song, P., Yu, Y., Zhang, X.: A tutorial survey and comparison of impedance control on robotic manipulation. Robotica **37**, 801–836 (2019)
2. Whitney, D.E.: Historical perspective and state of the art in robot force control. Int. J. Robot. Res. **6**(1), 3–14 (1987)
3. Hogan, N.: Impedance control: an approach to manipulation. In: American Control Conference. IEEE 1984, pp. 304–313 (1984)

4. Mason, M.T.: Compliance and force control for computer controlled manipulators. IEEE Trans. Syst. Man Cybern. **11**, 418–432 (1981)
5. Şucan, I.A., Moll, M., Kavraki, L.E.: The open motion planning library. IEEE Robot. Autom. Mag. **19**(4), 72–82 (2012)

Challenges in Situation Understanding and Scene Perception

Razane Azrou(✉), Selma Kchir, Raphaël Lallement, and Matteo Morelli

List, CEA, Université Paris-Saclay, Saclay, France
{razane.azrou,selma.kchir,raphael.lallement, matteo.morelli}@cea.fr

Abstract. During the execution of long-horizon tasks in complex environments, it is inevitable that a robot faces many situations that are not part of the nominal operation. To achieve long-term autonomy, one crucial capability of future robots is then to reason about their perception to detect unexpected situations and handle them. In this paper we discuss a number of challenges in situation understanding and scene perception and our current research towards the development of a safe adaptive, cognitive robotic deliberation system.

Keywords: Knowledge representation · situation understanding · situation learning · known anomalies · unknown anomalies

1 Introduction and Objectives

Long-term autonomy demands robots able to reason about their perception and prior knowledge to understand the context of execution and to adapt accordingly. Realizing this ability is a major challenge, as it combines research from i) situation modeling, ii) situation awareness, and iii) learning new situation models.

A recent survey [1] reviews ontologies to support robot autonomy in general, while others focus specifically on situation awareness [2–4]. A shortcoming is that each solution defines its own domain ontologies, resulting in incompleteness, inconsistences, and ambiguities, especially for those concepts that have a polysemantic nature (e.g., “Behavior”, “Activity”, “Skill”).

Situation understanding and scene perception focus on detecting discrepancies between predictions and observations, diagnosing their possible causes to propose recovery actions. Examples of projects that address this problem are SafeAdapt[1] and HORSE[2] respectively for the automotive domain and production lines. Existing works [5, 6] address faults anticipated at design time based on adaptive state machines and present limitations in unknown and dynamic environments. Other approaches [7] have performance limitations. In the case of high-dimensional data, data-driven approaches have

1 https://www.safeadapt.eu/.
2 http://www.horse-project.eu/Situation_Awareness.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 186–190, 2024.
https://doi.org/10.1007/978-3-031-76428-8_35

been proposed to detect and classify time series [8, 9], which can have a direct application to anomaly detection using reconstruction or classification techniques.

To develop a safe, adaptive and cognitive deliberation system capable of detecting unexpected situations and providing contingency plans to handle them, there are a number of key aspects related to situation understanding and scene representation:

- The development of a tool for inter-component knowledge exchange based on ontologies and knowledge graphs by means of common data models for real-time representation of the state of the environment. The knowledge base identifies relevant concepts and properties for situation understanding.
- Leveraging the developed knowledge representation formalism, state-of-the-art multi-sensory perception and the results on task and motion planning and verification to develop integrated software stacks for situation understanding. The first step to reach this objective is to identify the relevant data coming from the knowledge base and to attach monitors to the relevant properties to detect deviations. Then, when the software stack responsible for situation understanding is triggered, the situation analysis identifies the cause of the anomaly to generate the appropriate contingency plan.
- The development of a software for learning new strategies to deal with anomalies based on previous experiences. To reach this objective, further formalization of situation models and learning techniques need to be investigated.

In this paper, we give an overview of the challenges and of the current research we carry out on these aspects in the EU-funded project CONVINCE[3].

2 Challenges

2.1 Challenge 1 – A Tool for Inter-component Knowledge Exchange Based on Ontologies and Knowledge Graphs

The first challenge is the representation of the state of the environment, including the robot, the objects, the humans and the interactions between them in a consistent and common knowledge base for all the system components i.e., from verification to monitoring and planning. The knowledge base will then serve as a basis for analyzing situations in case of a deviation from the initial plan. Ontologies are commonly used to express knowledge as a set of concepts within a domain and the relations between them. In the literature, we can find several ontology-based knowledge representations for robotic systems [1, 11]. Some of them are related to a specific domain like manipulation [12]. Despite the various ontologies available in the literature, there are aspects that make the choice of a particular ontology challenging:

- The lack of consensus about how to model an entity. For instance, in the case of manipulation ontologies, affordance is defined as the possibility of an agent to perform actions with an object. In their survey about affordances in robotic tasks, authors of [10] highlight the ambiguity of the usage of the affordance concept in robotics tasks. Affordance can be modeled in different ways: as properties of an object, as a pair object-agent, as events, etc.

[3] https://convince-project.eu/.

- The type of ontology. In [1], authors distinguish three types: upper-level, reference and domain ontologies. Upper-level ontologies focus on widely applicable concepts like object, event, state, quality, and high-level relations like part hood, constitution, participation, dependence. On the other hand, reference ontologies focus on a discipline with the goal of fixing the general terms in it (e.g., medical, engineering, enterprise) and application ontologies contain all the definitions needed to model the knowledge required for a particular application.
- The relevance of entities to represent in the knowledge base. In general, it seems trivial to represent situation entities until we meet situations where we need to formalize properties to deduce a new knowledge. For instance, when encountering an unknown situation, we have to collect all possible data about the state of the environment and usually we do not know in advance what piece of information we need.

For these reasons, we cannot choose a particular ontology from the literature. Therefore, we start by building our own application ontology to model appropriate projects' use cases, like CONVINCE project, to then harmonize the developed concepts with respect to the literature.

2.2 Challenge 2 - Situation Understanding and Scene Perception

The second challenge is the implementation of components and algorithms for scene perception and situation understanding. These components will provide the robots with cognitive capabilities to detect and interpret discrepancies between predictions and observations (i.e., anomalies) and to reason regarding their symptoms and possible root causes (due to software faults, noise in the perception system, or erroneous assumptions, e.g., there are no blocks in some assembly tray), and decide about contingency actions to mitigate the anomaly. While scene perception can rely on state-of-the-art perception functions e.g., object/defect detection and pose estimation, or human detection to feed the knowledge base and the digital twin of the environment, the understanding of a particular situation when an anomaly occurs is more complex. We observed that we could distinguish two families of anomalies: known and unknown.

Known Anomalies refer to situations that are detected by the robot software and managed by generating appropriate contingency behaviors. These anomalies are considered known at design time, and mitigated by appropriate strategies.

Unknown Anomalies refer to situations that violate characteristic properties of the nominal operation and that run-time monitors detect as deviations for the expected behavior, but for which the designer does not provide a mitigation.

A major challenge in this case is to be able to generate mitigation plans for unknown anomalies. In CONVINCE an investigated approach is to start by ascertaining that the detected anomaly is really part of the unknown class, based on an extracted description of the anomaly given by the perception data and the knowledge base. If the anomaly is part of the known classes, the associated mitigation strategy is applied. Either way, an active plan procedure is launched to try finding a recovery strategy. Active plan tries several recovery strategies based on current knowledge base and similarities of the unknown anomaly case to known anomalies. Active plan also contributes into updating the current knowledge base to enrich the classification process and avoid falling into false unknown cases. Therefore, when an unknown anomaly case is encountered, active planning will take the upper hand to generate appropriate mitigation strategies.

An example with an assembly use case is while the robotic system is at picking station to pick a specific block and then monitors rise a flag of external property violation; in other words, the picking process was not successful. This anomaly detection launches the identification process: is the cause known or unknown? A known cause is *the block fell*. The knowledge base and the perception can give the information '*the block is not in gripper'*, with proprioception for instance, which orients the classification to known class *the block fell*. To recover the robot needs to pick the block again. On another hand, the description of the anomaly could be new and leads to an unknown classification. In this case, active plan will use current knowledge base to generate an effective recovery plan.

2.3 Challenge 3 – Learning New Situation Models

The third challenge is learning new models for situations that were not foreseen during design. A major focus will be on learning those representations that encode contingency strategies, or the knowledge to plan them based on the state of the world. In CONVINCE we will apply state-of-the-art learning techniques to learn models of situations, including descriptive models for planning and operational models for acting (e.g., based on the Behavior Tree formulation).

Another lead is by extending the situation understanding above described procedure. By generating new recovery strategies, it is possible to extract new anomalies descriptions. These generated (*Anomaly, Recovery)* couples are added to a database of known anomalies as new classes. Therefore, if the robotic system encounters the situation again it now knows how to recover from it. This means that the robot learnt a new situation model.

3 Outlook

In this paper, we introduced the objectives and the challenges for our research on situation understanding and scene representation. In CONVINCE we apply our research to three different use cases: a vacuum-cleaner robot, a museum robot guide and an industrial assembly robot. Most of the tools presented above will first be demonstrated in the industrial assembly domain, where a robot has to assemble refractory blocks on a cart with demanding constraints of precision and efficiency. Figure 1 presents the mock-up used.

The tooling we are developing to model situations and the associated algorithms that support it are also required to work in the other use-cases and more generally to be reused in as many intelligent-robotic domains as possible.

Fig. 1. Left: robot mobile base (arm is missing) and mock-up assembly; right: part of the expected assembly.

Acknowledgements. This work was supported by the European Union's Horizon Europe Research & Innovation Program under Grant 101070227 (CONVINCE). The industrial assembly robot use case is provided by CLEIA company.

References

1. Olivares-Alarcos, A., et al.: A review and comparison of ontology-based approaches to robot autonomy – ADDENDUM. Knowl. Eng. Rev. **35**, e29 (2020)
2. Matheus, A.J., et al.: A core ontology for situation awareness. In: Int. Conf. of Inf. Fusion, 545–552 (2003)
3. Dhouib, S., et al.: Control architecture concepts and properties of an ontology devoted to exchanges in mobile robotics. In: Nat. Conf. on Control Architectures of Rob. Grenoble, France (2011)
4. Farber, M., et al.: AWARE: an ontology for situational awareness of autonomous vehicles in manufacturing. In: Commonsense Knowledge Graph Workshop (CSKG'21) (2021)
5. Weiss, G., et al.: Safe adaptation for reliable and energy-efficient E/E architectures. In: Watzenig, D., Brandstätter, B. (eds.) Comprehensive Energy Management - Safe Adaptation, Predictive Control and Thermal Management, pp. 1–18. Springer International Publishing, Cham (2018)
6. Hussein, M., et al.: Safe adaptation of vehicle software systems. Microprocess. Microsyst. **52**, 272–286 (2017)
7. Ghezala, M.W.B., et al.: RSAW: A situation awareness system for autonomous robots. In: ICARCV, pp. 450–455 (2014)
8. Bohez, S., et al.: Sensor fusion for robot control through deep reinforcement learning. In: IEEE Int. Conf. Intell. Robots Syst., pp. 2365–2370 (2017)
9. Wang, Z. et al.: Time series classification from scratch with deep neural networks: a strong baseline. Int. Joint Conf. on Neural Netw. (IJCNN), 1578–1585 (2017)
10. Ardón, P.P.: Affordances in robotic tasks - a survey (2020). arXiv preprint arXiv:2004.07400
11. David Paulius, Y.S.: A survey of knowledge representation in service robotics. Robot. Auton. Syst. **118**, 13–30 (2019)
12. Wörgötter, F., Aksoy, E.E.: A simple ontology of manipulation actions based on hand-object relations. IEEE Trans. Autonom. Mental Develop. **5**(2), 117–134 (2013). https://doi.org/10.1109/TAMD.2012.2232291

Modeling Robot Control Architectures for Verification and Monitoring

Stefano Bernagozzi[1,2], Angelo Ferrando[1], Enrico Ghiorzi[1,2], Lorenzo Natale[2], and Armando Tacchella[1(✉)]

[1] Università degli Studi di Genova, Genoa, Italy
{angelo.ferrando,armando.tacchella}@unige.it, enrico.ghiorzi@edu.unige.it
[2] Fondazione Istituto Italiano di Tecnologia, Genoa, Italy
{stefano.bernagozzi,lorenzo.natale}@iit.it

Abstract. Model-based techniques are gaining widespread adoption in Robotics for their potential in easing the production and maintenance of complex control software. One of their advantages is that the availability of models and (formal) system requirements enables automated verification of system properties at design time and the generation of monitors to guard against unwanted behaviors at operation time. To guarantee consistency between the developers' work and the verification and monitoring results, it is mandatory to endow models with formal semantics, where the main challenge lies in striking a reasonable compromise between expressiveness and complexity of verification and monitoring.

Keywords: control architectures · verification · monitoring

1 Operational Models of Control Software

Contributions to formal verification from the robotics community often focuses on some form of verification for autonomous components. In the EU project CONVINCE we aim to formalize the overall architecture of a robot, as addressing individual components only would miss the point of integrating verification into the development lifecycle of a robot. In this paper we present our approach to formalizing the overall architecture of a robot and the initial efforts to endow models with formal semantics.

As mentioned in [7] (Chapter 12, Robotic Systems Architectures and Programming), the control software of a robot is most often organized in layers arranged in a stack of increasing abstraction from the hardware. Here, we consider a three layered architecture organized as follows—terminology from [2]:

- The *control layer* includes the embedded software for actuation and sensing, developed in close connection with the hardware it operates.
- The *functional layer* is made by software components (sometimes also called services) that provide functionalities to interface with the control layer (e.g., drivers), as well as processing units that rely on data acquired through drivers or other components to perform specific tasks (e.g., inverse kinematics, navigation, and object detection).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 191–195, 2024.
https://doi.org/10.1007/978-3-031-76428-8_36

– The *deliberative layer* coordinates the functional layer to perform specific activities (Tasks) and obtain specific objectives (e.g., fetch a glass of water, tidy up a room, fill a crate with specific components).

The control layer interacts directly with the environment which includes the physical robot and the actual environment in which it operates. Usually, the components of the different layers communicate among them through a middleware, e.g., ROS or YARP, that provides basic functionality for distributing computation among different modules, possibly across etherogeneous hardware.

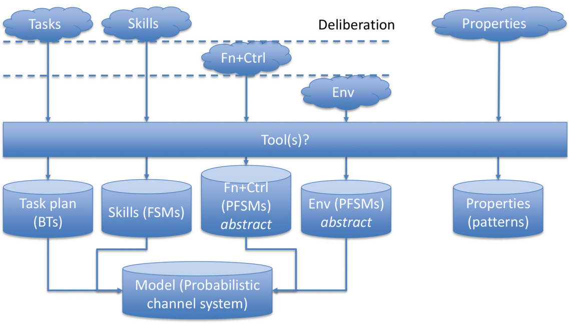


Fig. 1. From mental models and system properties to formal models and specifications.

Figure 1 represents the modeling process for the components of the three layers, as well as the environment and the required properties. Developers start from mental models, comprising the following:

– *Task plan* (policy in [6]): The highest-level form of structured planning of the robot's behavior. We assume that the task plan is initially represented as a Behavior Tree (BT) whose leaves are actions which could be subject to refinement. An executable task plan is a hierarchy of BTs encoded in a shared format where all the leaves correspond to skills.
– *Skills*: Functionalities obtained by orchestrating and configuring components. They are modeled as Finite State Machines (FSMs) and encoded in a shared format.
– *Functional and control* (commands and platform in [6]): The actual code running the system. Since it is out of the scope of our contribution to define models at the programming language level, an abstract model has to be provided here, possibly including probabilities (PFSM stands for Probabilistic FSM).
– *Environment*: Everything in the physical space where the system operates. Most of the times, it is not possible to provide a concrete and accurate model of the environment, therefore also in this case an abstract model has to be provided as a PFSM.

– *Properties*: propositions that the system has to satisfy, expressed using some temporal logic.

Note that, wherever an abstraction is employed, a reality gap is introduced. In other words, the abstract models of the running system and environment (see above) must correspond accurately to their real-world counterparts. This reality gap is typically addressed through monitoring [3,5], wherein a monitor is deployed to validate the abstractions against the actual systems.

The main idea is that developers are free to use one or more tools to formalize each mental model and encode it in a suitable format. The properties can be drafted either directly in logic or, more conveniently, using some kind of controlled natural language pattern [4]. Since all the elements interact with each other, we should also define the interaction between different components. We call the ensemble of models shown in Fig. 1 the *Operational Model* of a robot.

2 Designing Model Semantics

The system we model is a robot together with its control software at various layers and its hardware interacting with an external environment. The scenario is intrinsically *dynamic* and we postulate that it is also *time invariant*, i.e., the parameters governing the overall system behavior do not change significantly over time and random processes are characterized by stationary distributions. We expect that *nonlinear* models will have to be considered in order to take into account all kinds of interactions between the robot and the environment, but we restrict ourselves to *discrete time* models, i.e., in all cases where time needs to be considered explicitly (see below), we assume that it does so in discrete—albeit possibly very small—steps. Other elements for which we may consider alternatives are:

Continuous-state vs. Discrete-state: we expect that a combination of continuous and discrete quantities might characterize the overall state of the system; whenever possible, we will prefer discrete and finite domains for variables in order to preserve computational tractability of, e.g., algorithmic model verification.

Time-driven vs. Event-driven: In time-driven systems, the state changes as time changes. The models of physical interactions, e.g., between the robot and the environment, are typically modeled in a time-driven fashion. In event-driven systems, it is only the occurrence of asynchronously generated discrete events that forces instantaneous state transitions. The models of software interactions, e.g., among different control software components, are usually best modeled as such. Since we deal with combinations of software, hardware and environment, we expect that *hybrid models*, featuring both time-driven and event-driven models, will have to be considered.

Deterministic vs. Stochastic evolution: A model is stochastic whenever one or more of its parameters are random variables. In this case, the state of the system is described by a stochastic process, and a probabilistic framework is required to characterize the model behavior.

3 A Simple Case Study and Future Work

We now present a simple example to illustrate how the approach described in this paper can be used for modelling a complete system. We consider a simple system consisting of a stationary robot with a battery. The battery discharges with time, and an alarm has to be raised if the battery drops below a certain threshold.

In Fig. 2 we present the elements of the control architecture and the relationships among them. A short description of each element follows.

- *Task plan*: A BT encodes the behavior of the robot; the robot checks whether the battery level percentage is above the critical threshold. If so, an alarm is raised; otherwise, it performs an activity (for simplicity represented in this example as a generic action that discharge the battery).
- *Skills*: The skills used in the system are the battery level checker, that checks whether the battery level percentage is above the critical threshold; an alarm that notifies the user; and a battery drainer that waits for the battery to discharge by one percentage point.
- *Functional and control*: It includes the component that reads the value from the battery and publishes it on, for example, a ROS topic.
- *Environment*: It includes the physical robot, its battery, and a nearby user (not shown in Fig. 2).
- *Properties*: We want to monitor that the alarm is always raised when the battery drops below the threshold. This is a typical *safety* property ("nothing bad will happen").

In this example, a monitor will be generated automatically from the safety property. The monitor is a simple "watchdog" ensuring that the battery level never drops below a given threshold. In general, using (timed) temporal logic we are able to express more complex properties including *liveness* ("something good will happen") and include deadlines, e.g., a component will always respond to a request within a given time horizon. In the case of monitoring, the (timed) trace corresponding to system execution will be extracted from the robot, which in turn implies sniffing the messages that are exchanged among various components—see, e.g., [5]. In the case of offline verification, we will follow a falsification approach based on stochastic model checking techniques. In a few words, we will sample timed traces by executing the *models* of the components and simulating their interactions—technically, this amounts to simulate a channel systems with (a)synchronous communications [1]. From these traces, we will be able to extract either counterexamples to the required properties, or statements about the probability of their satisfaction, possibly together with some confidence bound.

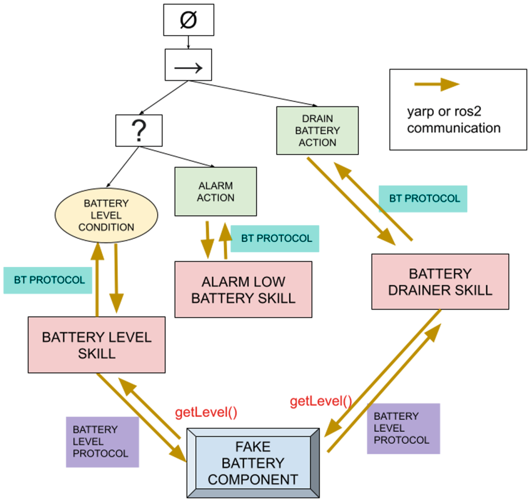


Fig. 2. A simple case study for the CONVINCE architecture.

Acknowledgements. The work was supported by the European Union's Horizon Europe Research & Innovation Program under Grant 101070227 (CONVINCE).

References

1. Baier, C., Katoen, J.P.: Principles of Model Checking. MIT Press (2008)
2. Colledanchise, M., Cicala, G., Domenichelli, D.E., Natale, L., Tacchella, A.: Formalizing the execution context of behavior trees for runtime verification of deliberative policies. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2021, Prague, Czech Republic, 27 September–1 October 2021, pp. 9841–9848. IEEE (2021)
3. Desai, A., Dreossi, T., Seshia, S.A.: Combining model checking and runtime verification for safe robotics. In: Lahiri, S., Reger, G. (eds.) RV 2017. LNCS, vol. 10548, pp. 172–189. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67531-2_11
4. Dwyer, M.B., Avrunin, G.S., Corbett, J.C.: Patterns in property specifications for finite-state verification. In: Proceedings of the 21st International Conference on Software Engineering, ICSE 1999, pp. 411–420. Association for Computing Machinery, New York (1999). https://doi.org/10.1145/302405.302672
5. Ferrando, A., Dennis, L.A., Cardoso, R.C., Fisher, M., Ancona, D., Mascardi, V.: Toward a holistic approach to verification and validation of autonomous cognitive systems. ACM Trans. Softw. Eng. Methodol. **30**(4), 43:1–43:43 (2021). https://doi.org/10.1145/3447246
6. Ingrand, F., Ghallab, M.: Deliberation for autonomous robots: a survey. Artif. Intell. **247**, 10–44 (2017)
7. Siciliano, B., Khatib, O. (eds.): Springer Handbook of Robotics. Springer, Cham (2016)

Intuitive Telemanipulation of DLOs Via Vision-Based Shared Control: A Pilot Study

Davide Chiaravalli(✉), Alessio Caporali, Anna Friz, Roberto Meattini, and Gianluca Palli

DEI - Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy
davide.chiaravalli2@unibo.it

Abstract. The handling of Deformable Linear Objects (DLOs) represents a crucial process where the integration of automation and autonomous systems remains limited. This paper introduces a teleoperation framework designed for the intuitive manipulation of DLOs with the assistance of visual cues. The proposed system holds potential for applications in hazardous scenarios and robot teaching tasks involving DLOs. A set of experiments involving multiple subjects were conducted. The results demonstrate that the suggested teleoperation framework simplifies the manipulation of DLOs, resulting in an average 20% reduction in task completion time.

Keywords: Shared Autonomy · Teleoperation · Deformable Linear Objects

1 Introduction

In the last decade, in the context of developing advanced interaction capabilities and flexibility in industrial control processes, great focus has been given to the modeling and manipulation of Deformable Linear Objects (DLOs). The manipulation of a DLO is performed by tracking the objects either exploiting a mechanical deformation model [1] or continuously estimating its shape through a dedicated vision system [2]. Given the complexity of the problem, several research studies have addressed DLO manipulation with a task-oriented strategy by either designing specific tools [3] or by directly defining separate modules to handle the different operations required [4]. Nonetheless, they lack the flexibility to allow task generalization in different scenarios. To this purpose,

This work was supported by the European Commission's Horizon Europe Framework Programme with the Project IntelliMan under Grant 101070136 and in part by MUR with the Project "Sustainable Mobility Center" under Grant CN00000023-CUP J33C22001120001.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 196–200, 2024.
https://doi.org/10.1007/978-3-031-76428-8_37

teleoperation still plays a central role for its capability to merge together the intuivity and flexibility of a human operator, and the autonomy of a robotic system. In this work, we illustrate a teleoperation framework for intuitive manipulation of DLOs by means of a robotic manipulator equipped with a vision system for DLO detection.

2 Methods

2.1 Perception and Modeling

The perception of the DLO is achieved by employing the algorithm named *FASTDLO* [5], which provides as output the representation of a generic DLO shape in the image space by a 3rd-order spline basis as a function of a free coordinate u representing the position along the cable starting from an endpoint ($u = 0$) to the opposite end ($u = L$) being L the length of the DLO (please refer to [5] for more details.) Afterward, the spline is evaluated in a number n of points and each of the obtained points is projected in the 3D cartesian space.

2.2 Teleoperation

The system alternates between two main control modes: *free teleoperation* and *cable targeting modes*. In *free teleoperation mode* the remote manipulator's behavior is directly controlled through the operator's motions. In *cable targeting mode* the system performs a scan of the environment to search for existing cables and by exploiting the points of the spline detected, forces the remote manipulator to move along the cable main dimension only. The operator during the task can freely switch between the two control modes.

Free Teleoperation Mode. In this control mode, the operator fully controls the remote manipulator motion. An edge drifting mapping technique is employed to efficiently map the whole task environment: given the haptic stylus and the robot end-effector positions $P_h(t) \in \mathbb{R}^3$ and $P_r(t) \in \mathbb{R}^3$ respectively, a virtual spherical region or radius r_h centered in the zero coordinate of the haptic workspace is defined. Inside the spherical region a position control mapping is applied such that the remote manipulator exactly replicates scaled motions of the operator in the local workspace. Outside the spherical region, a rate control mapping is applied, where the velocity of the manipulator $V_h(t) \in \mathbb{R}^3$ is controlled proportionally to the distance from the sphere. An elastic-like force, proportional to the distance from the position control area, is produced on the haptic device when moving in the rate control region to provide the operator with a clear perception of the commanded velocity. Eventually, the grasping action is activated by a specific button on the haptic stylus.

Cable Targeting Mode. Upon the operator's request, the system can enter in *cable targeting mode*. The vision system mounted on the robot end-effector acquires an image of the environment and localizes the DLOs. Then, the closest DLO is identified, and the robot motion is constrained along its main direction. In this mode, the haptic workspace is directly projected on the DLO surface so that any operator motion would result in a manipulator displacement along the object. Along the vertical direction, the edge drifting mapping is kept active to allow a quick motion to the correct height. Moreover, the manipulator end-effector is naturally oriented perpendicularly to the DLO main direction at the current point, guaranteeing an effective grasp position for the gripper. The operator's perception is enhanced by a virtual fixture generation through haptic feedback of the DLO shape. A potential map generating an attractive force F_{vf} toward the cable shape is generated as a function of the distance from the DLO according to an arctangent behavior

$$F_{vf} = -(\frac{1}{2} + \mathrm{atan}(d(q^*, P_h)K_{vf})) \tag{1}$$

with K_{vf} characterizing the gradient of the force moving toward the DLO.

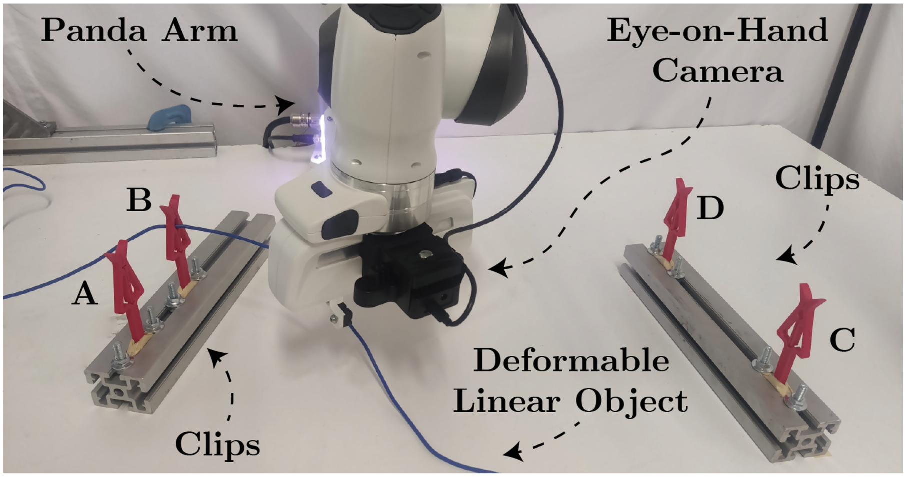


Fig. 1. Experimental setup.

3 Experiment and Results

The proposed teleoperation framework was evaluated using a Franka Emika Panda robot and a 3DSystems Geomagic haptic device. Visual perception of the DLO was achieved through 2D RGB images of an eye-in-hand camera, see Fig. 1.

Experimental Protocol. Ten healthy subjects, with no prior knowledge of the setup (2 females, 8 males; age: 29 ± 4; right-handed: 9 subjects, left-handed: 1 subject), participated in the test. The experiments adhered to the principles of the Declaration of Helsinki, with participants receiving thorough information about the experimental protocol and providing informed consent. The experimental protocol involved each subject performing manipulation tests, requiring multiple insertions of a cable into specific clips fixed in the environment, i.e. DLO routing task (see Fig. 1). The operator was instructed to move the robot to grasp specific points along the DLO shape and perform insertions in designated clips. The tests were conducted for both the *free teleoperation mode* and the *cable targeting mode*. Three sequences were performed for each mode, with different clip couplings selected. Before the tests, subjects received five minutes of initial training to familiarize themselves with the system. The overall task time for each sequence was measured to assess the effectiveness of the proposed approach.

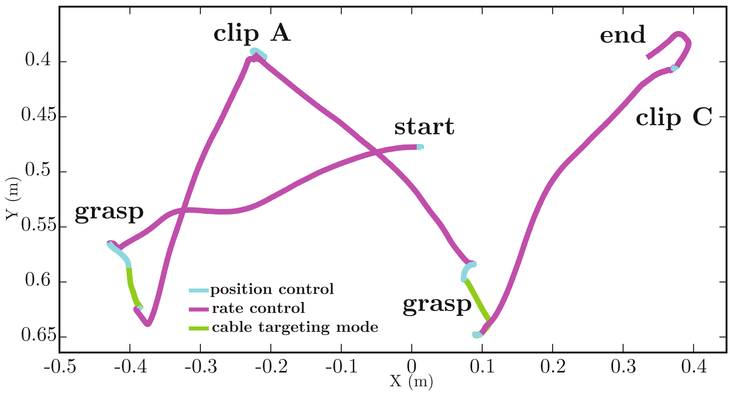


Fig. 2. Single trial experiment.

In Fig. 2, a comprehensive visualization of the remote manipulator's motion in the horizontal plane is displayed, concerning a single subject trial involving clips A and C. The graph illustrates that the majority of the motion involves exploiting rate control to move the spherical position control area across the environment. Position control is primarily utilized during high-precision tasks such as clip insertions or when traversing for grasping operations. Notably, the *cable targeting mode* is employed during the precise grasp of the DLO. The comprehensive results for the test, comparing all test subjects, are presented in Fig. 3. The analysis in Fig. 3 focuses on the Task Completion Time (TCT) metric to assess the performance of subjects with and without the assistance of the proposed shared control approach.

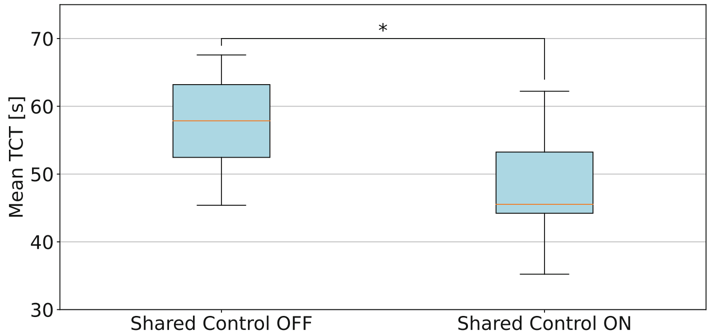


Fig. 3. Boxplot of the subjects' mean Task Completion Times (TCTs) with and without the shared control aid ('*' indicates statistical significance).

4 Conclusions

In this work a shared control framework aiming at enhancing the efficiency and user-friendliness of DLO telemanipulation tasks has been illustrated. The usage of the camera system for reconstructing the DLO shape simplifies the grasping process. An experimental session involving ten subjects demonstrated the effectiveness of the proposed framework in decreasing operator mental effort, significantly reducing task completion time. Future work will concentrate on extending the framework to cluttered environments with partially obscured cables and enhancing the vision system to achieve real-time 3D detection during operator motions.

References

1. Jin, S., Wang, C., Tomizuka, M.: Robust deformation model approximation for robotic cable manipulation. In: Proceedings of IEEE/RSJ International Conference on IROS. IEEE (2019)
2. Mou, F., Ren, H., Wang, B., Wu, D.: Pose estimation and robotic insertion tasks based on yolo and layout features. Eng. Appl. Artif. Intell. **114**, 105164 (2022)
3. She, Y., Wang, S., Dong, S., Sunil, N., Rodriguez, A., Adelson, E.: Cable manipulation with a tactile-reactive gripper. Int. J. Robot. Res. **40**(12–14), 1385–1401 (2021)
4. Gao, Y., Chen, Z., Lin, J., Li, X., Liu, Y.-H.: Development of an automated system for the soldering of USB cables. Robot. Comput.-Integr. Manuf. **79**, 102440 (2023)
5. Caporali, A., Galassi, K., Zanella, R., Palli, G.: Fastdlo: fast deformable linear objects instance segmentation. Robot. Autom. Lett. **7**(4), 9075–9082 (2022)

Sparse Optical Sampling in the Close Proximity of a Robotic Arm

Martin Laurenzis[1(✉)], Ante Marić[2], Emmanuel Bacher[1], Mateusz Pietrzak[1], Stéphane Schertzer[1], Francesco Grella[3], and Sylvain Calinon[2]

[1] French-German Research Institute of Saint-Louis (ISL), 68301 Saint-Louis, France
martin.laurenzis@isl.eu
[2] Idiap Research Institute, 1920 Martigny, Switzerland
[3] DIBRIS, University of Genoa, 16145 Genoa, Italy
https://www.isl.eu

Abstract. Close collaboration between humans and robots needs a sensing infrastructure to monitor the robot environment and secure human-robot interaction. In this context, we investigate sparse optical range sampling using a distributed network of robot mounted Time-of-Flight (ToF) sensors. We present an evaluation of sensor candidates, provide experimental characterization of an early prototype and show strategies for environment modeling and object reconstruction.

Keywords: sensor arrays and networks · robotic environment · proximity sensors · signed distance fields

1 Introduction

The effectiveness of Human-Robot Collaboration (HRC) strongly depends on the perceptual capabilities of the involved robotic system. To improve task performances and overall safety, a sensing architecture containing external (E1, E2) and robotic/robot-mounted (R1, R2, R3) sensors can enhance the robotic awareness, see Fig. 1(a). But, as shown in Fig. 1(b), external sensors can suffer from occlusions and blind spots as well as from the need of a perfect calibrated environment. Especially in close proximity of robot and human, the occlusion of the line of sight can have critical consequences. Alternatively, industrial robots can be equipped with additional sensors that are integrated directly into the robotic platform (Fig. 1(c)).

A distributed network of sensors can cover the robot body and allow the proximity of the robot to be analyzed. In the current paper we neglect tactile sensing and investigate a network of distributed laser ranging sensors (LIDAR), as in [1–3]. The presented work was carried out within the SESTOSENSTO project (HORIZON-CL4-Digital-Emerging Grant 101070310). This work reports the collaborative work of multiple partners within the consortium. Further reading can be found online [4].

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 201–205, 2024.
https://doi.org/10.1007/978-3-031-76428-8_38

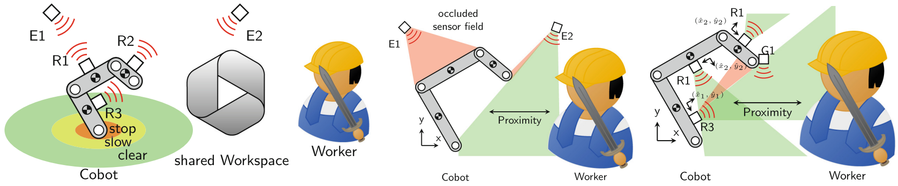


Fig. 1. Safe collaboration of human and robot in a production line can be realized, for instance, by (a) a Cobot equiped with (b) external ($E1, E2 \dots$) or (c) robot mounted ($R1, R2 \dots$) sensors.

Table 1. Non exhaustive list of optical time-of-flight proximity sensors with light detection and ranging (LIDAR) on integrated circuits (LoIC).

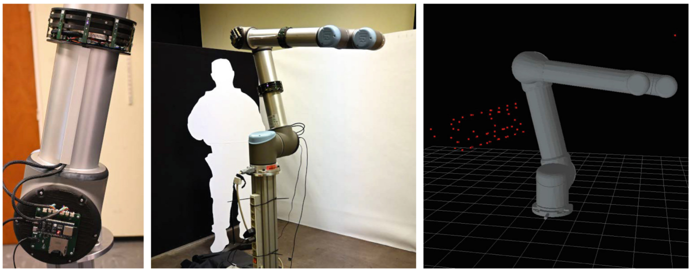
Role	Manufacturer Name	Model No.	Package Type	ToF	Sensor Type	Resol.	Range	Laser Type	λ [nm]
Proximity	STMicro.	VL53L8	LoC	pulsed	SPAD	8×8	4 m	on chip	940
	OSRAM	TMF8828	LoC	pulsed	SPAD	8×8	4 m	on chip	940
	SHARP	GP2APx	LoC	pulsed	SPAD	1×1	1.2 m	on chip	940
	OnSemi	MicroFC	Sensor	pulsed	SiPM	1×1	10 m	sep.	905

2 Optical Range Sampling in Robot Proximity

A task-oriented evaluation of optical range imaging sensors must be based on task-related constrains and, in addition to their physical performance, consider a hypothetic cost function $S_{\text{SWAP-C}}(x)$, that evaluates **S**ize, **W**eight **A**nd electrical **P**ower consumption as well as the sensor's **C**osts (SWAP-C). Here, the sensor's costs cover not only the economic trade value but incorporates also computational cost: data transfer loads and processing efforts.

Proximity monitoring has to cover the environment all around the robot within a hemisphere from very close (few centimeter) to, at least, the maximum range of the robotic arm. In this case, the optical resolution is initially of secondary importance, as it is only used for a rudimentary environmental model that is intended to distinguish open movement areas from fixed installations. In addition, and more important due to work safety issues, spatial coverage has to ensure reliable localization of human employees. Therefore, we decided to use a network of distributed sensors to provide complete coverage of the environment and avoid shadowing of areas by the robot itself.

Optical time-of-flight sensors are available in a compact size as LIDAR on integrated circuits (LoIC). In Table 1 we provide an overview of a selection of LoIC candidates (green highlighted parameters meet requirements).



(a) (b) (c)

Fig. 2. Sensor network consisting of a bracelet with (a) 10 distributed LIDAR sensors and a micro-controller unit (MCU). The sensor is mounted on a robotic arm (b). The data of the robotic environment can be displayed (c) as a sparse point cloud.

3 Experimental Evaluation

We have built an experimental setup to evaluate different sensors for proximity monitoring. The sensors were mounted on a robotic arm (UR10, Universal Robotics) and used in a synthetic scene consisting of a human silhouette and walls with light (white paper) and dark (black fabric) surfaces.

As illustrated in Fig. 2(a), we set up networks of LoIC sensors in the form of a bracelet or waistband. Each unit consisted of 10 LoIC sensors and a micro-controller unit (MCU, Espressif, EPS32). The MCU is connected to the individual LoICs to control the automated data acquisition and to read out data. The bracelets were mounted on the individual robotic arm segments as illustrated in Fig. 2(b). The recorded data is transmitted through a wireless network connection to a main processing unit (PC). High level data processing such as fusion and display of point clouds (Fig. 2(c)), building an environment model (e.g. SLAM), the reconstruction of objects (Sect. 4), control and adaption of the robot motion as well as human robot interaction (HRI) will be realized on the main processing unit.

In our experiments, we used the LoICs TMF8828 (ams OSRAM) and VL53L8 (ST Microelectronics), respectively. Each individual sensor measures a point cloud of 8×8 range values in a field of view of $40° \times 60°$. Thus, the group of 10 sensors cover a disk shaped area ($360° \times 40°$) perpendicular to the arm segment. In this area, we were able to monitor the proximity of the robotic arm from close range (ca. 2 cm) to a maximal distance of about 3 m with a point cloud update rate of up to 5 Hz. The maximum range is impacted by the ambient light level which can alter the signal to noise level. Further the surface characteristics such as reflectance and orientation can effect the results. A detailed analysis is pending. Again, in our first approach we were able to cover a disk shaped area. In later application, the position (place and orientation) of the sensors mounted on the robot has to be optimized for maximal coverage and to comply with work task specific requirements.

4 Object Reconstruction

Sampled distance information can be used to learn signed distance fields (SDFs) in order to reason about making contact with the environment or objects of interest. For this purpose, piecewise polynomial basis functions can be leveraged as an underlying model to represent continuous and smooth SDFs, with direct access to gradients. These properties make such representations directly usable for guiding movement in robotic manipulation tasks [5].

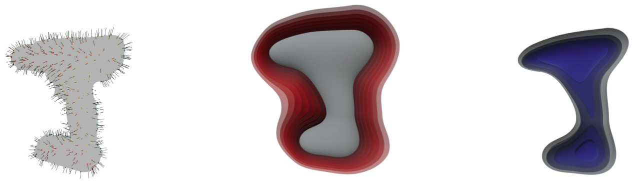


Fig. 3. Piecewise polynomial SDF of the `035_power_drill` object from the YCB dataset [6]. The ground truth mesh with 700 non-uniformly distributed training samples is shown on the left. The center and right image show the reconstructed mesh and level sets (positive in red and negative in blue).

Fitting a piecewise polynomial SDF model amounts to learning a number of basis function weights from sampled distance data. A simple way of doing this is through least squares regression. Learning only from surface points and normals requires additional regularization terms in order to provide valid representations of distance [7]. In an online setting, data collected in batches or point-by-point can be used to update an arbitrary prior model through an incremental formulation of least squares [8]. Figure 3 shows an example object reconstruction. The training procedure for a 2D case is illustrated in Fig. 4.

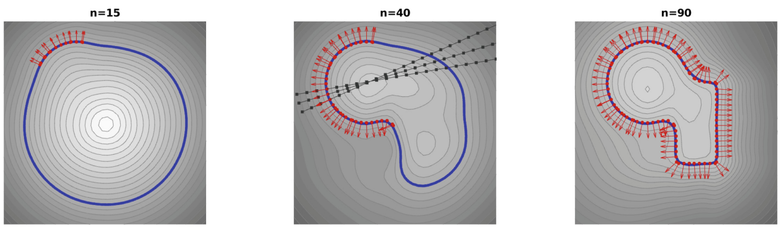


Fig. 4. Incremental learning of an SDF represented using piecewise polynomial basis functions, with level sets displayed as gray and blue contours. Starting from a spherical prior, the model is incrementally updated with n incoming samples, shown in red. Regularization is enforced on points shown in black.

5 Conclusions

We have evaluated different LoIC sensors to be used in a sensor network mounted on the robot skin. Our sensor networks can monitor the robot environment from close range several meters and sample a sparse point cloud which covers a disk shape area all around the robot arm without self occlusion or blind spots. In principle, these results comply with the requirements for a application in a safety and control system for human-robot-collaboration (HRC). Furthermore, we have investigated the reconstruction of object shapes from a small set of sample points. Our approach uses piecewise polynomial basis functions to implicitly represent shapes as signed distance fields (SDFs). Thus, without the need of a dense point cloud we are able to develop an environment model which can be used for further high level operations. Future developments foresee the design and implementation of a dedicated middleware architecture to provide efficient access to different proximity data representations. Moreover, the methods and technologies presented in this paper will be extended to real-world applications which are relevant for industrial manufacturing and agriculture scenarios.

References

1. Giovinazzo, F., Grella, F., Sartore, M., Adami, M., Galletti, R., Cannata, G.: From CySkin to ProxySKIN: design, implementation and testing of a multi-modal robotic skin for human-robot interaction. Sensors **24**(4), 1334 (2024)
2. Tsuji, S., Kohama, T.: Proximity skin sensor using time-of-flight sensor for human collaborative robot. IEEE Sens. J. **19**(14), 5859–5864 (2019)
3. Adamides, O.A., Modur, A.S., Kumar, S., Sahin, F.: A time-of-flight on-robot proximity sensing system to achieve human detection for collaborative robots. In: 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), pp. 1230–1236 (2019)
4. SestoSenso (2023). http://sestosenso.eu/
5. Li, Y., Zhang, Y., Razmjoo, A., Calinon, S.: Learning robot geometry as distance fields: applications to whole-body manipulation (2023)
6. Calli, B., et al.: Yale-CMU-Berkeley dataset for robotic manipulation research. Int. J. Robot. Res. **36**(3), 261–268 (2017)
7. Taubin, G.: Smooth signed distance surface reconstruction and applications. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 38–45 (2012)
8. Marić, A., Li, Y., Calinon, S.: Online learning of piecewise polynomial signed distance fields for manipulation tasks (2024)

The CONVINCE Perspective on Task and Motion Planning in Dynamic Environments

Masoumeh Mansouri(✉), Charlie Street, and Yassin Warsame

School of Computer Science, University of Birmingham, Birmingham, UK
m.mansouri@bham.ac.uk

Abstract. Combined task and motion planning (TAMP) has been extensively studied within robotics. This paper explores TAMP in dynamic environments in the context of the European Union Horizon project CONVINCE (context-aware verifiable and adaptive dynamic deliberation). CONVINCE aims to develop cognitive deliberation capabilities for safe and autonomous robot operation over extended periods. The technical contributions of CONVINCE, including solutions to TAMP, will be validated on three real-world use cases: a robot vacuum cleaner, an assembly robot, and a robotic museum guide. In this paper, we briefly explain our proposed TAMP strategies for these use cases.

Keywords: Task and motion planning · planning under uncertainty · active perception

1 Task and Motion Planning in Dynamic Environments

Combined task and motion planning (TAMP) is a well-established robotic problem, reaching a level of maturity in research that has resulted in several surveys, such as [1,6,7], and [10]. This paper focuses on TAMP in dynamic environments. The core challenge for both task and motion planning in complex, dynamic environments is the uncertainty which arises due to the unobservability and inherent stochasticity of the real world. This is addressed in the literature through approaches such as online planning, and planning under uncertainty [10]. Online planning techniques either assume offline plans are only preliminary until execution, at which points extra details are included [4]; or augment an offline plan with contingency plans to handle potential anomalies [18]. The main limitation of these approaches is their limited scalability, as computing all contingency plans is prohibitively expensive, even if the sources of all anomalies are known.

Prior work on automated task planning in uncertain and dynamic environments has focused on aligning the belief-states of the robot (i.e. a distribution over possible robot states) with the symbolic planning process. For example, [17] apply hierarchical task networks (HTNs) to partially observable environments, and [3] use probabilistic representations to infer the most appropriate

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 206–210, 2024.
https://doi.org/10.1007/978-3-031-76428-8_39

plan from a pre-defined library of plans. Decision-theoretic planning methods based around Markov decision processes (MDPs) and partially observable MDPs (POMDPs) are the most prevalent approaches for tackling different sources of uncertainty [8,14,15]. However, despite their common usage in robotics, these approaches often suffer from poor scalability and inaccurate modelling [12]. Inaccurate models produce inefficient robot behaviour, as our expectations of robot behaviour during planning diverge from what we observe during execution. An alternative approach for handling uncertainty is to use reactive formal methods which provide guarantees over robot behaviour in uncertain or adversarial environments (e.g., within a receding horizon paradigm [11]). Reactive approaches have limited performance when long-term autonomy is required, as the robot cannot predict what will happen beyond the range of its sensors, restricting its ability to plan to account for anticipated dynamics beyond the range of the sensors, limiting possibilities for reordering the sequence of task execution to achieve a goal. Finally, verification-based methods have been used to account for limited forms of uncertainty, e.g., delays and measurement errors in TAMP [9].

2 TAMP in CONVINCE

CONVINCE (context-aware verifiable and adaptive dynamic deliberation) is a project funded by the European Union Horizon Europe Programme. The main contribution is to develop and verify cognitive deliberation capabilities that ensure safe robot operation over extended periods without human intervention, and integrate these capabilities into a model-driven software toolchain for robot developers. CONVINCE will demonstrate the techniques developed in the project on three real-world use cases:

- **(UC1) Robot vacuum cleaner:** Here, a robot vacuum cleaner with limited capabilities must operate in diverse and dynamic environments which are impossible to know at design-time. End-users require vacuum cleaners that improve with time, i.e. learn which areas are problematic to clean and where objects are commonly placed to avoid getting stuck.
- **(UC2) Assembly robot:** Here, two manipulators undertake an assembly task. Parts are available in the environment in no prescribed order and should be assembled based on their shape. Occlusions or difficult lighting conditions may impair the visual system, and some parts may be missing, occluded, or restrained by other parts.
- **(UC3) Robotic museum guide:** Here, a humanoid robot must guide visitors inside a museum and describe the artworks. Part of the computation is performed on-board and part off-board. The robot's behaviour should change depending on the currently available functionality (i.e. if a network connection is available to utilise off-board computation), the status of the environment, and the behaviour of humans.

Each use case requires a TAMP solution. Although each use case is different in terms of complexity and requirements, each faces unique challenges. For instance, in UC2 a robot may place an object such as to block access to another object required for the task. In UC3, task plans (i.e. the order of visited artworks) and motions may be affected by an unexpected flow of visitors that is difficult to predict and model.

UC1 and UC3 require solutions for coverage planning in dynamic and unstructured environments. In UC1, the robot must autonomously clean the entire floor, where some areas may be temporarily inaccessible due to the presence of an object such as a toy. These difficult to predict dynamics make complete coverage impossible. For UC3, we face a similar problem, as a high concentration of visitors around an artwork may make it temporarily inaccessible. This may cause the robot to skip a piece of artwork, resulting in incomplete coverage. To solve coverage in these environments we must address two problems: the order areas of interest are visited, e.g. artworks or floor regions (task planning); and how, when, and where the robots move (motion planning). To ensure safe autonomy, the TAMP solutions we propose must account for the spatiotemporal dynamics of the environment, e.g. the flow of humans throughout a museum.

Recall that the objective of CONVINCE is to design a verifiable model-driven software toolchain for synthesising robot behaviour. To facilitate the integration of software into a toolchain, CONVINCE uses behaviour trees (BTs) as a high-level task specification provided by the end-users. BTs are a popular formalism for designing robot behaviour due to their flexibility, modularity, and reusability [5]. BTs are often hand-designed by domain experts. However, most robot environments have complex stochastic dynamics which affect robot performance and are hard for engineers to reason over. Therefore, in CONVINCE we refine hand-designed BTs by explicitly reasoning over these stochastic dynamics, where refinement restructures the BT and adds new nodes where necessary. This is an important aspect of TAMP in CONVINCE.

In CONVINCE, TAMP is also necessary to devise strategies for actively mitigating anomalies caused by perception failure. For example, we may need to synthesise a new trajectory which changes the robot's viewpoint when object detection fails due to the lighting conditions or object occlusion. In CONVINCE, situation understanding, execution monitoring, and active anomaly mitigation are tightly coupled. This opens up an integration challenge as each of these modules has different representations. In the next section, we briefly explain our strategies for addressing the challenges outlined so far.

3 Methodology

In CONVINCE we will develop combined TAMP techniques that incorporate multiple representations of uncertainty at different levels of abstraction. These representations may include semantic information, e.g. in UC1 toys may move but sofas will not; learned flow behaviours, e.g. in UC3 humans often move in a certain direction through the museum at a certain speed; safety standards; and long-term scene prediction. Learned representations of uncertainty improve TAMP by informing the search process, which allows us to synthesise robot behaviour that is robust to environmental dynamics. The TAMP methods we present will combine decision-theoretic approaches, formal methods, and constraint-based techniques such as to provide performance guarantees over the robot's behaviour.

3.1 Active Perception for Anomaly Detection and Recovery

In UC2, unknown anomalies may occur during the planning horizon, i.e. anomalies (e.g., faulty gripper, missing parts) that have not been accounted for beforehand. Our planner will identify the root cause of the anomalies by actively searching the workspace, updating our knowledge base of known anomalies, and using TAMP to recover from the failure. The TAMP solution for previously unknown anomalies will be translated back to a BT, enriching our models with the goal of achieving automatic recovery in the long run.

3.2 Coverage Planning in Dynamic Environments

Recall that in UC1, a vacuum cleaner with limited sensing must clean the floor each day by covering it with its brushes. Domestic environments are dynamic, where these dynamics are *a priori* unknown. In CONVINCE, we propose a *lifelong* approach to coverage planning, where a spatiotemporal dynamics model is learned over multiple days, or episodes, to improve coverage. This dynamics model can predict which locations will be occupied, and when [13]. With this, we employ online POMDP planning techniques [19] during each cleaning run which explicitly reason over the stochastic environment dynamics to efficiently cover the environment. After each cleaning run, we update our spatiotemporal dynamics model using the observations collected by the robot during coverage.

3.3 Behaviour Tree Refinement

BT refinement begins with an initial BT hand-designed by domain experts. To support refinement, we require formal stochastic models of the *skills* executed by the robot, which correspond to the action nodes of the BT. We obtain these models at design time from domain experts, or learn them from simulation. Using these skill models, we construct and solve a hierarchical Markov model which reasons over the skill durations and outcomes [16]. The solution to our Markov model is then converted back to a BT using existing techniques [2]. By refining the BT, we synthesise robot behaviour which is more robust to the probabilistic nature of the environment while exploiting expert domain knowledge.

Acknowledgments. This work was supported by the European Union's Horizon Europe Research Programme (grant #101070227) & UKRI (grant #10042096).

References

1. Antonyshyn, L., Silveira, J., Givigi, S., Marshall, J.: Multiple mobile robot task and motion planning: a survey. ACM Comput. Surv. **55**(10), 1–35 (2023)
2. Banerjee, B.: Autonomous acquisition of behavior trees for robot control. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3460–3467. IEEE (2018)
3. Beetz, M., et al.: Integrated, plan-based control of autonomous robot in human environments. IEEE Intell. Syst. **16**(5), 56–65 (2001)
4. Cacace, J., Caccavale, R., Finzi, A., Grieco, R.: Combining human guidance and structured task execution during physical human-robot collaboration. J. Intell. Manuf. **34**(7), 3053–3067 (2023)
5. Colledanchise, M., Ögren, P.: Behavior Trees in Robotics and AI: An Introduction. CRC Press (2018)
6. Garrett, C.R., et al.: Integrated task and motion planning. Ann. Rev. Control Robot. Auton. Syst. **4**, 265–293 (2021)
7. Guo, H., Wu, F., Qin, Y., Li, R., Li, K., Li, K.: Recent trends in task and motion planning for robotics: a survey. ACM Comput. Surv. (2023)
8. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artif. Intell. **101**(1–2), 99–134 (1998)
9. Liu, J., Ozay, N.: Finite abstractions with robustness margins for temporal logic-based control synthesis. Nonlinear Anal. Hybrid Syst. **22**, 1–15 (2016)
10. Mansouri, M., Pecora, F., Schüller, P.: Combining task and motion planning: challenges and guidelines. Front. Robot. AI **8**, 637888 (2021)
11. Raman, V., Donzé, A., Sadigh, D., Murray, R.M., Seshia, S.A.: Reactive synthesis from signal temporal logic specifications. In: Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control, pp. 239–248 (2015)
12. Russell, S.J., Norvig, P.: Artificial intelligence a modern approach. London (2010)
13. Saarinen, J., Andreasson, H., Lilienthal, A.J.: Independent Markov chain occupancy grid maps for representation of dynamic environment. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3489–3495. IEEE (2012)
14. Shah, N., Vasudevan, D.K., Kumar, K., Kamojjhala, P., Srivastava, S.: Anytime integrated task and motion policies for stochastic environments. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 9285–9291. IEEE (2020)
15. Şucan, I.A., Kavraki, L.E.: Accounting for uncertainty in simultaneous task and motion planning using task motion multigraphs. In: 2012 IEEE International Conference on Robotics and Automation, pp. 4822–4828. IEEE (2012)
16. Sutton, R.S., Precup, D., Singh, S.: Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning. Artif. Intell. **112**(1–2), 181–211 (1999)
17. Weser, M., Off, D., Zhang, J.: HTN robot planning in partially observable dynamic environments. In: 2010 IEEE International Conference on Robotics and Automation, pp. 1505–1510. IEEE (2010)
18. Yalciner, I.F., Nouman, A., Patoglu, V., Erdem, E.: Hybrid conditional planning using answer set programming. Theory Pract. Logic Program. **17**(5–6), 1027–1047 (2017)
19. Ye, N., Somani, A., Hsu, D., Lee, W.S.: DESPOT: online POMDP planning with regularization. J. Artif. Intell. Res. **58**, 231–266 (2017)

Towards Collaborative Grape Harvesting with a Mobile Manipulator

Edwin Pircher, Giovanni Carabin(✉), Marco Camurri, and Renato Vidoni

Libera Università di Bolzano, Piazza Università 1, 39100 Bolzano, BZ, Italy
{giovanni.carabin,marco.camurri,renato.vidoni}@unibz.it

Abstract. Modern agriculture is progressively integrating robotic systems for a wide variety of tasks. Guided by the agricultural use case of the Sestosenso EU project, we introduce a robotic system consisting of a manipulator with a custom-designed end-effector, positioned on top of a mobile robotic platform. The robot's primary function is to assist operators in hillside vineyards, facilitating grape harvesting. Key contributions include the development of a dedicated end-effector, implementation of computing hardware, and comprehensive robotic integration of the system. The work is completed with the final realisation and tests of the mobile robot prototype and assessment of the future work.

Keywords: Agriculture 4.0 · Agricultural Robotics · Collaborative Mobile Manipulator · Assistant Systems · Grape Harvesting

1 Introduction

The incorporation of automated robotic systems in precision agriculture offers opportunities to increase efficiency, sustainability, and productivity [5,16]. This transformative process comes with its own unique challenges in the different sectors of agriculture, such as fruit orchards and vineyards, as it has to combine advanced technologies and traditional agricultural practices.

With the rapid advancements in agricultural automation research towards Agriculture 4.0, numerous research trends have emerged. Examples of mobile robotic platforms used in agriculture can be found in Wall-Ye [4], Spot [12], Rowbot [3], Televitis [17], Bacchus [8], GRAPE [14], the Collaborative Harvest Robot [7], the Reconfigurable UGV [15] and the Dual-Arm Grape Harvesting Robot [13]. These contributions are valuable for assessing the technologies, hardware, and sensor selection in the agricultural context, in our case grape harvesting.

Motivated by the Sestosenso project [9], which aims to develop new sensing technologies for human-robot interaction, in this work we tackle the challenges faced by workers in hillside vineyard grape harvesting, who are particularly exposed to the risks of musculoskeletal diseases linked to manual labour. In particular, our goal is to developed a robotic system that assists the operator by carrying a collection basket at a comfortable height and distance for the grapes collection, which is performed by the operator.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 211–215, 2024.
https://doi.org/10.1007/978-3-031-76428-8_40

We also introduce an innovative design for an end-effector used in fruit collection, tailored to increase payload capacity and reduce harvesting cycle times. Collectively, these efforts mark a substantial step forward in the application of robotics to agriculture.

2 Mobile Manipulator Design Overview

The primary aim of this work is to design a mobile robotic platform integrated with a manipulator, a dedicated gripper, as well as all the necessary perception and control systems, for the assisted collection of grapes.

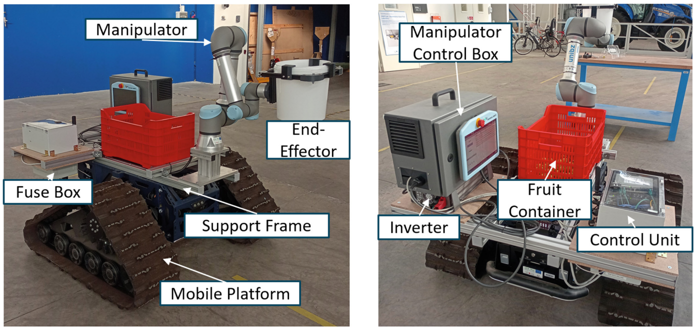


Fig. 1. Prototype of the Mobile Manipulator with the GRABEX end-effector.

The Hawe-Mattro ROVO 2 tracked mobile platform has been chosen as the mobile base of our robotic system. With a dimension of $1200 \times 1230 \times 530$ mm and a weight of 290 kg, thanks to its two 7.5 kW motors it is capable of transporting a payload of up to 500 kg at a maximum speed of 20 km h^{-1} and with slopes up to 45°. The tracked architecture ensures low ground pressure (5 kPa) despite the substantial weight, which guarantees rollover stability even when the manipulator is fully extended. The high-capacity battery (8.8kWh, 100V) provides an autonomy of 8 h or 40 km and powers all on-board equipment.

A dedicated aluminium frame has been designed secure the payload, considering several requirements: (1) Ensuring the system operates within the machine's limits of power, weight, dimensions and stability. (2) Maintaining acceptable machine manoeuvrability on rough terrain. (3) Protecting the electronics against environmental factors such as dust, humidity and sunlight [6]. (4) Compliance with safety standards, given the close collaborative human-robot interaction [1,2,10,11]. (5) Facilitating future expansion and adaptability with a modular design.

As shown in Fig. 1, the proposed frame spreads horizontally to avoid raising the centre of mass, thus reducing the risk of rollover instability. At the front is located the manipulator coupling flange, which is raised to extend the workspace of the manipulator as explained in Sect. 4. At the centre is the main fruit container, which can be quickly released for unloading. An additional container can be fitted to double the load capacity. At the rear are the two control cabinets for the manipulator and the mobile platform, respectively. Their location was chosen to minimise collisions with the manipulator.

3 GRABEX

The GRABEX (Grape Harvesting Robotic Assistant Basket End-Efector eXtension) is an end-effector specifically developed for the collaborative grape harvesting task. GRABEX consists of a basket for fruit collection, with optional attachments for a camera and proximity Time-of-Flight sensors. Its design aims to enable real-time monitoring of the basket's leftover capacity, either via the manipulator's joint sensors or the ones installed on the basket.

Several requirements have been considered for the design of GRABEX, including: 1) reliability and durability, 2) food-contact safety 3) ease of replacement and cleaning 4) standardised mounting.

A GRABEX prototype is visible in Fig. 1. It consists in an aluminium frame with a square shape embracing the container. The container is a 10 L food grade plastic basket (diameter 240 mm, height 260 mm). The size and the aspect ratio have been chosen as a trade-off between maximum capacity, manipulator payload limit, ease of access by the human operator, and minimum volume to avoid self-collisions. The basket is fixed to the structure by means of removable hooks for easy replacement while the frame is fixed to the manipulator via a vertical spacer and a standard flange (ISO 9409-1-50-4-M6). The spacer is designed to keep the basket's centre of mass (CoM) as close as possible to the manipulator 6th joint when full, to minimise the torque when unloading.

4 Manipulator Placement Optimisation

The manipulator's task is to constantly keep GRABEX close to the optimal operator's unload height, to facilitate the laying down of grapes. This position is located in front of the operator, at chest height. The manipulator's base mounting height plays an important role in the safety by keeping hazardous points, such as the tracks, away from the operator. However, when placed above the GRABEX target position, the manipulator's workspace is reduced. Simulations indicate that a mounting 160 mm above the robot's base offers a balanced solution, ensuring operator safety at an acceptable sacrifice in reachability.

5 Payload Estimation Task

In evaluating the robotic arm's interaction, a kinematic analysis is required, focusing on joint movement and interaction with the end-effector. Quantifying the end-effector forces utilising the joint motor torques ensures the precision

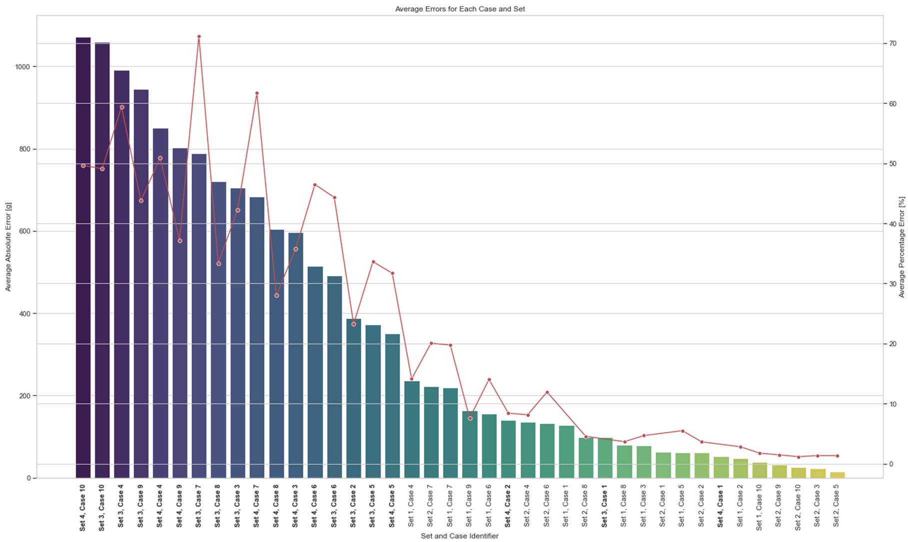


Fig. 2. Evaluation of the UR5e manipulator force measurement accuracy.

of the filling status estimation. Accurate assessment of the payload capacity is crucial to optimise operational efficiency and ensure mechanical integrity. With this in mind, an in-depth analysis of the accuracy of the (indirect) measurement of the weight force provided by the robot has been carried out. In particular, different known weights (up to 3 kg) have been tested, and placed at different distances from the TCP centre (i.e., to emulate the CoG displacement as the basket is filled). The manipulator has then been moved to different locations within its workspace. In Fig. 2 the error between the estimated weight and the real one is reported. This can be high under certain conditions (e.g. 70% in the case of an heavy weight fixed ad a high horizontal distance from TCP). The GRABEX design has been optimised considering also these aspects, to limit the measure error.

6 Conclusions and Future Works

In this paper we presented a prototype for a grape harvesting assistant consisting of a mobile base and a manipulator, discussing the critical design criteria. We also presented a novel end effector design, which was optimised to reduce payload estimation error while meeting safety standards and torque limits.

Future developments aim to incorporate navigation algorithms and sensor technologies to enable the robot to autonomously navigate and operate across varied agricultural landscapes, making real-time decisions, and adapting to changing condition while meeting all safety standards for autonomous and collaborative operations.

As the system evolves towards full autonomy, the integration of advanced safety features will be imperative to ensure a safe working environment for human-robot collaboration. This involves developing the robot's ability to detect and respond to human presence.

References

1. ISO 10218-1:2011 robots and robotic devices – safety requirements for industrial robots – part 1: Robots. International Standard (2011)
2. ANSI-RIA R15.06 American national standard for industrial robots and robot systems – safety requirements. American National Standard (2012)
3. Rowbot (2023). https://www.rowbot.com. Accessed 27 Nov 2023
4. Agency, A.N.: Wall-ye wine robot takes bow in burgundy (2012). YouTube video. https://youtu.be/VjeG8CSK0kE
5. Bechar, A., Vigneault, C.: Agricultural robots for field operations: concepts and components. Biosys. Eng. **149**, 94–111 (2016)
6. International Electrotechnical Commission. Degrees of protection provided by enclosures (IP code). Standard IEC 60529, IEC (2013)
7. Conejero, M.N., Montes, H., Andujar, D., Bengochea-Guevara, J.M., Ribeiro, A.: Collaborative harvest robot. In: Iberian Robotics Conference, pp. 415–426. Springer, Cham (2022)
8. Droukas, L., et al.: A survey of robotic harvesting systems and enabling technologies. J. Intell. Robot. Syst. **107**(2), 21 (2023)
9. European Commission: Physical cognition for intelligent control and safe human-robot interaction (2022). https://doi.org/10.3030/101070310
10. International Organization for Standardization. Robots and robotic devices – collaborative robots. Standard ISO/TS 15066:2016, ISO (2016)
11. International Organization for Standardization. Robots and robotic devices – safety requirements for industrial robots – part 1: Robots. Standard ISO/TR 20218-1:2018, ISO (2018)
12. Hill, E.: Spot the robotic dog helps out down on the farm (2022). Accessed 27 Nov 2023
13. Jiang, Y., Liu, J., Wang, J., Li, W., Peng, Y., Shan, H.: Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation. Front. Plant Sci. **13**, 881904 (2022)
14. Roure, F., et al.: GRAPE: ground robot for vineyArd monitoring and ProtEction. In: Ollero, A., Sanfeliu, A., Montano, L., Lau, N., Cardeira, C. (eds.) ROBOT 2017. AISC, vol. 693, pp. 249–260. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-70833-1_21
15. Saeed, R.A., Tomasi, G., Carabin, G., Vidoni, R., von Ellenrieder, K.D.: Conceptualization and implementation of a reconfigurable unmanned ground vehicle for emulated agricultural tasks. Machines **10**(9), 817 (2022)
16. Stafford, J.V.: Implementing precision agriculture in the 21st century. J. Agric. Eng. Res. **76**(3), 267–275 (2000)
17. Tardaguila, J., Stoll, M., Gutiérrez, S., Proffitt, T., Diago, M.P.: Smart applications and digital technologies in viticulture: a review. Smart Agric. Technol. **1**, 100005 (2021)

Reflective Understanding for Dependable Robots

Ricardo Sanz(✉) and Esther Aguado

Autonomous Systems Laboratory, UPM-CSIC Centre for Automation and Robotics, Universidad Politécnica de Madrid, Madrid, Spain
ricardo.sanz@upm.es

Abstract. To achieve autonomy in open, unstructured, changing environments while pursuing a specific mission ordered by human principals, robots shall control their bodies, adapt to different operational conditions, face unexpected situations, and decide about tradeoffs that may be presented to them. Achieving dependability implies keeping performance while guaranteeing other essential desiderata like safety, security or reduced environmental impact. In CORESENSE (http://coresense.eu) we seek a general, knowledge-based strategy to build resilient robotic systems. This strategy is based on the idea of making the robots *understand* their environments, themselves and their missions. We think that meagre understanding has always been the essential limitation of AI. In CORESENSE we target a science and technology of understanding to make robots more aware and capable of better dealing with the expected and the unexpected.

Keywords: Robotics · AI · autonomy · resilience · understanding · awareness · knowledge · models · ontology

1 The Case of Robot Dependability

The construction of controllers for autonomous robots is always a challenge because the design of the controller requires having good knowledge of the robot and its environment and this is not easy in open, changing environments. The *raison-d'etre* of autonomous systems engineering research is indeed the development of methods to overcome such lack of knowledge. Autonomous systems are designed to be deployed in environments where the reality may depart from what is initially known and expected. And beyond having the capability of performing specific tasks—e.g. movement or grasping—it is necessary to endow them with an additional property: dependability. Dependable robots keep their mission going –if it is feasible.

There exists only one way to system dependability[1]: Make the systems in such a way that they have an intrinsic dynamics that forces them into defined

[1] If considering only actions on the system.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 216–219, 2024.
https://doi.org/10.1007/978-3-031-76428-8_41

attractors no matter what the disturbances are. The ways of achieving this are multiple, depending on many factors: what the disturbances are, what we know about them, what can be changed in the system, what amount of effort can be used, what level of dependability is desired, etc. At the end of the day the concrete engineering strategy to follow will result from a trade-off analysis between all these factors.

Simplifying enormously, we can identify two big classes of strategies to change the system behaviour: $S1$) change the extant system structure, hence forcing a different system behaviour; and $S2$) adding a suppletory system that forces a different aggregate system behaviour. This last is indeed the role of control systems. SESTOSENSO attempts a combination of $S1$ and $S2$, while CONVINCE and CORESENSE are essentially focused on $S2$.

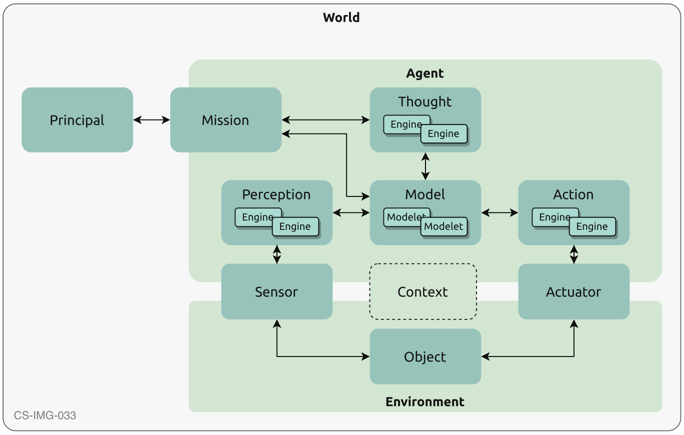


Fig. 1. The domain entities to be understood are a supersystem composed by an agent, an environment, and a mission to be performed.

Robustness is a system's ability to maintain persistency of service in the presence of disturbances. This is commonly achieved using $S1$ but $S2$ can also be effective if the temporal response is fast enough. *Resilience* is a system's ability to recover from a fault and maintain persistency of service in the face of harsh disturbance. This is commonly achieved using $S2$ but $S1$ is also a possibility[2].

Designing better robot control strategies will improve both R&R, but it is also necessary to have good engineering methods to build such controllers. Instead of using *ad-hoc* system design methods to overcome R&R challenges of concrete robots, we seek developing *principles-based methods* that should be more general and effective.

[2] Indeed, system maintenance and evolution does this and a larger timescale, whether human-done or system-intrinsic.

2 The CORESENSE Approach

In CORESENSE we wholeheartedly adopt the "knowledge is power" lemma. We try to empower robots with the deep knowledge necessary for understanding what is going on in the robot-environment relation and compute the actions that can keep the mission going. This strategy is obviously not new [3], we just only try to ground it in fundamental system-cognitive principles.

The controllers that implement the robot actions may be of hybrid nature that implies a need of a more abstract conceptual perspective. In low-level controllers (Level 0), these actions change values of variables; in learning-level controllers (Level 1), the actions change the knowledge that the controller is using; in adaptive-level controllers (Level 2), the actions change the controller itself[3].

Three ideas guide the CORESENSE project:

– That models are the essential forms of knowledge that an agent needs for meaningful interaction.
– That it is possible to build a system-architectural infrastructure for general, modular, hybrid, knowledge based controllers.
– That this infrastructure can be used by humans to build new systems and by the systems themselves to endow them with *deep understanding*.

The knowledge used in the CoreSense architecture represents not only the state of the world and the agent[4] procedures but a complete agent-environment-mission tripartite relation (see Fig. 1). This knowledge is created in the form of modular, integrated models in an MBSE lifecycle that integrates the engineering and operation phases. In a sense, what we seek in CORESENSE is a customisable cognitive architecture that can support runtime system self-engineering—EngOps—lifecycles for autonomous robots (see Fig. 2).

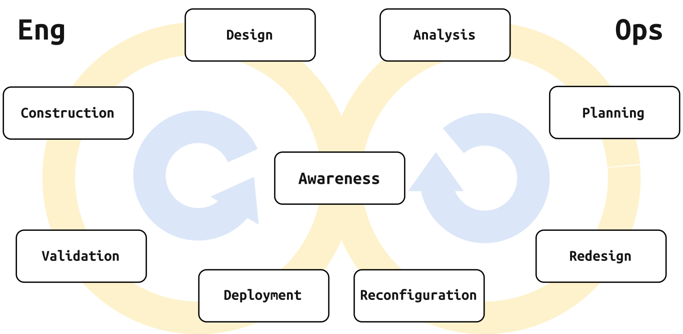


Fig. 2. An EngOps lifecycle breaks the barrier between engineering time and runtime.

[3] Note that indeed Level 0 $\subset$ Level 1 $\subset$ Level 2.
[4] We commonly use the term "agent" to refer to the software part of the robot.

3 The Concept of Understanding

To achieve an increased level of dependability we need to endow the system with the capacity of understanding well the situation and be aware of it and the effects of its own actions. *Understanding* and *awareness* are two central objectives of the development of the CoreSense model-based architecture. This work is just another step in a long tradition on model-based understanding and action that has been central in many domains like science [2], cognition [1] or control [4] or more specifically, robotics.

The following definition for understanding has been initially proposed:

> A **subject** S understands a **phenomenon** ϕ if it has a set of **models** M_ϕ of ϕ and those models can be used to make **sound inferences** about the phenomenon.

The *subject* is the individual (human or machine) that performs the action of understanding. Examples: aerial robot, social robot, physicist, economist, sociologist. *Sound inferences:* are those that follow the modelling relation, i.e., that implications in the model **correspond** to causality in the system that is modelled. The conclusions inferred from premises using the model correspond to object states of the modelled system.

A variant definition under consideration addresses the mission-centric nature of purposeful understanding for agents that shall decide on actions:

> A subject S understands a phenomenon ϕ if the subject S is able to infer facts from observations of that phenomenon (using a set of models M_ϕ of ϕ) that help the subject S **judge the observed situation** in order to assess possible decisions about what to do in that situation.

The project implementation work is just starting, and these ideas under development as reusable designs, software assets and methodologies. In this talk we will address the objectives, strategy and current status of the project. The author acknowledges the support of the European Commission through the Horizon Europe project #101070254 CORESENSE.

References

1. Craik, K.J.W.: The Nature of Explanation. Cambridge University Press, Cambridge (1943)
2. Magnani, L., Bertolotti, T. (eds.): Springer Handbook of Model-Based Science. Springer, Heidelberg (2017)
3. Perlis, D.: How can a program meant. In: International Joint Conference on Artificial Intelligence (IJCAI-87), pp. 163–166 (1987)
4. van den Hof, P.M.J., Scherer, C., Heuberger, P.S.C.: Model-Based Control: Bridging Rigorous Theory and Advanced Technology. Springer, Heidelberg (2009)

Robotic Applications

Experimental Tests of a Motion Planning System Based on Multi-Objective Optimisation for Nuclear Decommissioning Practice Using Long-Reach Systems

Kaiqiang Zhang[1](✉), Vijay M. Pawar[2], Faiz Rahman[1], Alice Cryer[1], Tomoki Sakaue[3], Fumiaki Abe[3], Masaki Sakamoto[3], Yoshimasa Sugawara[3], David Marquez-Gamez[4], Ricardo J. Louro Rei[4], Bahadir Kocer[5], Shu Shirai[3], Wataru Sato[3], Ipek Caliskanelli[1], Matthew Goodliffe[1], Harun Tugal[1], Salvador Pacheco-Gutierrez[1], and Robert Skilton[1]

[1] UK Atomic Energy Authority, Abingdon OX13 4DB, UK
kaiqiang.zhang@ukaea.uk
[2] University College London, London WC1E 6BT, UK
v.pawar@ucl.ac.uk
[3] Tokyo Electric Power Company Holdings, Fukushima, Japan
[4] AtkinsRéalis Company, Epsom KT18 5BW, UK
[5] University of Bristol, Bristol BS8 1QU, UK

Abstract. Motion planning for long-reach robotic systems is a complicated laborious task reliant on operators with specialist expertise in nuclear decommissioning practice. Thus, a motion planning system is developed in this work, motivated to improve the efficiency of planning motions in such safety-critical environments. The system can assist operations by generating viable routing solutions for long-reach robots with high degrees of freedom in nuclear decommissioning operations using multi-objective optimisation. It was experimentally tested using a hyper-redundant manipulator, which was for remote maintenance of an experimental fusion-energy tokamak device. The results demonstrate the feasibility of applying optimisation-based planning to complex nuclear tasks, overcoming current excessive reliance on manual methods to improve operational efficiency and safety assurance in decommissioning.

Keywords: Nuclear Decommissioning · Remote Operation · Long-Reach Robot · Motion Planning · Multi-Objective Optimisation

1 Motion Planning Problem in Decommissioning Practice

Long-reach robotic manipulators are critical systems enabling access to hazardous environments inside nuclear facilities during decommissioning activities. At the Fukushima Daiichi plants, an extremely slender robotic manipulator with 18 degrees of freedom (DOF) and approximately 22 m in length was custom-built to retrieve fuel debris from the reactor cores through very narrow openings [1]. Similarly, at Sellafield, long-reach

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 223–228, 2024.
https://doi.org/10.1007/978-3-031-76428-8_42

manipulators are required to access legacy radioactive sites via small ports [2]). The UK Atomic Energy Authority (UKAEA) is planning to use specialised hyper-redundant long-reach manipulators for decommissioning the Joint European Torus (JET) fusion tokamak device [3, 4]. These long-reach robots allow remote operations at safe distances to avoid hazard (e.g., radiation) exposure [5]. Their high dexterity enables intricate tasks in confined spaces.

However, manually directing them through convoluted passages is difficult [1, 6]. Precise motion planning algorithms are needed to optimise their trajectories through constraint spaces. In current practice, planning long-reach robot motions is a highly complex, manual, and time-intensive process. Teams of specialised operators with years of experience are required to plan multi-DOF long-reach manipulator motions while considering many factors, including but not limited to the following important considerations [6, 7].

- **Accounting for various environmental conditions and constraints**, like radiation hotspots, obstacles, and openings, to maximise robot lifespan and prevent collisions, which is challenging [8].
- **Frequently re-planning for changing environments** as decommissioning work progresses over long timescales [9].
- **Avoiding risks as a result of control uncertainties and mechanical flexibility** considering the manipulators' kinodynamic limits.

Theoretically, given full knowledge of conditions and constraints, motion planning reduces to a multi-objective optimisation problem. Although planning has advanced for industrial and mobile robots [10, 11], limited research exists to assist operators of long-reach systems for decommissioning. In response, this work presents a motion planning system using multi-objective optimisation algorithms to generate efficient, reliable motion sequences considering objectives like smoothness, clearance, and irradiated dose.

2 Motion Planning System Using Multi-Objective Optimisation

To enable automated motion planning for the long-reach robots, a modular software system was developed using the Robot Operating System 2 (ROS2) framework. ROS2 allows for integrating different software components in a flexible manner, facilitating convenient maintenance, upgrades, reconfiguration, and reuse for various research and development needs. A specialised operator-focused user interface (UI) was also implemented with Unity 3D (see Fig. 1). Following human factors guidelines, this UI exposes only essential control and visualisation interfaces needed for motion planning whilst not overwhelming operators with irrelevant information.

The core motion planning capabilities are developed based on the open-source CHOMP algorithm [12]. The motion planner generates optimised manipulator trajectories along operator-specified waypoints, considering multiple objectives, including 1) **Smoothness** of the planned motions (described in terms of velocity, acceleration and jerk), 2) **Clearance** from obstacles and space constraints, and 3) **Total** irradiated air-karma **dose exposure** on specific robot joints during the planned motions.

The multi-objective motion planner was tested in simulated environments modelled after real-world nuclear decommissioning sites. Additional software functions have been implemented to support operator awareness, including 1) a **singularity avoidance monitoring** function to warn when the manipulator reaches a singular configuration, and 2) a real-time **collision-risk monitoring** function to warn if any part of the robot breaches a minimum clearance threshold.

Note that the environmental conditions are assumed to be pre-known (e.g., measured by survey operations or estimated from building design data) or measurable information (e.g., collected by sensors online). The kinematic design, the geometric data of the bodies, and the joint actuation limitations of the long-reach manipulator are also known from the design. Simulated sensors provide enhanced perception of the environments and tasks.

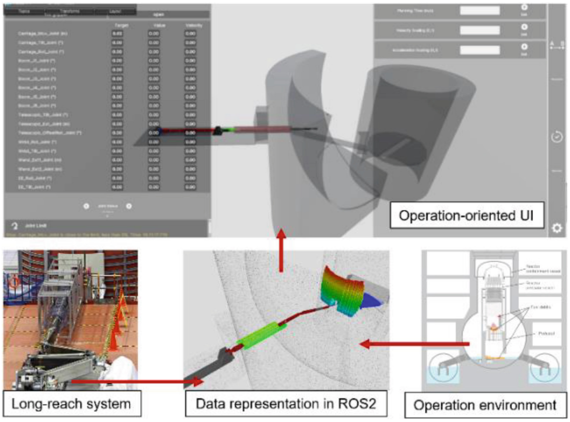


Fig. 1. An operation-oriented user interface (UI) for a decommissioning task at Fukushima Daiichi. The UI communicates with the motion planning function in ROS2. The environment information and the long-reach system's design data are used for setting up the operation scene.

3 Experimental Setup and Tests

Motivated to verify the feasibility of using the system in practical operations, TARM [13] was adopted in the experimental tests (see Fig. 2). Before being renovated into a test platform, TARM was deployed and operated to maintain JET, as an experimental fusion reaction tokamak, including radiation hazards. TARM is a long-reach system with its hardware limits and special design considerations for operating in challenging environments. It is with redundant DOF especially giving significant freedom of in-horizontal-plane motions. Thus, an experimental setup was configured (see Fig. 2) using TARM, as a representative device that could reflect typical challenges when deploying long-reach systems in decommissioning-related environments.

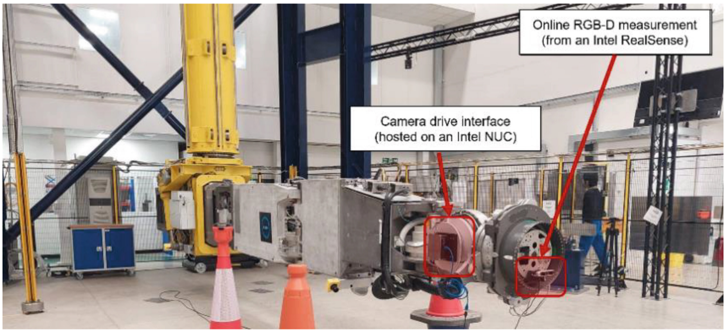


Fig. 2. Experimental setup with an RGB-D camera integrated with TARM [13].

A mock-up scene (see Fig. 3) was facilitated to conduct a series of test cases as follows. *Case 1* defined waypoints to move TARM (considering the hardware limits, including joint ranges and velocity/acceleration constraints). By incorporating real robot parameters, the planner produced achievable trajectories adhering to kinodynamic restrictions. *Case 2* considered the presence of a virtual wall, and drove TARM through the challenging narrow opening at the wall and into a confined workspace beyond. *Case 3* added known static obstacles (Obstacle A, B, and C in Fig. 3) that were perceived by the RGB-D camera. The system avoided these obstacles when generating trajectories, showing effective utilisation of sensor measurements online. *Case 4* evaluated the response to a dynamic environment by removing an obstacle during planning. The system adapted the resulting trajectories accordingly, highlighting its adaptability to changing environments. *Case 5* used a simulated Gamma-radiation dose rate map to simulate radiation presence, and it proved the ability to consider non-standard optimum objectives, e.g., irradiated dose.

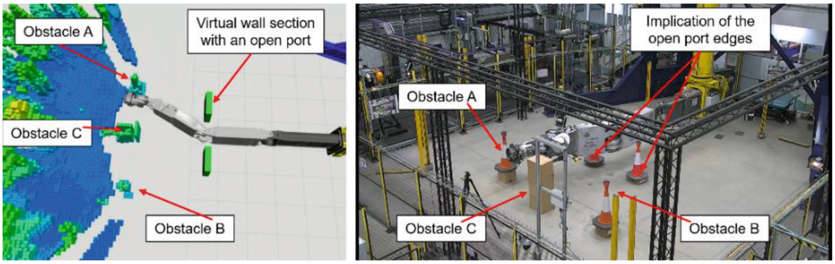


Fig. 3. Experimental mock-up scene for the experimental tests.

These results provided a proof-of-concept validation of the system on representative hardware in environments mimicking real-world challenges (as summarised in Table 1). While further development is required for fully autonomous operations, this represents a major milestone toward integrating automated intelligence to assist operators.

Table 1. Experimental test cases.

Case	Plan motions via a given set of waypoints	Robot hardware limits	Access via the virtual wall	Existence of pre-known obstacles	Variation (removal/addition) of obstacles	Known radiation map
1	✔	✔				
2	✔	✔	✔			
3	✔	✔	✔	✔		
4	✔	✔	✔	✔	✔	
5	✔	✔	✔	✔		✔

4 Summary

This paper presents a motion planning system tailored for long-reach manipulators operating in complicated nuclear decommissioning operations. The motion planning system can utilise a multi-objective optimisation-based approach for planning motion sequences, when a long-reach system is commanded to move along given waypoints, considering the motion smoothness, the robot-environment clearance, and the irradiated total dose. The experimental results demonstrated the viability of motion planning techniques based on multi-objective optimisation to enhance the safety, reliability, and efficiency of real-world robotic deployment in nuclear decommissioning.

Acknowledgement. The associated work is conducted within the LongOps programme that is funded by the UK Nuclear Decommissioning Authority, the UK Atomic Energy Authority, Tokyo Electric Power Company Holdings, and UK Research and Innovation (ref. 107463). AtkinsRéalis Co Ltd. and UCL Consultants Ltd. were contracted and subcontracted, respectively, to develop the software. The views and opinions expressed herein do not necessarily reflect those of the organisations.

References

1. Tepco, H.D.: Status of R&D for the trial removal of fuel debris from Unit 2, TEPCO HD (2020)
2. NDA and Sellafield Ltd: End of reprocessing at thorp signals new era for Sellafield (2018). https://www.gov.uk/government/news/end-of-reprocessing-at-thorp-signals-new-era-for-sellafield. Accessed online: 14 Sep. 2023
3. Sykes, N., et al.: Design for high productivity remote handling. Fusion Eng. Des. **86**(9–11), 1843–1846 (2011)
4. Wilson, K., Stevens, K.: Decommissioning planning for the Joint European Torus fusion reactor. WM'07 conference (2007)
5. Stedman, H., et al.: Evaluating immersive teleoperation interfaces: Coordinating robot radiation monitoring tasks in nuclear facilities. In: 2023 ICRA. IEEE (2023)

6. Buckingham, R., Loving, A.: Remote-handling challenges in fusion research and beyond. Nat. Phys. **12**(5), 391–393 (2016)
7. David, O., et al.: Operational experience feedback in JET remote handling. Fusion Eng. Des. **75**, 519–523 (2005)
8. Brownridge, M.: Remote decommissioning experiences at Sellafield. In: WM'06 conference (2006)
9. Kaneko, T., et al.: Evaluation of corrosion behaviors for reactor pressure vessels/primary containment vessels in Fukushima Daiichi units 1–3 nuclear power plant. J. Nucl. Sci. Technol. **52**(6), 773–783 (2015)
10. Du, Z.C., et al.: A review on kinematic, workspace, trajectory planning and path planning of hyper-redundant manipulators. In: 10th IEEE CYBER), pp. 444–449. IEEE (2020)
11. Kavraki, L.E., LaValle, S.M.: Motion planning. In: Springer handbook of robotics, pp. 139–162. Springer (2016)
12. Ratliff, N., et al.: CHOMP: Gradient optimization techniques for efficient motion planning. In: 2009 IEEE ICRA, pp. 489–494. IEEE (2009)
13. McGarrigle, D., et al.: Analysis of existing and proposed maintenance deployment systems toward DEMO MPD development. IEEE Trans. Plasma Sci. **50**(11), 4485–4490 (2022)

Manufacturing Logistics Optimization Using the SPECTER Task Planner: A Shoe Manufacturing Logistics Case Study

Anatoli A. Tziola(✉) and Savvas G. Loizou

Department of Mechanical Engineering and Materials Science and Engineering, Cyprus University of Technology, Limassol, Cyprus
{anatoli.tziola,savvas.loizou}@cut.ac.cy

Abstract. This paper presents an application of SPECTER task planner for shoe manufacturing logistics. In this case study we propose an abstract model of the work flow of a shoe manufacturing company taking into account: i) the workflow stages; ii) the time costs (i.e. a machine operation, products time construction, worker transitions between work-cells); and iii) the agents involved in the production line (such as machines, humans, materials, products etc.). Based on the derived abstraction, optimal solutions are provided for utilizing 1, 2 or 3 workers in the production line. The paper validates the modeling power of the SPECTER framework, while demonstrating its potential applications in providing solutions for the industry.

Keywords: Task planning · shoe manufacturing logistics · SPECTER

1 Introduction

Solving and automating complex manufacturing logistics tasks in order to produce a variety of products, has been one of the areas of research interest in the field of robotics and in particular their applications to manufacturing logistics. Shoe making is such a labour intensive process that requires many different sizes, styles and materials for shoes. The shoe manufacturing workflow undergoes various and complex steps before a quality product arrives at a customer, making it hard to automate the planning of such a production. Among the major challenges affecting shoe manufacturing are inefficient utilization of resources, including energy, materials and human resources, leading to increase the cost of the final product.

The manufacturing of shoes consists of multiple workflows with intermediate stages until a pair of shoes is produced. Some of the workflows are the shoe designing, cutting from leather to textiles, stitching, final assembly and packing. Each workflow is distinguished in more individual steps. This work focuses on the optimization of the final assembly workflow and its individual steps of a shoe manufacturing industry in order to optimize the intra-factory logistics. The

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 229–234, 2024.
https://doi.org/10.1007/978-3-031-76428-8_43

challenge is that a single production line consists of many intermediate steps and multiple types of production processes depending on the customer order. Utilizing the SPECTER task planning framework proposed by the authors in [1,2], different scenarios are studied, to determine the resources and the time required for shoes production depending on the available human workers in the factory as well as the sequence of actions that need to be performed (i.e. the machines/robots operation sequence, the materials utilization, worker actions). The presented results depend on the assumptions, the accuracy of the data and on the level of abstraction and should be considered as a preliminary result that demonstrates the capabilities of our modeling framework.

The rest of the paper is organized as follows: Sect. 2 presents the description of the workflow, and the workflow abstract model, while Sect. 3 presents the results of 3 different scenarios. Section 4 concludes the paper.

2 Analysis and Modeling

2.1 Description of the Final Assembly Workflow

Final assembly is the workflow where upper leather is shaped, lasted and assembled with sole to get final the product i.e. the shoes. The final assembly is performed before the packing procedure. The main techniques of final assembly used in the shoe manufacturing industry, presented in Fig. 1 are: **Toe Shaping:** Unprocessed leather or back-part shaped leather is doubled shaped for preserving the shape and the original appearance of the shoe. **Back-Part Shaping:** Unprocessed leather or toe shaped leather is doubled shaped for preserving the shape and the original appearance of the shoe. **Lasting:** The unprocessed leather or back-part shape leather or back-part and toe shaped leather can be processed. In this stage, the upper leather is attached to the bottom of the shoe. **Side and**

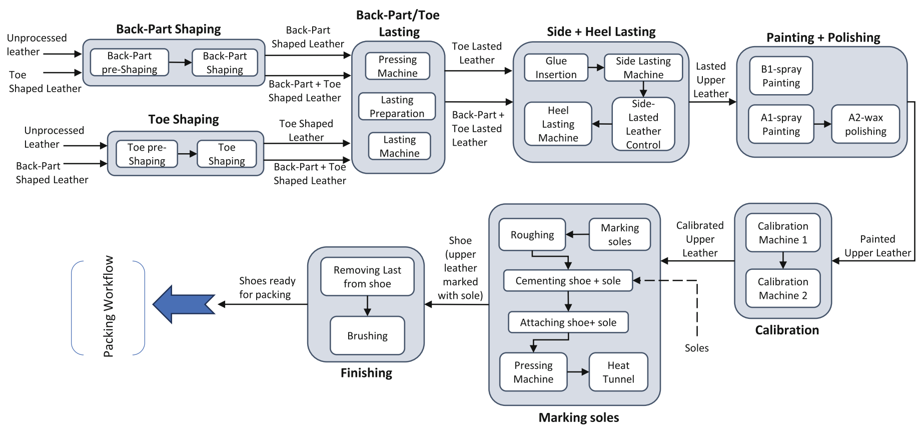


Fig. 1. Diagram of the final assembly workflow.

Heel Lasting: Setting the final shape of the shoe and holds it in place so the outsole can be permanently attached. **Painting and Polishing**, **Calibration:** Calibrating the shoe last. **Marking soles:** Marking the sole in the upper lasting leather, then roughing and cementing shoe with sole. When shoe with sole is attached, the shoe is pressed and then passed through the heat tunnel. **Finishing:** Removing last from shoe's inside, then brushing. Finally, the shoe is produced and continues to packing workflow.

2.2 Abstract Model

In order to model the workflow shown in Fig. 1 an abstraction is required to cast the problem in the discrete processes domain of the SPECTER framework. Due to space restrictions, an outline of the process abstraction is depicted in Fig. 2. The nodes of the diagram represents the processing steps and buffer zones, whereas the arrows indicate the input and the output products of each process and buffer zone.

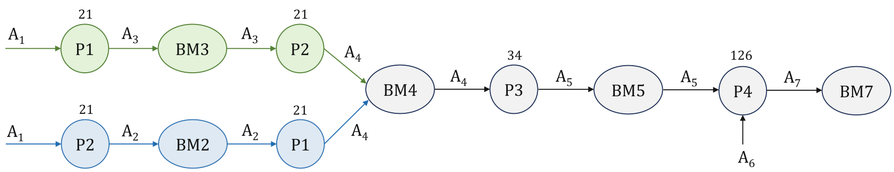


Fig. 2. Workflow abstraction. Numbers above circles indicate duration (sec).

Workers and items are modelled as agents expressed as ϵ_0-NFAs [1,2]. The detailed representations of each agent are beyond the scope of the current paper. More specifically, the unprocessed leather is modeled as A_1; the toe-shaping leather is modeled as A_2; the back-part shaping leather is modeled as A_3; the back-part and toe-shaping leather is modeled as A_4; the lasted leather is modeled as A_5; the soles are modeled as A_6; and the shoes are modeled as A_7. Additionally, 3 workers are considered working in the factory and modeled as A_8, A_9, A_{10} respectively.

Step $P1$ represents the back-part shaping, $P2$ the toe shaping, $P3$ the back-part and toe lasting processes. The processes of side and heel lasting, painting and polishing, calibration, marking with soles and finishing are grouped in step $P4$ since no sequence changes are allowed. Moreover, $BM2$ represents the buffer zone for items modeled as A_2, $BM3$ the buffer for A_3, $BM4$ the buffer for A_4, $BM5$ the buffer for A_5 and $BM7$ the buffer for A_7. The processing steps and the buffer zones are considered as agents locations.

The unprocessed leather A_1 can be converted to the back-part shaping leather A_3 at $P1$ and then, A_3 placed at buffer $BM3$. Also, A_1 can be converted to A_2 at $P2$ and then, A_2 is placed at buffer $BM2$. Item A_4 can be produced at $P1$ using item A_2 or at $P2$ using item A_3. Then, A_4 is placed at $BM4$. Item A_5 is produced at $P3$ using item A_4. Item A_7 is produced at $P4$ using items A_5 and A_6. Finally, item A_7 is placed at $BM7$.

3 Case Study Results

We run 3 scenarios utilizing 1, 2 or 3 workers respectively. We assume that there are unlimited resources of unprocessed leather at $P1$ and $P2$; and soles are at $P4$. For the initial conditions we assume that toe-shaping leather, back-part shaping leather, back-part and toe-shaping leather, lasted leather and shoes have not been produced yet, the workers could be anywhere in the factory.

1st Scenario: 1 Worker. For the 1st scenario, we considered 8 agents in total; 7 items and 1 worker. The cardinality of the environment's state space is 14,400 states. The objective is to produce A_7 and locate it at $BM7$ utilizing 1 worker.

In the solution computed by SPECTER, the objective is fulfill after 20 steps. In words, (T_1) worker goes at work-cell $P1$, (T_2) worker inserts A_1 in machine at $P1$, (T_3) A_3 is produced at $P1$ after 21 s, (T_4) worker places A_3 at $BM3$, (T_5) worker goes at work-cell $P2$, (T_6) worker inserts A_1 in the machine at $P2$, (T_7) A_2 is produced at $P2$ after 21 s, (T_8) worker places A_2 at $BM2$, (T_9) worker inserts A_3 in machine at $P2$, (T_{10}) A_4 is produced at $P2$ after 21 s, (T_{11}) worker places A_4 at $BM4$, (T_{12}) worker goes at work-cell $P3$, (T_{13}) worker inserts A_4 in machine at $P3$, (T_{14}) A_5 is produced at $P3$ after 34 s, (T_{15}) worker places A_5 at $BM5$, (T_{16}) worker goes at work-cell $P4$, (T_{17}) worker inserts A_5 in machine at $P4$, (T_{18}) worker inserts A_6 in machine at $P4$, (T_{19}) A_7 is produced at $P4$ after 126 s, (T_{20}) worker places A_7 at $BM7$. Thus, the task is fulfilled; shoes produced and places at buffer zone $BM7$. The time required to implement the solution is 269 s utilizing 1 worker for the whole process.

2nd Scenario: 2 Workers. For the 2nd scenario, we considered 9 agents in total; 7 items and 2 workers. The cardinality of the environment's state space is 72,000 states. The objective is to produce A_7 and locate it at $BM7$ utilizing 2 workers.

In the solution computed by SPECTER, the objective is fulfill after 16 steps. In words, (T_1) worker A_8 goes at work-cell $P2$; worker A_9 goes at work-cell $P1$, (T_2) worker A_8 inserts A_1 in machine at $P2$, while worker A_9 inserts A_1 in machine at $P1$, (T_3) after 21 s, A_2 is produced at $P2$ and A_3 is produced at $P1$, (T_4) worker A_9 places A_3 at $BM3$, while worker A_8 places A_2 at $BM2$, (T_5) worker A_8 inserts A_3 in the machine at $P2$, while worker A_9 inserts A_2 in the machine at $P1$, (T_6) 2 entities of A_4 are produced after 21 s; 1 entity at $P1$ and 1 entity at $P2$, (T_7) worker A_8 places A_4 entities at $BM4$, (T_8) worker A_8 goes at work-cell $P3$, (T_9) worker A_8 inserts A_4 in machine at $P3$, (T_{10}) A_5 is produced at $P3$ after 34 s, (T_{11}) worker A_8 places A_5 at $BM5$, (T_{12}) worker A_8 goes at work-cell $P4$, (T_{13}) worker A_8 inserts A_5 in machine at $P4$, (T_{14}) worker A_8 inserts A_6 in machine at $P4$, (T_{15}) A_7 is produced at $P4$ after 126 s, (T_{16})

worker A_8 places A_7 at $BM7$. The time required to implement the solution with concurrent execution capability requires 238 s.

3rd Scenario: 3 Workers. For the 3rd scenario, we considered 10 agents in total; 7 items and 3 workers. The cardinality of the environment's state space is 360,000 states. The objective is to produce A_7 and locate it at $BM7$ utilizing 3 workers. The solution provided by SPECTER consists of 16 steps from (T_1) to (T_{16}) as described in scenario 2 but in this case the solution utilizes worker A_9 instead of A_8 and worker A_{10} instead of A_9. The time required to implement the solution with concurrent execution capability requires 238 s.

Figure 3 demonstrates the results from the 3 case studies.

Fig. 3. Optimal cost (time) for the three case studies.

4 Conclusions

In this paper a case study for a manufacturing logistics optimization problem is presented, utilizing the SPECTER task planning framework. An abstraction was proposed that modeled key features of the work flow and the resulting model abstraction was implemented on the SPECTER task planner. Three case studies were investigated, with 1, 2 and 3 workers working on the production line. Interestingly a decrease in the production time was observed only when a second worker was added in the workflow but no change was registered when a third worker was added. The results concluded that an increase above the optimal number of workers will not decrease the production time. Further investigations with different machines/robots/buffers and factory layout arrangements are under consideration as future work.

Acknowledgements. This work was partially supported by European Union's research and innovation programs under grant agreement 951813 (Better Factory, H2020) and 101092295 (CIRCULOOS, Horizon Europe). The authors would like to acknowledge Mr. Aleksander Zur, CEO of TAPI NERO shoe manufacturing, for the discussions and valuable insights in shoe manufacturing logistics.

References

1. Tziola, A.A., Loizou, S.G.: Autonomous task planning for heterogeneous multi-agent systems. In: IEEE International Conference on Robotics and Automation (2023)
2. Tziola, A.A., Loizou, S.G.: Autonomous task planning for heterogeneous multi-agent systems. arXiv Preprint arXiv:2209.08611 (2022)

SmartOfflineNG: An Off-line Programming Solution for the Ceramic Industry

Matteo Ragaglia[1(✉)], Nicola Battilani[2], Antonio Castellano[1], Silvia Costi[2], Joao Marcos Da Silva Araujo[2], Cesare Fantuzzi[2], Gabriele Masotti[1], Mirko Mattioli[3], Giorgio Motta[1], and Umberto Scarcia[4]

[1] Gaiotto Automation S.p.a (SACMI Group), Piacenza, Italy
matteo.ragaglia@sacmigroup.com
[2] Industria Tecnologica Italiana (IT-I), Reggio Emilia, Italy
[3] IPREL Progetti S.p.a (SACMI Group), Imola, Italy
[4] Innovation Lab - SACMI Imola S.c., Imola, Italy

Abstract. Today robots represents a key factor in several industry fields as far as productivity and competitiveness are concerned. To this purpose, off-line programming approaches and tools are clearly needed in order to allow fine development and fast adaptation of robot programs while minimizing machine downtime. In this context, this work introduces the off-line programming software developed by SACMI, named "SmartOffline NextGen".

Keywords: Robot Programming · Off-line Programming

1 Introduction and Problem Setting

Nowadays, robots are becoming key elements in increasing manufacturing competitiveness. Nevertheless, widespread adoption of robotic technologies in small-to-medium sized enterprises is still undermined by certain well-known factors, among which the inherently complex and time-consuming nature of robot programming surely plays a crucial role Pan et al. (2012). Traditional methods for robot programming typically consist in either using the robot teach pendant or in simulating the robot task inside an off-line programming (OLP) environment Wang and Zhang (2002), Erdos et al. (2020). In the first case, the whole programming time turns out to be machine downtime, with significant impacts on the robot's productivity. On the other hand, the OLP approach allows the programmer to develop at least a first draft of the complete program without having to stop the robot, thus limiting machine downtime to the time needed to realize

This work is partly supported by Regione Emilia-Romagna under the agreement "Innovazione della value proposition di Gaiotto Automation 2021–2025" (grant number/CUP: E32C21001090009).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 235–239, 2024.
https://doi.org/10.1007/978-3-031-76428-8_44

the final validation of the program Zheng et al. (2022). Independently from the specific approaches, OLP is almost unanimously considered the most efficient solution to the problem of robot programming, as argued in Castor (2020). To this regard, this work presents SACMI's OLP solution, named "SmartOffline NextGen" (SmOffNG) Ragaglia et al. (2022).

2 General Concepts

The main goal behind the development of SmOffNG consists in offering a comprehensive OLP platform for the robotics applications currently developed by SACMI in the context of the ceramic-sanitaryware sector, that will facilitate the development of complex programs by users that do not have a strong background in robotics and robotic technologies. Given this goals, the design and the development of SmOffNG have revolved around the following key concepts:

- **Cell**: root of the tree structure modelling the virtualized cell (Fig. 1(a));
- **Robot**: a machine endowed with multiple DoFs, either anthropomorphic manipulator or an auxiliary machine (i.e. a linear axis or a turning table);
- **Work-Item**: a generic physical object that can model either the work-piece that is processed during the program or an auxiliary object, like for instance a prohibited area. It is modelled by means of a triangular mesh;
- **Via-Point (VP)**: a destination that a robot needs to reach, either in joint or operational space coordinates. In case the virtualized cell contains more than one robot, a VP will contain a destination for each virtualized robot;
- **Segment**: an interpolated trajectory connecting two consecutive VPs;
- **Trace**: an ordered list of VPs, connected by Segments;
- **Program**: an ordered list of Traces, see Fig. 1(b);

As far as the definition of VPs and the interpolation of the Segments are concerned, SmOffNG was developed in order to identify the following problems:

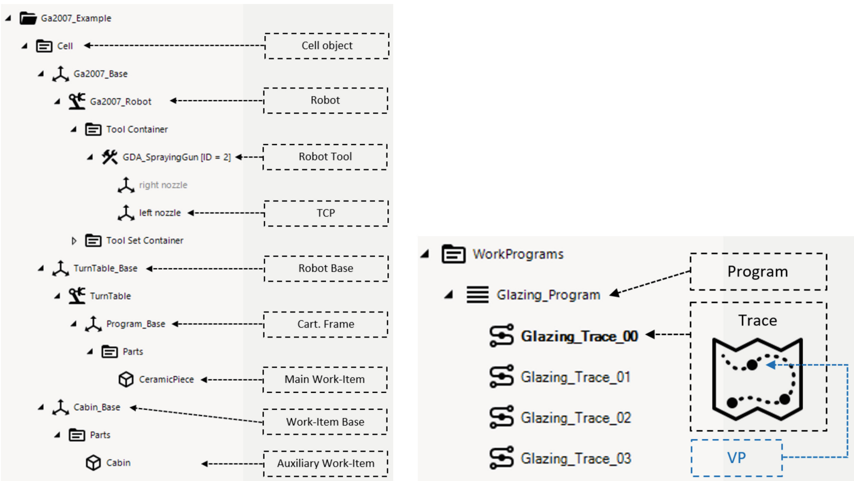
- **Reachability**: either a VP or an interpolated point lies outside the reachable workspace of at least one robot;
- **Singularity**: during a motion interpolated in the Cartesian Space an anthropomorphic manipulator assumes a singular configuration;
- **Collision**: either a VP or an interpolated point are characterized by an interference between two physical objects (links, tools, work-items).

Finally, beside the definition of VPs, Traces and Programs, the main features offered by SmOffNG can be listed as follows:

- **Interactive 3D View**: 3D rendering of the cell;
- **Process Simulation**: physics-based simulation of the task, currently available only for glazing applications;
- **Exporting**: conversion of a defined Program in a format that can be executed on a physical robot. SmOffNG can export programs for Gaiotto proprietary robots and for several commercial vendors.

3 Software Architecture

The main idea behind the design of SmOffNG consists in decoupling the algorithmic portion from the desktop application. The first component is called "BuildServer" (BS) and it is responsible for: i) computing direct and inverse kinematics of the robots; ii) interpolating a continuous trajectories among the various VP defined by the user. On the other hand the desktop application, simply named "Client", acts mainly as HMI, allowing the user to build and modify the virtualized cell, instantiate programs, traces and VPs, and simulate the execution of either programs or traces. The information exchange between the two components is realized via a gRPC Google (2023) interface. In addition, this solution allows to realize the deployment diagram pictured in Fig. 2. Let us assume that the user develops a SmOffNG Program for a specific model of work-piece. According to the scheme, each time an instance of this work-piece enters the plant, the inspection system measures its displacement with respect to the nominal position defined inside SmOffNG. This information is then forwarded to the BS instance running on the so-called "Panel-PC". The BS loads the Program, configures the coordinates of the main work-item's base according to the measured displacement, and finally re-compiles the program taking into account this offset. In case recompilation is successful, the BS will export the robot program that will be transferred to the robot controller for execution. Otherwise, it will output one or more error messages to inform the user about reachability, singularity or collision problems that emerged after the offset was applied.



(a) (b)

Fig. 1. 1(a) Example of Tree Structure modelling a cell, 1(b) Program breakdown.

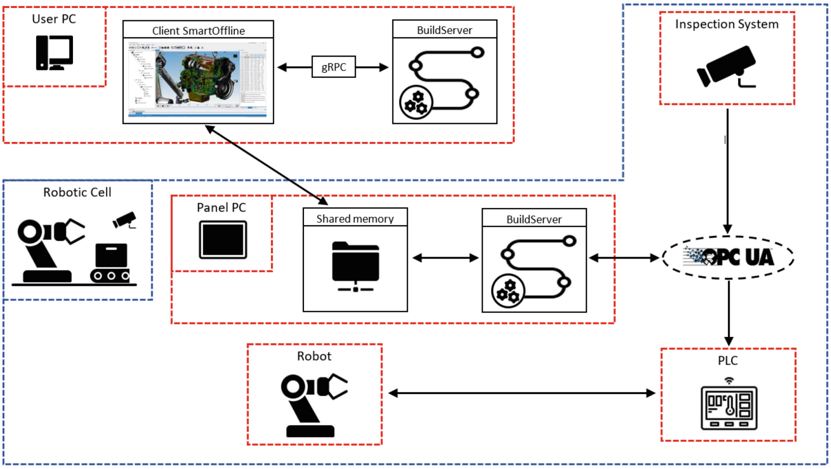


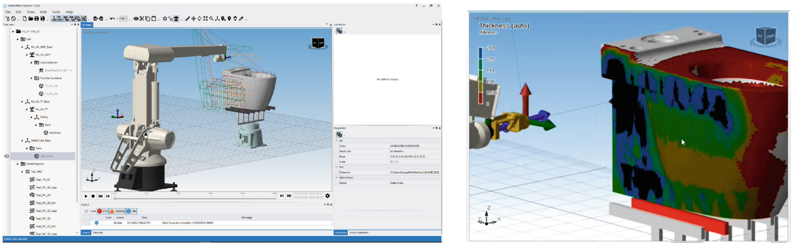
Fig. 2. SmOffNG SW/HW deployment diagram.

4 Applications

As far as the applications are concerned, at the moment SmOffNG is used to develop glazing and superficial finishing programs for applications in the ceramic industry. In the following, specific features related to each task are discussed. With respect to glazing applications, SmOffNG implements two dedicated features: the templatization of the analog outputs and process simulation. The first feature allows the user to create combinations of values of the analog outputs (also referred as technological parameters) so that the combination of values can be assigned to several VPs. In the context of glazing, the following analog outputs have been considered: i) **Flow** (mass flow of the product to be sprayed); ii) **Cap** (atomization pressure); iii) **Fan** (spraying cone opening pressure). Moving to process simulation, Fig. 3(a) shows an example of a glazing program. The simulation is based on a physical model of the spraying process that estimates the amount of material transferred to the work-item on the basis of: i) a probabilistic distribution; ii) the geometry of a spraying cone, which in turn is defined on the basis of the Cap and Fan parameters; iii) some physical parameters of the sprayed fluid, like for instance its density and drying factor. Moving to the superficial finishing task, the distinctive feature implemented by SmOffNG consists in the management of the collisions between the abrasive tools (that are characterized by passive mechanical compensation acting along their approach direction) and the work-item. More in detail:

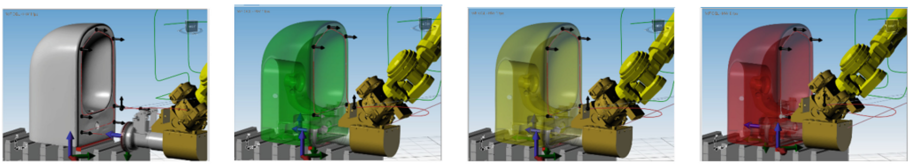
- **No Collision**: no contact between tool and work-item, see Fig. 4(a);
- **Desired Contact**: tool is in contact with the work-item and compensation is either entirely (Fig. 4(b)) or partially (Fig. 4(c)) available;
- **Critical Collision**: tool is in rigid contact with the work-item, i.e. no more compensation is available, see Fig. 4(d);

Finally, these are the technological parameters related to finishing: i) **Rotation Pressure** (pressure to activate rotation of the abrasive tool); ii) **Compensation Pressure** (pressure to activate compensation of the abrasive tool).



(a) (b)

Fig. 3. 3(a) Example of glazing program. VPs are highlighted as black arrows, 3(b) Example of spraying simulation output, with a 5-level thickness gradient on the left.



(a) (b) (c) (d)

Fig. 4. Types of collision checking outputs for compensated tools.

References

Castor, M.: The need for robotic offline programming in a covid-19 world (2020). https://www.controleng.com/articles/the-need-for-robotic-offline-programming-in-a-covid-19-world/

Erdos, G., Paniti, I., Tipary, B.: Transformation of robotic workcells to digital twins. CIRP Ann. **69**(1), 149–152 (2020)

Google. Google Remote Procedure Call (2023). https://grpc.io/

Pan, Z., Polden, J., Larkin, N., Van Duin, S., Norrish, J.: Recent progress on programming methods for industrial robots. Robot. Comput.-Integrat. Manuf. **28**(2), 87–94 (2012)

Ragaglia, M., et al.: SmartOffline NextGen: SW architecture and applications. In: 2022 I-RIM Conference, pp. 16–20 (2022)

Wang, H.F., Zhang, Y.L.: Cad/cam integrated system in collaborative development environment. Robot. Comput.-Integrat. Manuf. **18**(2), 135–145 (2002)

Zheng, C., et al.: Hybrid offline programming method for robotic welding systems. Robot. Comput.-Integrat. Manuf. **73**(102), 238 (2022)

Towards the Adoption of Augmented Reality-Based Digital Twin Approaches in Real Industrial Scenarios

Andrea Di Spigno[1], Alessio Giordano[1], Roberto F. Pitzalis[1,2(✉)], Andrew Cowell[3], Octavian Niculita[3], and Giovanni Berselli[1,2]

[1] Università degli Studi di Genova, 16145 Genova, GE, Italy
{andrea.dispigno,alessio.giordano}@studenti.unige.it, giovanni.berselli@unige.it

[2] IIT - Istituto Italiano di Tecnologia, 16163 Genova, GE, Italy
roberto.pitzalis@iit.it

[3] Glasgow Caledonian University, Glasgow G4 0BA, UK
{a.cowell,octavian.niculita}@gcu.ac.uk

Abstract. In recent years, continuous technological change is influencing and modifying almost all sectors. The industrial one, it can be said, is one of the first to be affected by these changes, through the adoption of new techniques and working methodologies aimed at making all kinds of activities more efficient. This paper focused on researching and implementing enabling technologies of Industry 4.0 in the field of oil and gas industrial plant design and management by presenting and describing two real applications. Augmented Reality and Digital Twin technologies have been implemented, using the software PTC Creo Parametric and Vuforia Studio, with the aim of realising experiences that show the potential of the technology itself and the software that realises its use.

Keywords: industry 4.0 · smart factory · digital twin · augmented reality

1 Overview on Industry 4.0 and Smart Factories

Industry 4.0 refers to the digitization of the industrial sector through the merging of the real and virtual world and the pervasive interconnection of people and things [1]. It is based on the *Smart Factory* paradigm, a new idea of factory that moves away from more traditional automation and towards a fully connected and adaptable system [2]. Nowadays, alongside old technologies, manufacturers are integrating new ones outlining a complex architecture interconnected as a network of devices that can communicate with each other, and exchange data with people through the use of Internet [4]. These technologies include Industrial Internet of Things (IIoT), big data, adaptive robotics, additive manufacturing, virtual simulations, cloud computing and analytics, artificial intelligence (AI) and machine learning (ML), Augmented Reality (AR) and Digital Twins

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 240–245, 2024.
https://doi.org/10.1007/978-3-031-76428-8_45

(DT) [3–6]. They in turn need support by other technologies, such as cyber-security and real-time data management to protect the non-physical part of computing, including data, passwords and identities [7]. As can be seen in the next chapter, all these principles and technologies enable designers, production and maintenance experts to easily predict and monitor progress of industrial process by making plants and parts of operations automatic or quasi-autonomous, remotely doing inspections and tracking all operations performed, thus reducing costs, saving general consumption, and preventing accidents at work [8].

2 Digital Twin and Augmented Reality Experiences in Real Industrial Scenarios

In industrial design constant revisions and changes, at various stages, could be time-consuming, very difficult or sometimes impossible. Thus, the adoption of digital models (DT), fitting the real ones, allows to manipulate the products or their processes before production with less difficulties and costs. In this perspective AR can improve design quality and efficiency by manipulating and interacting with these models displayed into the real world [9]. Vuforia Studio is a tool that permits to design, create and publish AR experiences by overlapping 2D and 3D elements within real world and integrating IIoT data. In manufacturing, this helps live visualisation of internal components or design criteria of a product (e.g. plan how and where to place equipment in relation to space) to quickly update or modify production plans. Hereafter are proposed, in collaboration with Universtiy of Genova and Glasgow Caledonian University, case studies which approach the world of AR-based Digital Twins for smart factories' applications related to maintenance and services.

2.1 Case Study 1 – A Flow Circuit of an Industrial Oil Plant

The following section illustrates a oil flow plant designed to support sensors that can detect the percentage of inhibiting substances (hydrated compounds (THI)) which obstruct the passage of oil in extraction and transportation pipelines. The fluid flow is managed by an EBARA DWOHS/E 200 pump (10 bar) and an Endress and Hauser Proline Promag flow meter (50 W) [10].

The Digital Twin structure consists of components and fieldbus network connections already proposed in [11]. A flow meter measures the fluid flow rate, while pressure transmitter detects the pressure before and after the pump. Thermocouples measure the temperature on different points of the system and on the pump. All sensors are connected to PLC modules via cables, which in turn run an Open Platform Communications (OPC) server and send measurements for remote logging through OPC client connection over industrial ethernet.

To make the experience as immersive as possible for the final user, a detailed CAD model of the plant is created using PTC Creo Parametric software (Fig. 1).

While Vuforia Studio software is used to create the Digital Twin. The experience provides support in the installation, maintenance and management of the

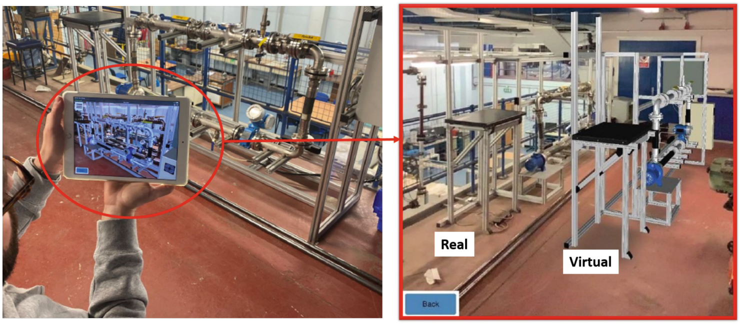


Fig. 1. It shows the real oil plant with its virtual model placed side-by-side.

oil plant through a simple, clear, and easy to use interface. This includes a menu composed of three parts each one linked to specific CAD contents and sensors. The first (*"Whole system"*) relates to the overall system and includes the characteristics of the pump such as technical aspects. The second part (*"Sensors"*) contains the sensors and electrical cabin characteristics. The third part (*"Functionality"*) shows the pressure and temperature gauges data, in real-time or not. After completing the experience, it can be associated to a QRCode. This code can be printed and placed on the object in question for its maintenance. Then it is possible to scan it, using a mobile device, and open Vuforia View application to see the virtual model placed next to the physical one. The user can thus move around the system to see more details of specific components, as shown in Fig. 1. The authors assume this experience is of great support as it provides technical aspects of the asset to better train people working in different departments of a company (such as workers, technicians, engineers, sales managers), or share IIoT data in real-time with custom AR animations for guided services and maintenance.

2.2 Case Study 2 – A Planetary Gearbox

This section refers to the MT 122 unit, a single-stage epicyclic reducer with a transmission ratio of 5:1. It works with a brushed 12 V DC geared motor, a Thurlby LB15 power supply (6–12 V and max 5.5 A), and two tachometers to monitor input and output rotational speeds. The objective is to compare input and output rotational speeds respect to the voltage, at a fixed current of 4 A. A NI 9201 voltage data acquisition module acquires data from the tachometers, and is connected to a CompactDAQ controller NI 9133 to store and share data with a computer. LabVIEW software is used as a graphical interface in which to specify and monitor each input and output quantity (e.g. voltage, current, speeds), while Matlab is used to manage the results.

Realization of a virtual system starts from CAD models using Creo Parametric software. Once the CAD models are assembled, Creo Mechanism is used to simulate the kinematics and dynamics of the mechanism, enabling the validation of the virtual model by finding a theoretical linear function which better fits real measurements This results in a Digital Twin AR system for predictive maintenance through the use of Vuforia Studio software in which the CAD models and sensors are connected via a menu organized in three parts (Fig. 2): the first (*"Details and view specific parts"*) allows to visualise the system and its internal components with all their characteristics. The second (*"Test Rig Simulation"*) shows the internal components during operation. The third (*"Rotation speed results"*) displays the results obtained from the tachometers, at different voltages, to monitor the status of the system and provide alert notification in case problems occur.

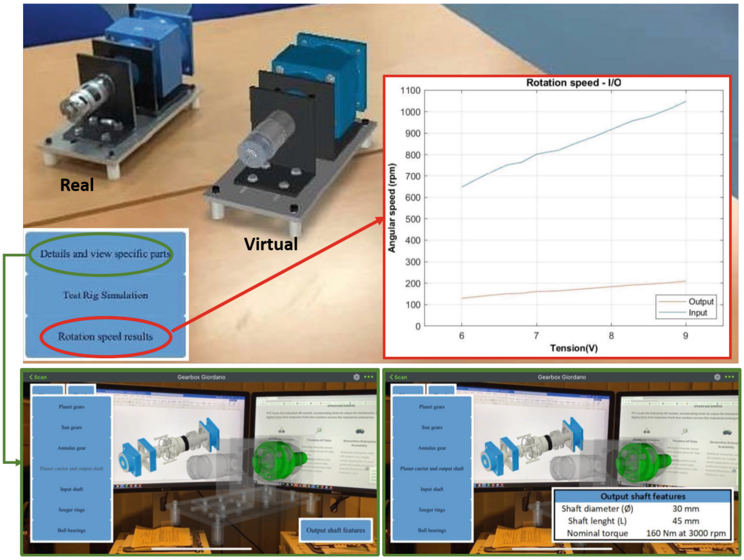


Fig. 2. Results of the real and virtual systems shown in Vuforia View app.

It is essential to scan the QRCode generated by the application and attached to the real system to give access to this experience. Additionally, a flat surface is required to view the test rig model with the three main menus.

The authors believe that the development of such applications can be useful for remote customer support services, providing work instructions, improving the effectiveness of project and processes reviews through knowledge transfer.

3 Conclusions

Key technologies in Industry 4.0 include smart sensors, data, cloud computing, automation, artificial intelligence and many others, all connected as in a dense network which use the same Internet and communication technologies to store, share and transmit data. The projects presented in this paper are intended to show the potential of some these emerging technologies. These have been applied on a oil plant flow loop and on a motorized gearbox, but they could also involve any type of object/physical asset, whether industrial or non-industrial. This because the use of software such as Vuforia Studio allows the creation of content targeting different applications and it is up to the creator of the AR-experience to work out what information to include in response to customers' requirements. According to some estimates, a lot of Universities or companies can benefit from these technologies by leading to a reduction in time to market, customisation costs, maintenance costs and further productivity benefits such as assembly inspection, workers training, a decrease in work-related accidents, travel costs, energy and general consumption [12].

References

1. Kagermann, H.: Change through digitization-Value creation in the age of Industry 4.0. In Management of Permanent Change, pp. 23–45. Springer Fachmedien, Wiesbaden (2014). https://doi.org/10.1007/978-3-658-05014-6_2
2. Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., Yin, B.: Smart factory of industry 4.0: key technologies, application case, and challenges. IEEE Access **6**, 6505–6519 (2017). https://doi.org/10.1109/ACCESS.2017.2783682
3. Wang, L., Wang, G.: Big data in cyber-physical systems, digital manufacturing and industry 4.0. Int. J. Eng. Manuf. (IJEM) **6**(4), 1–8 (2016). https://doi.org/10.5815/ijem.2016.04.01
4. Martinelli, A., Mina, A., Moggi, M.: The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. Ind. Corp. Change **30**(1), 161–188 (2021). https://doi.org/10.1093/icc/dtaa060
5. Schleich, B., Anwer, N., Mathieu, L., Wartzack, S.: Shaping the digital twin for design and production engineering. CIRP Ann. **66**(1), 141–144 (2017). https://doi.org/10.1016/j.cirp.2017.04.040
6. Tao, F., Zhang, H., Liu, A., Nee, A.Y.: Digital twin in industry: state-of-the-art. IEEE Trans. Ind. Inf. **15**(4), 2405–2415 (2018). https://doi.org/10.1109/TII.2018.2873186
7. Gray, G.R.: Blockchain Technology for Managers, pp. 3–184. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-030-85716-5
8. Ustundag, A., Cevikcan, E., Akdil, K. Y., Ustundag, A., Cevikcan, E.: Maturity and readiness model for industry 4.0 strategy. In: Industry 4.0: Managing the Digital Transformation, pp. 61–94. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-57870-5_4
9. An, J., Fan, J.S., Hua, R., Feng, D.P.: Augmented product modeling technique in new product development. Adv. Mater. Res. **156**, 36–41 (2011). https://doi.org/10.4028/www.scientific.net/AMR.156-157.36
10. McDowell, G.R., Uttamlal, M., Holmes-Smith, A.S., Graham, A.: A novel in-line fibre-optic sensor for the detection of hydrate inhibitors within the oil and gas industry. In 2016 IEEE SENSORS, pp. 1–3 (2016). https://doi.org/10.1109/ICSENS.2016.7808603

11. Lindsay, G., Nobakht, B., Marshall, C.: Reducing costs through digitisation and condition based monitoring. In: SPE Offshore Europe Conference and Exhibition. OnePetro (2019). https://doi.org/10.2118/195768-MS
12. Hahn, G.J.: Industry 4.0: a supply chain innovation perspective. Int. J. Prod. Res. **58**(5), 1425–1441 (2020). https://doi.org/10.1080/00207543.2019.1641642

Optimization Framework of a Robotic Pick and Place System for Waste Sorting

Konstantinos Kokkalis(✉), Fotios K. Konstantinidis, Georgios Tsimiklis, and Angelos Amditis

Institute of Communication and Computer Systems (ICCS), National Technical University of Athens, 9 Iroon. Polytechniou Street, 15773 Zografou Athens, Greece
konstantinos.kokkalis@iccs.gr

Abstract. This work proposes an open, distributed framework for planning and controlling robotic systems capable of separating waste items based on their material and/or geometry. A central planner derives valid picking poses and efficiently generates feasible, collision-free trajectories for picking and placing items into the corresponding bins. The planned trajectories are then transferred to a controller tasked with overseeing the overall process, controlling the robot's motion, and managing the gripper. *Robot Operating System (ROS)* serves as the middleware, and *Gazebo* is selected as the simulator, enabling the integration of a wide variety of robots, grippers, path planners, control and planning methods. Finally, numerous valuable metrics are leveraged to demonstrate how different hardware and algorithmic choices of the robotic cell affect its performance, enabling the selection of an optimal configuration.

Keywords: Pick and place · Industrial robotics · Waste sorting

1 Introduction

While EU directives [1] have been encouraging policies to reduce waste generation, little progress has been made in the last decades [2]. Simultaneously, there has been a significant improvement in waste recovery, with a notable 21.5% reduction in waste disposal and a 33.9% increase in recycling. However, the efficiency of recycling processes is largely affected by the homogeneity of the processed streams [3]. Consequently, separating them into different fractions of high purity poses a challenge of significant importance. Traditional recycling methods involve human workers in the initial stages of the sorting task, an unappealing solution due to high costs, biological risks, and human labor shortages. On the other hand, robotic sorting systems have garnered attention in recent years, with many companies introducing commercial products [4]. Most of these automated systems have a high initial cost and operate for limited target waste items, restricting their widespread adoption [5]. Historically, waste sorting systems have comprised three main components: a conveyor for moving waste items, a sensor unit above the conveyor to characterize the waste and a robot for pick & place

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 246–251, 2024.
https://doi.org/10.1007/978-3-031-76428-8_46

motions [6]. Immense advancements in machine learning algorithms and sensors have rendered waste characterization and detection easier [7], while a variety of robots have been used to accelerate the trajectory execution [4]. Although robot planning benchmarking has been conducted in multiple studies [8], these have been application-agnostic, focusing on assessing different planners. Planning for pick and place tasks has attracted attention in the literature [9], mainly studying scheduling rules to optimize throughput and focusing on simple robot classes. Therefore, a gap exists in the literature for benchmarking and evaluating pick and place applications with realistic geometric and kinematic constraints. This work aims to bridge this gap by presenting a framework for the holistic optimization of robotic systems, with the goal of segregating multiple classes of waste more efficiently.

2 Methodology

In this section, the planning/controlling methodology of the robotic waste sorting system will be described in detail, along with the tools used to assess it. However, it is crucial to note that prior to initiating any planning processes, vision-based inference is employed to classify the waste item on the conveyor, determining its position, and extracting other significant geometric features [10].

2.1 Sorting Controller

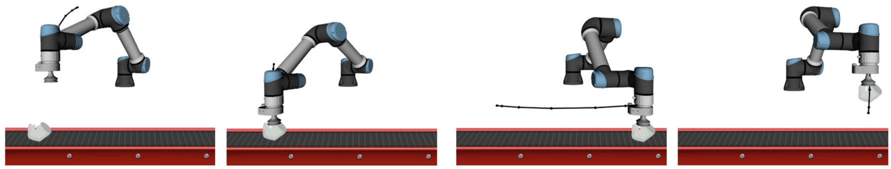
The tasks of a sorting robot are essentially: (i) picking a specific moving item from the conveyor, and (ii) placing it in the appropriate bin. A high-level task controller coordinates the execution of these two tasks in a cycle while being able to handle abnormalities, such as inability to grasp the object due to slip or poor synchronization. This controller processes the inference output and devises a set of candidate end-effector poses for each new item along the conveyor line. These poses, commonly called grasps, also depend on the gripper type (e.g., suction, finger), its properties, the geometry of the object and its material. All of them are cached into a queue and compose the input to the pickup pipeline.

The place pipeline requires the position in Cartesian space that should be reached by the end-effector as input. This position depends on the item class as provided by the inference, and the relation between them is stored in the controller. As soon as a new item is detected, the corresponding place position is retrieved and passed to the place pipeline. For maximum efficiency, motion planning and execution are always running in parallel. Consequently, when a placing trajectory is executed, a trajectory for the pickup of the next item is planned and vice versa. The gripper, like most present grippers, is able to provide feedback on whether the object is picked or released, and the controller utilizes this information to adjust its operation in case of a missed item or an error.

2.2 Planning Pipelines

Given several grasps by the sorter controller, a pipeline of multiple planning stages is deployed to derive a feasible, collision-free trajectory for picking the object from the conveyor. This motion of the end-effector can be divided into four separate parts shown in Fig. 1. The gripper is powered during the linear motion above the conveyor.

A critical factor for the success of the pickup pipeline is the synchronization between the robot and the conveyor. During step (c), the end-effector should be able to move in Cartesian space with a velocity equal to the conveyor velocity. Furthermore, during trajectory planning, the displacement of an object along the conveyor should be projected at a specific point in time. Thus, the conveyor velocity is chosen to be constant and is well-known to the pickup pipeline.



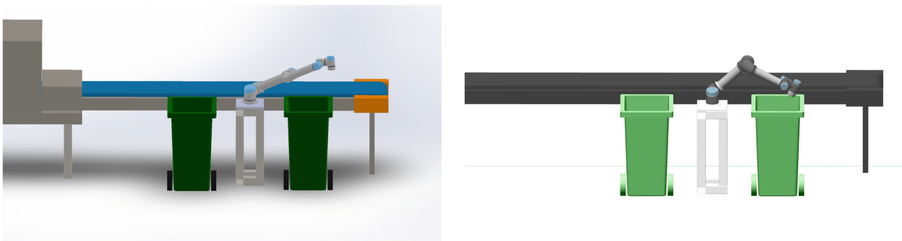
(a) Reach above the conveyor (b) Approach vertically (c) Move linearly along the conveyor (d) Lift the object from the conveyor

Fig. 1. The planned pickup motion

The online picking strategy is based on the FIFO scheduling rule. More precisely, the grasps contained in the queue are evaluated in order, with the oldest evaluated first. Multiple computing threads are used to process more than one grasps concurrently. As soon as a kinematically feasible, collision-free trajectory is found, the rest of the running threads are stopped, and the trajectory is passed to the controller. The place planning pipeline, on the other hand, is simpler and returns as output a trajectory that brings the end-effector above a bin and then optionally approaches and retreats. At the end of the approach motion, the gripper is activated accordingly to release the object into the bin.

2.3 Benchmarking

Since it is now possible to perform multiple cycles of picking and placing waste items, a solid architecture to evaluate the different parameters of the system is required. This architecture should be hardware- and software- independent to assess different robots, grippers, kinematic and planning algorithms, as well as overall cell designs. Meanwhile, modularity is also crucial to be able to adapt to different requirements and environments. These requirements can be fulfilled by ROS, a set of software libraries and tools that can help to build any robot application. Furthermore, *MoveIt!*, a general stack of ROS packages specialized for industrial and collaborative robotic arms, is also utilized as a control and planning framework. The implementation of the planning pipelines and the sorting controller is designed to fit into this flexible framework. The simulation environment chosen for this application is *Gazebo*, which can easily interact with the ROS infrastructure through highly flexible plugins. In Gazebo, physics, kinematics and control of each robot and its environment can be simulated and adjusted to the needs of the waste sorting application. The robot environment in Gazebo and the original CAD design of the robotic waste system are shown in Fig. 2. The waste items moving on the conveyor are simulated using the contours provided by the inference to generate their geometry, while they are spawned in random positions and time intervals using multiple distributions or a Poisson Process. With all the aforementioned tools, a wide variety of metrics are logged during simulation to evaluate the performance of each setup. Metrics such as pick cycles per time unit, pick and place execution time and picked items per class can be assessed offline or online.



(a) CAD designed environment (b) Gazebo simulated environment

Fig. 2. Robot environment for waste sorting in CAD and simulation

3 Results

In this section, multiple hardware and software configurations are examined and compared to demonstrate the flexibility of the framework. For each configuration, three simple metrics are selected: (i) items picked per minute, (ii) mean pickup motion execution time in seconds, (iii) mean pickup planning time in seconds (Table 1).

A series of scenarios are devised by deviating on a single parameter from a baseline scenario. The latter includes a UR5e cobot with a suction gripper mounted on a pedestal near the conveyor, separating objects of two classes in bins to its left and its right. The inverse kinematics are calculated with an analytic solution, while collision-free plans are devised using a deterministic Point-To-Point planner common for industrial robotic arms. Scenario 1 describes the same configuration, but the cobot is mounted above the conveyor. Scenario 2 includes

Table 1. Metrics for designed scenarios

Scenario	Items picked per minute	Pickup execution time (s)	Pickup planning time (s)
Baseline	11.09548	3.33211	0.057559
1	10.26890	4.40681	0.06626
2	17.09232	2.56467	0.14678
3	10.60532	3.59931	0.06113
4	9.37012	3.71946	0.33866

a FANUC M3iA-6S delta robot instead of the cobot, while scenario 3 is the same to the baseline scenario with a two-finger gripper instead of the suction gripper. In Scenario 4 the analytic inverse kinematics solver is replaced with *TRAC-IK*.

4 Conclusions and Future Work

In this paper, we introduced a highly modular and flexible framework for assessing robotic-based waste sorting systems. The framework's value has been demonstrated by evaluating different scenarios, involving diverse algorithmic and hardware selections. Testing and benchmarking the chosen tools enhance the development process, enabling the identification of an optimal configuration customized to the specific requirements of each waste sorting application.

Future work will include the integration of a multi-robot configuration into the existing framework. Additionally, the framework will be evaluated with real hardware controlled by the central controller. Finally, more complex scheduling strategies will be developed and integrated into the planning pipeline.

Acknowledgements. This research was financially supported by the European Union's Horizon Europe research and innovation program under grant agreement No. 101091801 (project DigInTraCE).

References

1. Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives (2008)
2. Eurostat. Generation of waste by waste category, hazardousness and nace rev. 2 activity, 2020. data retrieved from Eurostat on 31/10/2023 (2023)
3. Konstantinidis, F.K., Myrillas, N., Mouroutsos, S.G., Koulouriotis, D., Gasteratos, A.: Assessment of industry 4.0 for modern manufacturing ecosystem: a systematic survey of surveys. Machines **10**(9), 746 (2022)
4. Lukka, T.J., Tossavainen, T., Kujala, J.V., Raiko, T.: Zenrobotics recycler - robotic sorting using machine learning (2014)
5. Kiyokawa, T., Takamatsu, J., Koyanaka, S.: Challenges for future robotic sorters of mixed industrial waste: a survey. IEEE Trans. Autom. Sci. Eng. **21**(1), 1023–1040 (2022)

6. Konstantinidis, F. K., Myrillas, N., Tsintotas, K. A., Mouroutsos, S. G., Gasteratos, A.: A technology maturity assessment framework for industry 5.0 machine vision systems based on systematic literature review in automotive manufacturing. Int. J. Prod. Res., 1–37 (2023)
7. Konstantinidis, F.K., Sifnaios, S., Tsimiklis, G., Mouroutsos, S.G., Amditis, A., Gasteratos, A.: Multi-sensor cyber-physical sorting system (cpss) based on industry 4.0 principles: a multi-functional approach. Procedia Comput. Sci. **217**, 227–237 (2023)
8. Liu, S., Liu, P.: Benchmarking and optimization of robot motion planning with motion planning pipeline. Int. J. Adv. Manuf. Technol. **118**, 949–961 (2022)
9. Mattone, R., Divona, M., Wolf, A.: Sorting of items on a moving conveyor belt. Part 2: performance evaluation and optimization of pick-and-place operations. Robot. Comput.-Integrat. Manuf. **16**(2–3), 81–90 (2000)
10. Konstantinidis, F.K., Sifnaios, S., Arvanitakis, G., Tsimiklis, G., Mouroutsos, S.G., Amditis, A., Gasteratos, A.: Multi-modal sorting in plastic and wood waste streams. Resour. Conservat. Recycl. **199**, 107244 (2023)

Next-Generation Robotic Vision Systems for Quality Assurance in Plastics and Packaging

Alessandro Galdelli[1(✉)], Gagan Narang[1], Adriano Mancini[1], Alessandro Ciancaglione[2], Marco Brutti[2], Gianluca Di Buò[2], and Emanuele Frontoni[3]

[1] Vision Robotics Artifical Intelligence Lab, Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Ancona, Italy
{a.galdelli,a.mancini}@univpm.it, g.narang@pm.univpm.it
[2] Idea, Ancona, Italy
{a.ciancaglione,m.brutti,gianluca}@idea-on-line.it
[3] Università degli Studi di Macerata, Macerata, Italy
emanuele.frontoni@unimc.it

Abstract. This paper introduces an innovative robotic solution to quality assurance and product packaging to reduce human effort while ensuring high-quality results. The proposed system efficiently handles and inspects the product by utilizing the mounted cameras that feed the advanced computer vision models that can detect any fault in the product, which it can then package. A unique feature of our system is the custom-designed palletising island and vacuum gripper to handle products during quality control and packaging. Rigorous testing and experiments with plastic products have shown that our system matches, and in some aspects, surpasses the proficiency of experienced human operators in both efficiency and accuracy.

1 Introduction

In the last decade, the industrial landscape has improved exponentially, benefitting from the incorporation of state-of-the-art technologies like automation, computer vision, and machine vision robotics, among others, catering to the evolving demands of consumers [1]. Industrial manufacturing is undergoing a transformative phase marked by unprecedented growth at an extraordinary pace. There is fierce competition to ensure that the best quality products remain cost-effective along with minimum resource utilisation for sustainability. Large-scale production makes quality control a crucial parameter to secure any product's success. However, traditional manual methods proved to be inadequate in this high-paced environment. Furthermore, rising labour costs and a lack of skilled personnel in recent years have exacerbated the challenges to the manufacturing industry's quality assurance and packaging process. This scenario underscores a pressing need for innovative solutions to navigate these complexities effectively.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 252–257, 2024.
https://doi.org/10.1007/978-3-031-76428-8_47

This paper introduces an autonomous robotic system that meets these critical requirements by combining robust Automated Optical Inspection (AOI) and product packaging processes. The proposed system efficiently handles and inspects the product by utilizing the mounted cameras that feed the advanced computer vision model. Originally introduced in [6,7], this model has undergone transfer learning and fine-tuning to align precisely with our specific objectives, enhancing its capability to detect any faults in the products. A unique feature of our system is the custom-designed palletising island and vacuum gripper, which are integral in handling products with precision during both quality control and packaging stages. After quality inspection, the system employs a robotic arm to package the product, significantly streamlining the final packaging phase. Extensive experiments and testing with plastic products reveal that our novel system offers comparable performance to skilled human operators regarding efficiency and accuracy. The integration of these specialized components ensures a seamless, efficient workflow from inspection to packaging, embodying a significant advancement in robotic automation technology.

2 State of Art

Continuous efforts to enhance and automate quality inspection and packaging in Industry 4.0 have led to the utilization of optical tools and computer technology across many industries. A decade-long meta-review categorized ongoing research on quality inspection workflows based on automation levels, highlighting the transition from static machines to systems with cognitive abilities [2]. Currently, quality inspection and packaging automation progress independently, focusing on fault detection or packaging as separate tasks in distinct production stages. Various studies employ automated fault detection methods tailored to specific fault types. For instance, researchers proposed a deep architecture utilizing YOLO V3 for damage detection, followed by a designed level set algorithm for further investigation [3]. To address the lack of application-centric public datasets, an RGB-D camera-based quality inspection mechanism for food products was tested [4]. This system extracts features from RGB-D images to train a classifier for defect detection, showing successful results across various food products. Additionally, advancements in vision-guided robotics have improved product placement automation on packaging lines [5]. While existing literature showcases promising performance in quality inspection and product packaging systems, a comprehensive examination of a hybrid quality assurance and product placement automation mechanism is lacking. Our system addresses this gap by employing a sophisticated deep computer vision technique for identifying product imperfections, enhancing the quality assurance process. The inclusion of a custom palletising island and vacuum gripper further ensures effective packaging, activated only when the product meets predefined quality standards.

3 Methodology

The proposed AOI and packaging system includes an anthropomorphic robotic arm and three cameras, each with its own illuminator. The Universal Robot UR10, an industrial collaborative robot with a reach of up to 1300 mm and the ability to handle a payload of 12.5 kg, is used for this project. All three cameras feature a Sony IMX264 global shutter sensor, which captures color images at a rate of 7.7 fps and with a resolution of 5 megapixels (2464 × 2056 pixels). Three cameras are used in a quality control system for inspecting a rotating object. Two cameras are positioned 300 mm from the turntable to capture the object from the sides and below, while the third camera is placed approximately 850 mm above the workpiece to provide an overhead view (Fig. 1).

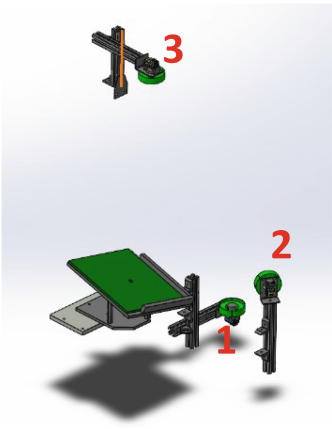


Fig. 1. Schematic positions of the (1) low, (2) side, and (3) top cameras.

An f8 mm manual zoom video lens is used for detecting micro flaws on the plastic surface from 300 mm away, and an f16 mm manual zoom video lens is used for the top-mounted camera. To ensure consistent quality under all lighting conditions, four LED illuminators with white light (6300 K) are used. In Fig. 2, a flowchart is presented to explain the quality analysis and pick-and-place operations. The process begins with the start of AOI and conveyor belt. The vision system detects the box and then performs object detection. The flaw detection object, as introduced in [6,7], has been transfer-learned and fine-tuned to suit the specific objective. The input image is passed through the Blurring Filter, which consists of the Gaussian blur to reduce noise and negligible details, and the Bilateral Filter to remove noise in flat areas. This approach reduces noise without blurring high-frequency details, preserving object clarity. Sequentially following the blurring filter, our fault detection algorithm follows a sequence of steps to identify defects in an image accurately. The Adaptive Threshold algorithm separates the objects of interest from the background based on strong variations in pixel intensity. Then, Morphological Operations (such as Opening for erosion process and dilation, while Closing is a dilation process followed by erosion) is used to remove small objects and holes. The Contour Search algorithm detects objects in the image and establishes the hierarchical relationship between them. It identifies and evaluates object contours in images and checks for containment. Further, to eliminate outliers, three filters are applied - Bounding-Box Aspect Ratio, Bounding-Box Angle, and Defect Area. The result of this process is a set of regions or bounding boxes representing potential defects, with parameters used for classification. These Binary images highlight flaws, with an optimal threshold intensity critical for accurate detection, avoiding both missed flaws and noisy images. In order to ensure proper alignment of the plastic object with the UR10 robot's end-effector, calibration of the robot is initially required. This is performed by using a calibration camera and grid. The calibration phase guides the robot's end-effector to various positions, capturing images to detect the table's centre. Through Singular Value Decomposition (SVD), a transformation matrix adapts the end-effector position to the camera origin, minimising quadratic errors between point sets and obtaining extrinsic parameters. Finally, to achieve our dual objective of quality control and palletising, a specialised vacuum gripper (Fig. 3a) was designed, while Fig. 3b depicts the fully operational system.

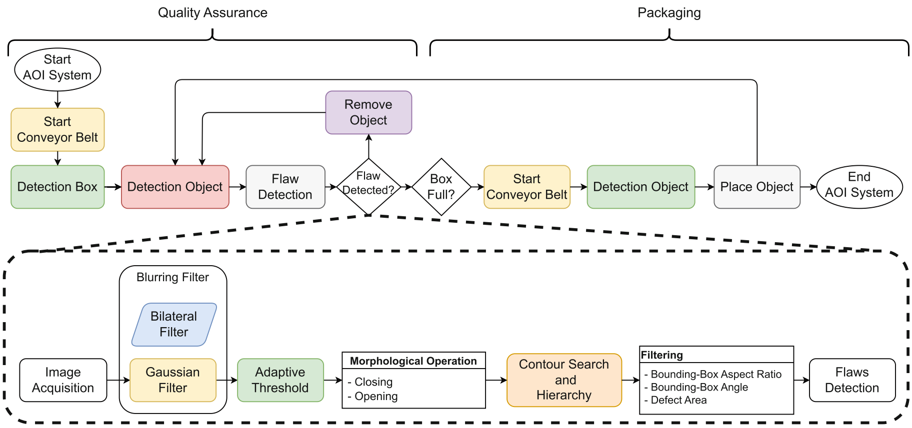
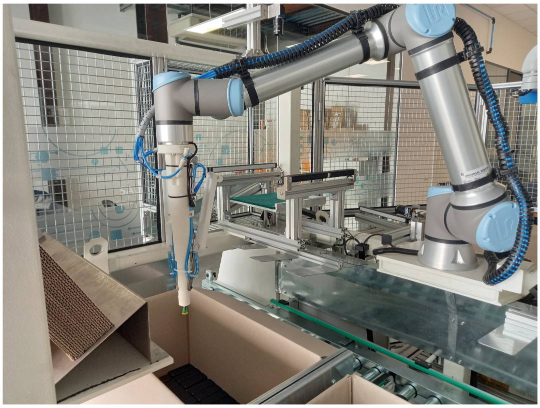


Fig. 2. Key modules in the proposed framework for combined quality analysis and pick-and-place operations.

4 Results and Discussion

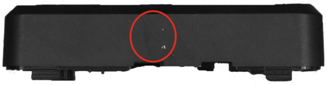
The crucial phase of the proposed AOI system is to implement a quality control and palletising system within an industrial production cycle that is accurate, robust and efficient without causing delays. As of current technology, the AOI system can identify various defects, such as burrs and opacity in certain areas, with a maximum resolution below one millimetre. Figure 4 showcases initial results upon experimentation with plastic covers. The system accurately detects a fault in the plastic cover and removes the cover from the conveyor belt. It immediately iterates to another object and continues the loop without human intervention. Maintaining a consistent duration of 15.2 s across various tests, the complete quality control and object placement process involves a 3.2 s duration for the conveyor belt operation, 8.3 s for quality control, and 3.7 s for placing the object. This duration significantly outpaces human capabilities, accelerating production and preserving resources as the operator can be reassigned to less repetitive tasks.

(a) Proposed gripper.



(b) AOI proposed system.

Fig. 3. a) showcases the envisioned robotic hand, whereas b) reveals the complete system in action.



(a) Plastic cover.

(b) Mask.

Fig. 4. a) Plastic cover and b) mask obtained from our proposed detection flaw algorithm (red circle highlights the flaws detected).

5 Conclusion

This research paves the way for a new era in industrial manufacturing, where intelligent automation plays a pivotal role in ensuring product quality, operational efficiency, and sustainability. Our system is innovatively designed with state-of-the-art image processing technology and a robotic arm, featuring a specialized vacuum gripper. This combination not only allows for effective inspection of plastic products for any faults but also ensures their precise and efficient packaging, streamlining the entire process with remarkable accuracy and speed. This autonomous robotic system offers a sustainable solution by eliminating the limitations of traditional manual methods and overcoming challenges related to a shortage of skilled personnel.

References

1. Reyna, M., Delgado, G., Akundi, A., Luna, S., Chumacero, E: Product digital quality inspection using machine vision systems–a categorical review. In: 2022 17th Annual System of Systems Engineering Conference, pp. 37–42. IEEE (2022)

2. Azamfirei, V., Psarommatis, F., Lagrosen, Y.: Application of automation for in-line quality inspection, a zero-defect manufacturing approach. J. Manuf. Syst. **67**, 1–22 (2023)
3. Tao, J., Chen, Q., Xu, J., Zhao, H., Song, S: Utilization of both machine vision and robotics technologies in assisting quality inspection and testing. Math. Prob. Eng. (2022)
4. Guo, Z., Zhang, M., Lee, D.J., Simons, T.: Smart camera for quality inspection and grading of food products. Electronics **9**(3), 505 (2020)
5. Naeem, M., Aslam, S., Suhaib, M., Gul, S., Murtaza, Z., Khan, M.J.: Design and implementation of pick and place manipulation system for industrial automation. In: International Conference on Artificial Intelligence and Mechatronics Systems (2021)
6. Galdelli, A., et al: A novel remote visual inspection system for bridge predictive maintenance. Remote Sens. **14**(9), 2248 (2022)
7. Galdelli, A., Pagnotta, D.P., Mancini, A., Freddi, A., Monteriù, A., Frontoni, E.: Empowered optical inspection by using robotic manipulator in industrial applications. In: International Conference on Intelligent Robots and Systems, Macau (2019)

Evaluating Xenomai and KVM for Real-Time Virtualization in Industrial Automation

Andrea Testa[1(✉)], Marco Valli[2], Gianluca Palli[1], and Ivan Ragazzini[2]

[1] Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy
{a.testa,gianluca.palli}@unibo.it

[2] IMA S.p.A., Ozzano dell'emilia, Bologna, Italy
{marco.valli,ivan.ragazzini}@ima.it

Abstract. During the last years, industrial automation is more and more witnessing the usage of virtualization technologies to run real-time tasks. On the one hand, virtualization is a consolidated technology that aims at improving the manageability and scalability of industrial control systems. On the other hand, real-time tasks often require strict timing guarantees and determinism in the response time. In this paper, we consider a real-time virtualization platform based on the Xenomai real-time operating system and the KVM virtualization technology. The chosen technologies are open-source and fully compatible with Linux-based systems and ensure seamless integration and coexistence with existing applications. We assess the performance of the proposed architecture with a set of benchmarking experiments run on a B&R industrial computer.

Keywords: Industrial automation · real-time operating systems · virtualization

1 Introduction

Real-time operating systems (RTOSs) play a key-role in several scenarios, including industrial control, automotive and robotics. During the last years, a plethora of native, real-time operating systems, such as VxWorks, NuttX and Zephyr, has been proposed to ensure determinism and reliability for these tasks. Yet, the Linux Kernel is receiving more and more attention thanks to its support for several hardware devices and peripherals and for its software community. Thus, several technologies have been proposed to patch general-purpose, Linux-based systems into real-time systems. A notable example is the `PREEMPT_RT` patch, that is being widely employed in several industrial settings [1].

One of the alternatives to the PREEMPT_RT patch is the Xenomai framework [2]. Xenomai employs a co-kernel approach, and has been shown to provide better response-time to eternal events with respect to the PREEMPT_RT kernel patch [3]. Xenomai is being more and more employed in several control and

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 258–262, 2024.
https://doi.org/10.1007/978-3-031-76428-8_48

robotics settings [4,5]. It has been also combined with the well-known Robot Operating System (ROS), [6,7], and with PLC programming standards, [8]. Industrial control is also recently experiencing the adoption of virtualization techniques [9], in which the control application is implementend in an operating system compounded into a virtual machine [10] or into a container [11,12]. In the context of industrial control, a comparison between Xenomai and PREEMPT_RT with and without VirtualBox and an Amazon Web Services hypervisors is provided in [13]. A benchmark of the Kernel based Virtual Machine (KVM) hypervisor for self-driving tasks is discussed in [14], while a comparison between KVM and the Xen hypervisor is proposed in [15]. Authors in [16] benchmark the NuttX RTOS when executed over the Jailhouse hypervisor.

Virtualization of real-time systems offers a wide range of benefits, as flexibility and development cost reduction. In this work, we provide a virtualization platform to run real-time tasks in a virtualized environment. The proposed platform is based on the Xenomai RTOS and on the KVM virtualization technology, is open source and compatible with classical Linux-based systems. This ease the virtualization of real-time control processes while allowing to reuse existing codebases. We benchmark the proposed methodoloy over a B&R Industrial Computer. As a baseline, we consider the setup in the RTOS is executed directly on the hardware. Then, we consider the setting in which the RTOS is run as a guest operating system inside KVM, leveraging a general-purpose operating system as the host. Finally, we consider the setting in which both the host and the guest operating systems provide real-time capabilities. To the best of the authors' knowlegde, the combination of these technologies in industrial control settings has not been investigated in the literature.

2 A Virtualization Platform Architecture

In the considered architecture (we refer to Fig. 1 for a sketch), we make use of the KVM hypervisor. As we detail next, KVM is integrated into the Linux kernel, and virtual machines are treated as Linux processes. Thereby, in the proposed platform, a general-purpose operating system (Ubuntu 20.04) in our case, acts as a bare-metal hypervisor. In a KVM virtual machine, we execute the Xenomai RTOS, also detailed next. It is worth noting that the choice of the host kernel has an impact on the response time of the virtualized RTOS.

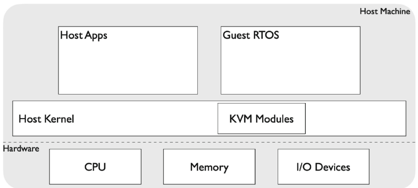


Fig. 1. Snapshot of the considered virtualization architecture and its main components.

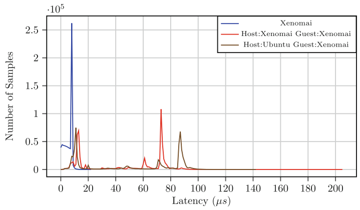
We now provide more details on the Xenomai RTOS and KVM. Xenomai is a real-time framework employing a dual-kernel approach. That is, a real-time kernel runs alongside the standard Linux one. A hardware abstraction layer called ADEOS handles the coexistence of these two kernels. To ensure deterministic, real-time performance, Linux tasks can be executed only if there are no real-time tasks to be served. To ease development in the Linux environment, Xenomai comes with a set of APIs, also called skins. These skins act as interfaces to the real-time kernel, so that users can easily implement real-time tasks. Xenomai also comes with a Posix skin, allowing to reuse Posix-compliant code. For our experimental results, we make use of this feature to compile a well-known benchmark software in order to run on the Xenomai real-time co-kernel. The Kernel based Virtual Machine (KVM) is a virtualization technology directly integrated into the Linux kernel. Thus, it is open-source and already available in Linux-based systems. Moreover, it benefits from all the updates and improvements released by the Linux development team, and can be run on top of all Linux-certified hardwares. Thanks to its integration with the Linux kernel, it can also exploit real-time extensions of the kernel to improve its response time. In Sect. 3, we will show how the performance of KVM varies when changing the underlying kernel, and how this aspect can be exploited to improve real-time capabilities in terms of response time and latency.

3 Benchmarking Tools and Results

We assess the performance of the proposed platform on a B&R industrial platform. We make use of an APC910 computer, with a 2.7 GHz Intel i5 quad-core CPU (QM170 chipset), equipped with two 8 GB DDR4 RAM modules. The platform is installed on a 32GB CFAST memory. As host operating system, we choose Ubuntu 20.04. As guest RTOS instead, we pick Xenomai 3.1. We isolated the virtual machines on two of the four cores, so that the host OS cannot interfere with the guest OS tasks. Moreover, we give the KVM vcpu the SCHED_FIFO policy and highest possible priority. This choice sensibly reduces worst-case latencies of the guest RTOS. In order to assess the performance of our system, we make use of the well-known `rt-tests` suite. This suite is made of a set of different programs, aiming at measuring different quantities of interest in a RTOS. In this work, we aim at assessing the *latency* and jitter performance of the RTOS kernel under different conditions. To this end, we make use of the `cyclictest` program contained in `rt-tests`. `cyclictest` generates a periodic real-time thread, with a period specified by the user. This thread may simulate, e.g., a control task. `cyclictest` measures the difference between the thread expected wake-up times and the actual wake-up times. In this way, the program captures the effects of different sources of delay. This procedure is repeated multiple-time in order to capture a statistically relevant description on the latency of the system. The benchmark tests have been run on one of the two cores assigned to the RTOS. We consider the following two settings. In the first one, the host kernel is the Linux kernel delivered with Ubuntu. In the second one, the host has a dual-kernel configuration with Xenomai 3.1. As a baseline, the same tests have also been executed on a Xenomai RTOS running as host RTOS on the same hardware. The results are reported in Table 1 and Fig. 2. In Table 1 we depict minimum, maximum, average and the standard deviation of the performed tests. The value 0 means that the latency was less than $1 \mu s$, which is the default clock tolerance for `cyclictest`. Figure 2 (left) depicts how many times a certain latency as occurred during the benchmarking. The blue line represents the baseline, the red line the setup in which both host and guest are real-time systems while the brown one represents the scenario in which the host is a general-purpose operating system. In Fig. 2 (right) instead, the same data are reorganized in a boxplot fashion, in order to better highlight mean and jitter of the latencies. The setup in which both the host and guest are RTOSs comes with smaller latencies. However, it comes with a set of high-latency, spurious values. Instead, the setup in which the guest is a general-purpose OS has higher average latencies.

Table 1. Numerical Results

Setting	min (μs)	max (μs)	avg (μs)	std (μs)
Xenomai	0	22	5	2.91
Host: Xenomai Guest: Xenomai	0	205	45	28.89
Host: Ubuntu Guest: Xenomai	0	142	51	34.82



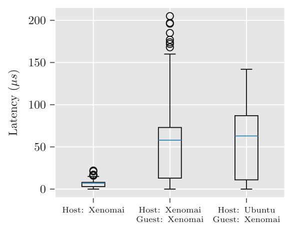


Fig. 2. (left) Amount of samples for which a certain latency was measured in all the considered settings. (right). Boxplot results for the different settings.

4 Conclusions and Future Work

We proposed a virtualization platform to run real-time activities in a virtual machine with real-time capabilities. The platform is based on the Xenomai and KVM open source solutions. We provided experimental benchmarking results on an industrial platform. The results show that the proposed virtualization environment provides deterministic responses, and is amenable to be employed in a real application. Future work includes testing the platform on a real control task, as well as exploring the performance of novel versions of Xenomai.

References

1. Reghenzani, F., Massari, G., Fornaciari, W.: The real-time linux kernel: a survey on preempt_rt. ACM Comput. Surv. (CSUR) **52**(1), 1–36 (2019)
2. Gerum, P.: Xenomai-implementing a rtos emulation framework on gnu/linux. White Paper, Xenomai, p. 81 (2004)
3. Huang, C., Lin, C.H., Wu, C.K.: Performance evaluation of xenomai 3. In: Proceedings of the 17th Real-Time Linux Workshop (RTLWS), pp. 21–22 (2015)
4. Laurenzi, A., Antonucci, D., Tsagarakis, N.G., Muratore, L.: The xbot2 real-time middleware for robotics. Robot. Autonom. Syst. **163**, 104,379 (2023)
5. Boltov, Y., Skarga-Bandurova, I., Kotsiuba, I., Hrushka, M., Krivoulya, G., Siriak, R.: Performance evaluation of real-time system for vision-based navigation of small autonomous mobile robots. In: IEEE International Conference on Dependable Systems, Services and Technologies, pp. 218–222 (2019)
6. Delgado, R., You, B.J., Choi, B.W.: Real-time control architecture based on Xenomai using ROS packages for a service robot. J. Syst. Softw. **151**, 8–19 (2019)
7. Zhang, G., Li, Z., Ni, F., Liu, H.: A real-time robot control framework using ROS control for 7-DOF light-weight robot. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1574–1579 (2019)
8. Kim, Y., Lee, S.Y., Lim, S.: Implementation of plc controller connected gazebo-ros to support IEC 61131-3. In: IEEE International Conference on Emerging Technologies and Factory Automation, vol. 1, pp. 1195–1198 (2020)
9. Lozano, S., Lugo, T., Carretero, J.: A comprehensive survey on the use of hypervisors in safety-critical systems. IEEE Access (2023)
10. Scordino, C., et al.: Real-time virtualization for industrial automation. In: IEEE International Conference on Emerging Technologies and Factory Automation, vol. 1, pp. 353–360 (2020)
11. Walker, M., Fischer, M., Lechler, A., Verl, A.: Evaluation of isolation and communication mechanisms for real-time containers. In: IEEE International Symposium on Industrial Electronics, pp. 1–8 (2023)
12. Sollfrank, M., Loch, F., Denteneer, S., Vogel-Heuser, B.: Evaluating docker for lightweight virtualization of distributed and time-sensitive applications in industrial automation. IEEE Tran. Ind. Inf. **17**(5), 3566–3576 (2020)
13. Hofer, F., Sehr, M.A., Iannopollo, A., Ugalde, I., Sangiovanni-Vincentelli, A., Russo, B.: Industrial control via application containers: Migrating from bare-metal to IAAS. In: IEEE International Conference on Cloud Computing Technology and Science, pp. 62–69 (2019)
14. Müller, T., Askaripoor, H., Knoll, A.: Performance analysis of kvm hypervisor using a self-driving developer kit. In: Annual Conference of the IEEE Industrial Electronics Society, pp. 1–7 (2022)
15. Abeni, L., Faggioli, D.: Using xen and kvm as real-time hypervisors. J. Syst. Archit. **106**, 101,709 (2020)
16. Yang, C.F., Shinjo, Y.: Obtaining hard real-time performance and rich linux features in a compounded real-time operating system by a partitioning hypervisor. In: ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, pp. 59–72 (2020)

Evaluating Image-Based Visual Servoing Techniques for Robotic Manipulation In Space

Lina María Amaya-Mejía[1,2(✉)], Andrej Orsula[1], Mohamed Ghita[2], Miguel Olivares-Mendez[1], and Carol Martinez[1]

[1] Space Robotics (SpaceR) Research Group, Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Esch-sur-Alzette, Luxembourg
lina.amaya.001@student.uni.lu

[2] Redwire Space Europe, Esch-sur-Alzette, Luxembourg

Abstract. In-Space Servicing, Assembly, and Manufacturing (ISAM) capabilities will power the next big step for enabling sustainable exploration and commercialization of space. Making these capabilities autonomous is a priority for the space industry. Visual Servoing (VS) is a promising approach to enable autonomy by using visual information to control the motion of a robot. This work presents a comparative study of four image-based VS (IBVS) approaches, evaluating their potential for autonomous robotic manipulation tasks in space; while demonstrating ISAM capabilities in a complex assembly scenario.

Keywords: Visual Servoing · Robotic Manipulation · Space Robotics

1 Introduction

Robotic arms are essential for a multitude of extraterrestrial applications. Planetary rovers (Fig. 1) use robotic arms to gather scientific data and handle samples. Similarly, orbital robotic arms semi-autonomously assist operations at the International Space Station (ISS) for berthing, inspection, and assembly (Fig. 2). Most of the existing manipulation systems in space need to be teleoperated by astronauts or ground controllers, which can be challenging due to communication delays and the need for highly skilled operators. Moreover, the fast-growing in-space industrialization efforts require autonomous manipulation systems to adapt to dynamic scenarios with minimal human intervention [1].

Developing visuomotor skills for space robotic manipulators can significantly enhance autonomous operations. These skills enable robots to recognize and track objects, as well as to navigate through complex and dynamic environments, enhancing flexibility and robustness when executing manipulation tasks. A robot can acquire visuomotor skills by using Visual Servoing (VS) strategies to control the robot's relative motion based on visual observations. The resulting increase

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 263–268, 2024.
https://doi.org/10.1007/978-3-031-76428-8_49

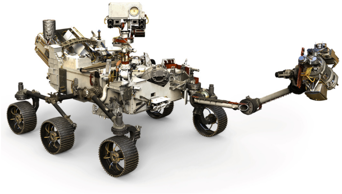


Fig. 1. Perseverance Rover (src: NASA)

Fig. 2. Canadarm2 and Dextre (src: CSA)

in precision will enable critical in-orbit servicing missions, such as refuelling, repositioning, component replacement, or repairing a failing satellite.

This work presents the comparison of four IBVS techniques that could be used to enhance autonomous space robotic manipulation. We evaluate different depth estimation methods, sensor modalities, features, and control laws in a complex roto-translational scenario. Additionally, we evaluate ISAM capabilities with an assembly scenario.

2 Visual Servoing

VS enables a robot to interact with its environment by controlling its motion based on visual feedback [2]. The latter may be acquired from a camera mounted on a robot manipulator (eye-in-hand) or fixed in the workspace (eye-to-hand). As shown in (1), a proportional controller typically defines the VS control scheme.

$$V_c = -\lambda \hat{L_e^+} e \tag{1}$$

where the visual features error e drives the camera velocity V_c. λ is a constant gain and $\hat{L_e^+}$ is the approximation of the Moore-Penrose pseudo-inverse of the interaction matrix L_e. L_e is constructed depending on the selection of the visual features and depth information [1]. Three popular choices exist: 1) Using current features when depth of each point is always known, 2) Using the desired features when only desired point depths are available, 3) Using the mean of desired and current features (which provides superior performance [2] and therefore, it is the one used in this work).

The feature space guiding the VS control law defines two classes: Position-Based Visual Servoing (PBVS), where the error is in 3D Cartesian space; and Image-Based Visual Servoing (IBVS), where the error is in 2D image space. Current VS methods for space object manipulation and non-tumbling spacecraft docking rely on classical PBVS [4,6] and IBVS [3,5] approaches. Advanced IBVS techniques like 2.5D VS and partitioned IBVS, use both 2D and 3D features or additional 2D features to decouple translation and rotation control [2]. IBVS approaches are known for their robustness against robot and camera model uncertainties, and for their ability to keep the target within the camera's field of view, compared to PBVS. This motivated our comparative study to explore their potential in enhancing autonomous space robotic manipulation.

All IBVS approaches require depth information that is not directly obtained from image measurements in order to construct L_e. Depth can be estimated either by using geometrical methods or measured directly from a depth sensor. Accurate depth estimation is crucial in VS as it affects the camera's motion during task execution. Therefore, the comparative study presented here analyses various IBVS methods and explores different depth data acquisition techniques for in-space robotic manipulation.

3 Comparative Study of IBVS Approaches

A 7-DoF Kinova Gen2 robot with an Intel RealSense L515 depth camera mounted on its end-effector, aiming to reach a pose over an AprilTag marker (Fig. 3), is used for the assessment. The visual features to control are the four corners of the marker. The ViSP C++ library [7] is used to design each IBVS control law.

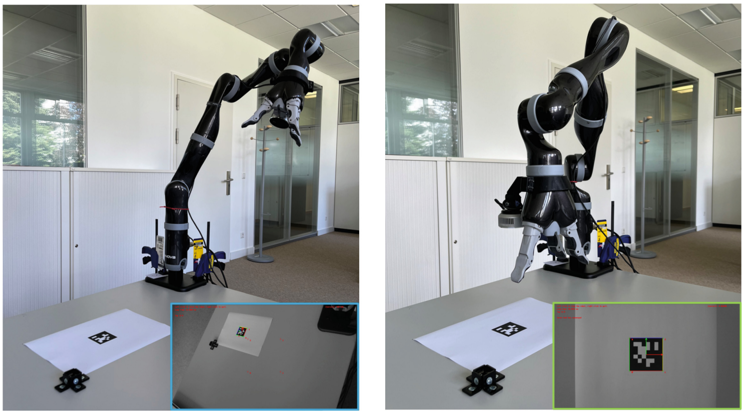


Fig. 3. Initial and desired poses and visual features

Table 1 shows the tested VS approaches. IBVS uses only image features to control all camera motions, 2.5D VS combines 2D and 3D features to decouple rotation and translation, and partitioned IBVS decouples motions in the Z axis using only image features. Depth estimation is done either by homography [2] (H) or directly from camera's depth sensor (D). These approaches have unique strengths and weaknesses since they were designed to solve different tasks. Therefore, the goal is not to rank them but to establish a basis for comparison to select the appropriate system for space assembly tasks based on specific requirements.

Table 1. Features of IBVS approaches tested

Control Scheme	Features	Depth Estimation
IBVS+H	– Image points (x, y)	H
IBVS+D	– Image points (x, y)	D
2.5D VS	– Extended image points $(x, y, logZ)$ for translation	H
	– Camera-object rotation (θ_u) for rotation	
Partitioned IBVS	– Image points (x, y) for XY motion	H
	– Polygon area (σ) for Z translation	
	– Angle (α) between horizontal axis and a segment for Z rotation	

For a fair comparison, a complex scenario was created introducing pose errors across all 3D axes. The most significant errors we included to assess the approaches were along the Z-axis, as IBVS relies on depth estimation and faces challenges in handling large rotation errors. The errors between initial and desired poses (Fig. 3) were $(e_t x : -0.12, e_t y : -0.13, e_t z : 0.50)$ meters for translation and $(e_r x : -16.22, e_r y : -21.88, e_r z : 68.51)$ degrees for rotation.

4 Results and Analysis

Each VS approach was tested three times. The system gains (λ) were adjusted to avoid oscillations and ensure convergence (average error of the feature points is < 1 pixel) within 1500 iterations. For the qualitative assessment, the visual feature trajectories and the camera 3D trajectory were recorded, while for the quantitative evaluation, the following metrics were used: 1) Number of iterations to converge, 2) Average pixel error at convergence, 3) Camera 3D position, and 4) Average time of each iteration. Table 2 shows the main results per method.

Table 2. Performance metrics results for each IBVS approach

Control scheme	Gain (λ)	Error (px)	Iter. until convergence	Max. deviation (px)	Max. camera deviation (m)	Iter. time (ms)
IBVS + H	11	0.797	614	323	0.206	32.9
IBVS + D	11	0.980	1320	323	0.203	46.3
2.5D VS	9.5	0.913	618	233	0.177	34.9
Partitioned IBVS	9	0.834	622	378	0.292	34.9

The IBVS+H demonstrated the fastest convergence with the smallest error, indicating its effectiveness in complex scenarios. The 2.5D VS approach excelled at keeping features centered within the camera's field of view while minimizing camera and feature deviations. On the other hand, the IBVS+D showed slower performance, and the partitioned IBVS method showed significant deviations in both camera and feature trajectories. Nevertheless, all methods guided the error < 1 pixel with similar real-time performance (see videos here).

Assembly Scenario: The 2.5D VS method was chosen for the scenario. It demonstrated key characteristics for space robotic manipulation, including rapid convergence, low error, controlled camera and feature trajectories, and efficient iteration times. The task entails executing a sequence of actions (pre-grasping, grasping, and post-grasping motions) over a beam as part of an assembly task of a hexagonal module, which is commonly employed in space structures. Figure 4 shows the poses and the corresponding visual features for each action (see video here).

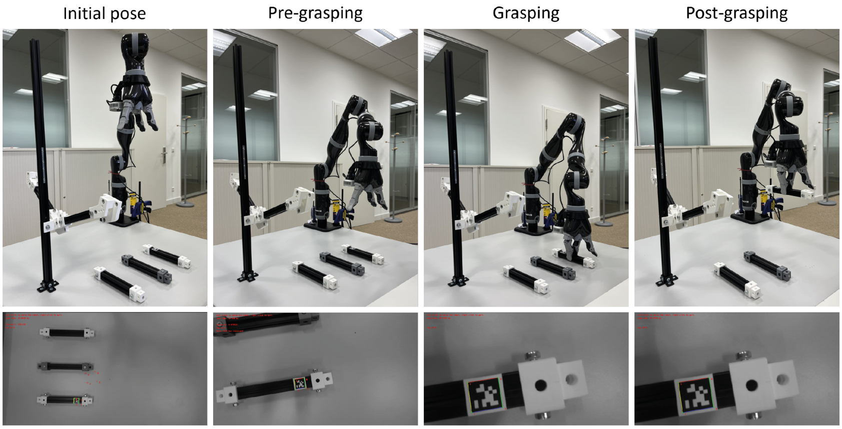


Fig. 4. External and camera view of an assembly task using 2.5D VS

Figure 5 shows the features (left plot) and camera (right plot) trajectories. Feature points remained within the camera's field of view. The right plot shows that the 2.5D VS strategy separates translational from rotational motions. First, it reduced the translational error; at the end, the rotation was corrected.

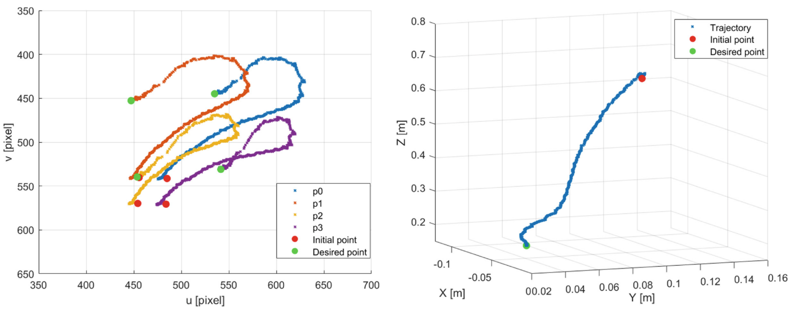


Fig. 5. 2.5D VS: Assembly task visual features and camera trajectories

5 Conclusions and Future Work

In response to the rapidly expanding in-space industrialization that requires manipulation systems capable of handling complex and dynamic scenarios, we conducted a comparative study of four IBVS techniques and demonstrated their potential for autonomous robotic manipulation tasks in space. The 2.5D VS method's rapid convergence, as well as minimal camera and feature motions make it well-suited for in-orbit operations, like autonomous assembly tasks. Future work includes exploring hybrid VS strategies, testing them in space-relevant scenarios, and identifying potential challenges for space missions.

References

1. Papadopoulos, E., Aghili, F., Ma, O., Lampariello, R.: Robotic manipulation and capture in space: a survey. Front. Robot. AI. **8** (2021)
2. Siciliano, B.: Springer Handbook of Robotics (2016)
3. Ramón, J., Pomares, J., Felicetti, L.: Direct visual servoing and interaction control for a two-arms on-orbit servicing spacecraft. Acta Astronautica **192** (2022)
4. Liang, C., Wang, R., Liu, X.: A visual-based servo control method for space manipulator assisted docking. In: 2022 IEEE 6th Information Technology And Mechatronics Engineering Conference (ITOEC) (2022)
5. Shademan, A., Farahmand, A., Jägersand, M.: Robust uncalibrated visual servoing for autonomous on-orbit-servicing (2010)
6. Wang, C., Wang, R., Mandlekar, A., Li, F., Savarese, S., Xu, D.: Generalization through hand-eye coordination: an action space for learning spatially-invariant visuomotor control. ArXiv (Cornell University) (2021)
7. Visual Servoing Platform - ViSP https://visp.inria.fr/visual-servoing/

Efficient Deadlock Detection and Resolution Algorithm for AGV Fleet Management

Alessandro Bonetti[1(✉)], Simone Guidetti[2], and Lorenzo Sabattini[1]

[1] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Via Amendola 2, Pad. Morselli, 42122 Reggio Emilia, Italy
alessandro.bonetti@unimore.it

[2] Gruppo TecnoFerrari S.p.a. con socio unico, Via Ghiarola Nuova, 105, 41042 Fiorano Modenese, MO, Italy
http://www.tecnoferrari.it/

Abstract. A well-known challenge in managing fleets of Automated Guided Vehicles (AGVs) is the issue of deadlocks. A deadlock occurs when two or more AGVs become mutually blocked, rendering the completion of their respective missions impossible. This article introduces a novel algorithm designed to detect and resolve deadlocks. The detection of deadlocks is achieved by employing the well-established Depth First algorithm, which analyzes the precedence relationships among the vehicles assigned by the traffic manager. If the deadlock condition is identified, a Conflict Based Search (CBS) algorithm is initialized for the involved vehicles. CBS is applied to the road map to find new paths in space-time without conflicts, effectively resolving the deadlock. Thus, some examples of deadlocks that the algorithm succeeds in resolving are presented. By implementing this strategy, AGV fleets can mitigate the degradation of performance caused by the occurrence of deadlocks and ensure smoother operations in automated logistics.

Keywords: Deadlock · AGV · Traffic management

1 Introduction

In automated logistics, Automated Guided Vehicles (AGVs) improve material handling and transportation efficiency. However, a key challenge in AGV systems is deadlock, which is a situation in which two or more vehicles are trapped in a blocking scenario, rendering them incapable of advancing in their missions.

Deadlock scenarios can be categorized as either cyclic or acyclic [4] based on how the resources (spatial parts of the workspace) are locked between the participating AGVs. Cyclic deadlocks occur when the immobilized vehicles exhibit

This work was supported by the AI-DROW Project through the Italian Ministry for University and Research under the PRIN 2022 program, funded by the European Union – Next Generation EU.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 269–274, 2024.
https://doi.org/10.1007/978-3-031-76428-8_50

cyclic precedence structure. This situation is characterized by the presence of a chain of AGVs, where each of them locks in the resources associated with the portion of space another vehicle in the chain wants to reserve. Acyclic deadlocks are scenarios in which the precedence structure can be represented with a tree or directed acyclic graph (DAG). They typically arise when a vehicle is halted due to an error and it encumbers the next portion of space another vehicle wants to traverse, so no cycle structure is formed.

Numerous studies have explored the challenge of managing deadlocks in the context of centralized AGV systems. For instance, in a study by the authors in [5], the transport road network is partitioned into non-overlapping zones. They introduce a Structural Online Control Policy (SOCP) to ensure that vehicle transitions between zones do not lead to deadlock situations. Additionally, in the study [1], an event-triggered colored elementary net (ETCEN) is employed to illustrate how AGVs occupy spatial resources, facilitating the avoidance and resolution of deadlocks. In another notable contribution [2], a time-expanded graph is used to model precedence relationships among vehicles, enabling the prediction of deadlock formation and attempts to resolve it by reordering precedence relationships.

In this paper, we present a novel solution for detecting and resolving deadlocks in AGV systems. In Sect. 2, we describe the traffic management feature in our case study and the method used for detecting deadlocks. In Sect. 3, we introduce the strategy for resolving deadlocks using the Conflict Based Search algorithm. Additionally, we provide two highly representative examples of AGV fleet system deadlocks and demonstrate how our proposed solution effectively resolves them.

2 Traffic Management System and Deadlock Detection

In our traffic management system, vehicles traverse a road map consisting of curve segments connected to each other at their external points. Each vehicle follows a predetermined path associated with its mission. To ensure efficient coordination and collision avoidance, the traffic manager uses space-time precedence assignment rules on the curved segments of the road map. When a vehicle is assigned a segment, it travels along it until reaching its endpoint. Moreover, vehicles can stop only at points and not within segments.

Precedence rules are critical for assigning the order in which vehicles traverse the curve segments. However, these rules may not always have universal applicability, and unexpected deadlocks can arise. In our traffic management system, a deadlock occurs when a cycle of precedence dependencies forms among two or more vehicles. This situation prevents the affected vehicles from progressing along their routes.

To resolve such deadlocks, it becomes necessary to re-plan the paths of one or more agents. The goal is to find compatible paths in space-time, ensuring that a solution exists for the Multi-Agent Path Finding (MAPF) problem. Moreover, using the precedence assignments made by the traffic management system, we can construct a precedence graph, denoted as G, at the current time instant. In this graph, each node represents an agent, and directed arcs indicate the precedence relationship between two AGVs. Once the precedence graph has been constructed, we employ the widely used Depth-First Search (DFS) algorithm, a common technique for detecting deadlocks in various systems, including operating system processes.

DFS operates by starting from an initial node and recursively delving as deeply as possible along a single branch before backtracking. It efficiently explores graphs and trees and keeps track of the visited nodes in a list. If a node to be visited is already in that list, a cycle exists. In our case, this algorithm serves the specific purpose of identifying a cycle of precedences.

It is important to note that resolving these cyclic deadlocks effectively addresses the vast majority of blocking situations that a traffic manager might encounter. Therefore, the identification and resolution of these specific deadlocks play a crucial role in maintaining smooth traffic flow within the system.

3 Deadlock Resolution

This chapter describes the conflict resolving algorithm we propose, which is based on the Conflict Based Search algorithm. After that, we will describe two examples of very common deadlocks in our Automated Guided Vehicle systems.

Conflict Based Search [3] (CBS) is an algorithm designed to solve the Multi-Agent Path Finding (MAPF) problem, which involves finding paths for a group of agents in a space-time environment while avoiding collisions and deadlocks. CBS is considered an optimal and complete algorithm, but it can be computationally expensive. Nevertheless, this factor is not a problem in AGV deadlock resolution since computational burden is not a critical constraint as vehicles are stationary during deadlock resolution. CBS employs a hybrid approach consisting of two levels: a high-level perspective that views the system holistically and a low-level perspective that deals with individual vehicles. The high-level component of CBS utilizes a best-first approach to explore a binary tree known as the Constraint Tree (CT). Each node N within the CT comprises three key elements:

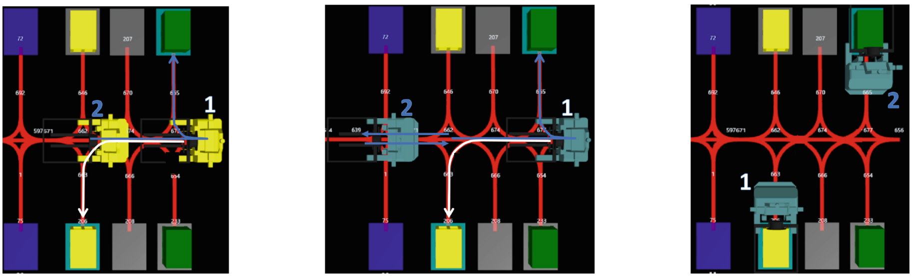
- **Constraints** (N.constraints): These constraints can either be vertex constraints, restricting an agent's presence at a specific vertex and time, or edge constraints, prohibiting agent movement between two vertices during specific time intervals.
- **Solution** (N.solution): This includes a set of cost-minimal paths for each agent, satisfying the constraints defined in N.constraints.
- **Cost Value** (N.cost): This represents the total cost, calculated as the sum of the costs associated with the paths in N.solution.

The process starts with the root CT node, which begins with an empty set of constraints. When CBS selects a CT node (denoted as N) for expansion, it checks for conflicts within N.solution. If there are no conflicts, CBS stops and returns N.solution as the optimal solution. However, if conflicts exist, CBS selects one of these conflicts, resolves it by splitting node N into two child nodes, and adds an additional constraint to each. The path of the agent involved in the conflict must be re-planned using a low-level search method. CBS explores both possible ways of resolving each conflict until a conflict-free solution is found or all possibilities are exhausted.

In our scenario, CBS is initiated when a deadlock is detected with the initial poses and goal points of the vehicles involved. Hence, the low-level planner employs the space-time A* algorithm to find paths for each AGV on the road map without any constraints. Next, high-level analysis in CBS involves comparing grid map cell sets of curve segments in space-time. If intersecting cell sets are found at a particular time step, a conflict is detected. Constraints are then generated for both vehicles, blocking the segment of one in space-time for the time the other encumbers it. In this procedure, two child nodes are created and analyzed, as explained above. This iterative process continues until a global conflict-free solution is found, provided it exists as CBS is complete. Finally, the result of CBS is handed to the traffic manager, that applies it to the vehicles, defining precedence relationships for initial deadlock resolution. Once resolved, the traffic manager returns to normal operations while maintaining the new road map routes.

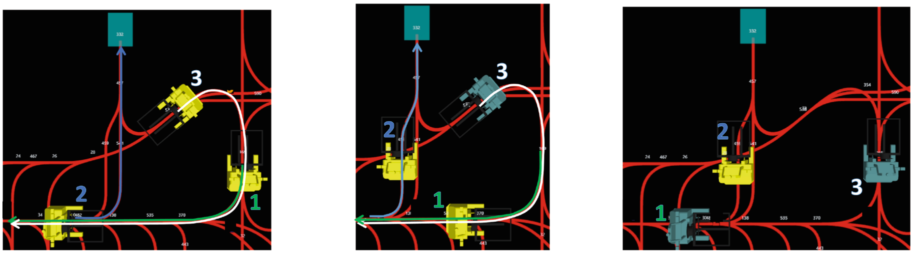
To show the potential of the suggested approach, two examples of frequent deadlocks in our case study's traffic manager are provided below. The first case occurs when a vehicle is given an update mission within a narrow bidirectional corridor with another AGV chasing it. If the new mission involves a path that obstructs that of the pursuing vehicle, a deadlock occurs, and re-planning of one or two vehicles within the corridor is required. In Fig. 1, vehicle 1's goal point is upgraded from 233 to 206. The new mission and vehicle 2's mission conflict, and thus a deadlock develops. The path of vehicle 2 is then re-planned so that it will backtrack and bring both missions to completion.

The second scenario occurs when traffic manager rules are applied and three or more vehicles establish a precedence loop. As shown by the arrows in Fig. 2, vehicle 2 must proceed to end point 332, while the other two must continue to the left part of the plant. In this situation, vehicle 3 cannot proceed because it



(a) (b) (c)

Fig. 1. Deadlock caused by updated mission in bidirectional narrow corridor



(a) (b) (c)

Fig. 2. Deadlock caused by cyclic precedence assigned by the traffic manager

is obstructed by vehicle 1, vehicle 2 cannot proceed because it is obstructed by vehicle 3, and vehicle 1 cannot proceed because it grants precedence to vehicle 2. To resolve the deadlock, the path of vehicle 2 is re-planned, as depicted in Fig. 2b. In this way, a solution in space-time is found, and all vehicles can continue on their missions.

4 Conclusion

In conclusion, this paper has introduced a promising deadlock detection and resolution strategy that holds the potential to enhance the operational efficiency of AGV systems. While the initial results are encouraging, further validation through extensive testing across multiple industrial settings is essential to assess its real-world applicability. Looking ahead, a future research direction involves extending deadlock detection to instants of time greater than zero so as to achieve the deadlock prevention function and further improve AGV fleets performance.

References

1. Chen, X., Xing, Z., Feng, L., Zhang, T., Wu, W., Hu, R.: An ETCEN-based motion coordination strategy avoiding active and passive deadlocks for multi-AGV system. IEEE Trans. Autom. Sci. Eng. **20**(2), 1364–1377 (2023). https://doi.org/10.1109/TASE.2022.3175909
2. Pratissoli, F., Brugioni, R., Battilani, N., Sabattini, L.: Hierarchical traffic management of multi-AGV systems with deadlock prevention applied to industrial environments. IEEE Trans. Autom. Sci. Eng. 1–15 (2023). https://doi.org/10.1109/TASE.2023.3276233
3. Sharon, G., Stern, R., Felner, A., Sturtevant, N.R.: Conflict-based search for optimal multi-agent pathfinding. Artif. Intell. **219**, 40–66 (2015). https://doi.org/10.1016/j.artint.2014.11.006, https://www.sciencedirect.com/science/article/pii/S0004370214001386

4. Shimomura, T., Ikeda, K.: Two Types of Deadlock Detection: Cyclic and Acyclic, pp. 233–259. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-04702-7_14
5. Zajac, J., Małopolski, W.: Structural on-line control policy for collision and deadlock resolution in multi-AGV systems. J. Manuf. Syst. **60**, 80–92 (2021). https://doi.org/10.1016/j.jmsy.2021.05.002, https://www.sciencedirect.com/science/article/pii/S0278612521000996

An Efficient Architecture Fulfilling Safety

Andrea Pupa(✉), Marco Minelli, Giorgio Battiato, and Cristian Secchi

Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Reggio Emilia, Italy
andrea.pupa@unimore.it

Abstract. In real-world industrial settings, human operators and robots often work closely together. In this scenario it is necessary to guarantee safety taking into account that the robot's path sometimes can not be changed, e.g. welding task. To improve robot performance while maintaining safety standards, a two-layer safe architecture is introduced. The first layer generates joint reference velocities to preserve the task path and utilize robot redundancies for obstacle avoidance. The second layer scales these velocities to comply with ISO/TS 15066 safety regulations. The framework has been experimentally validated with a Kuka LWR4+.

Keywords: HRC · Safety in HRI · Optimization and Optimal Control

1 Introduction

Collaborative safety standards delineate four distinct collaborative modes: *safety-rated monitored stop* (SMS), *hand guiding* (HG), *speed and separation monitoring* (SSM), and *power and force limiting* (PFL). For each of them, the technical specification ISO/TS 15066 [1] provides comprehensive guidelines for assessing the associated risks within each of these collaborative modes. The problem of how to ensure a safe robot behaviour has been widely analyzed in literature, addressing safety both at the planning level and at the control level, see, e.g., [2,3]. Some of these works, explicitly take into account safety standards in order to achieve a more efficient collaboration. In [4], the authors address the PFL constraint using an energy tank-based approach for safe interaction. Conversely, in [5], the focus is on enhancing flexibility and efficiency under the SSM velocity constraint with a two-layer architecture. The first layer ensures collision-free trajectories, while the second layer dynamically adjusts robot speed without altering joint space paths. When speed reduction is excessive, the second layer can trigger the first for replanning. Finally, in [6], the authors combine the SSM and PFL modalities into a single velocity constraint, shaping the robot trajectory online using a control barrier functions (CBFs) based architecture. However, some tasks require a perfect execution of the planned path, such as welding operations. In these cases, the proposed strategies may fail, leading to a deviation

Project funded under the National Recovery and Resilience Plan (NRRP), Mission 04 Component 2 Investment 1.5 - NextGenerationEU, Call for tender n. 3277 dated 30/12/2021. Award Number: 0001052 dated 23/06/2022.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 275–279, 2024.
https://doi.org/10.1007/978-3-031-76428-8_51

from the planned path or a decrease in performance. It would be more beneficial to exploit robot redundancies to increase the distance from the human operator while still completing the task. To achieve this, the hierarchical quadratic programming (HQP) formulation, as described in [7], may be exploited to ensure that the robot follows the desired path, allowing secondary tasks to be carried out in the null-space.

In this work we propose a two layer architecture for trajectory execution that explicitly considers the ISO/TS 15066 constraints to ensure a safe collaboration. The first layer takes as input a desired reference trajectory in the task space and solves online a HQP problem to generate the corresponding joint velocities. The goal of this layer is to maintain the path execution while trying to leverage the more degrees of freedom (DoFs) to increase the distance with the operator, i.e. allowing an higher speed. The second layer, instead, takes such joint velocities and modulate them online ensuring that the implemented behaviour is compliant with the safety standards. Summarizing, the main contributions of this paper are:

– A novel two layer architecture that explicitly considers redundancy to improve the performances of the robot without affecting the safety.
– A validation of the architecture in a real environment, proving the effectiveness of the framework.

2 Problem Statement

Consider an HRC application where a n-DoFs velocity-controlled robot manipulator has to collaborate with a human operator in order to accomplish a common job. The robot model is defined as:

$$\dot{\boldsymbol{q}} = \boldsymbol{u}, \tag{1}$$

where $\dot{\boldsymbol{q}} \in \mathbb{R}^n$, and $\boldsymbol{u} \in \mathbb{R}^n$ are the joint velocities and the controller input, respectively. During the collaboration, the robot has to perform a set of tasks associated with a desired trajectory $\boldsymbol{x}_{des}(\cdot) \in \mathbb{R}^m$, with $m < n$, that has to be executed from an initial time $t_i \in \mathbb{R}$ to the final time $t_f \in \mathbb{R}$. Such tasks are considered executed if and only if the robot follows exactly the same path, while the velocity can be modulated online if necessary. In such cases, this trajectory can be decomposed by applying a path-velocity decomposition:

$$\boldsymbol{x}_{des}(t) = \boldsymbol{x}_{des}(s(t)) \quad \wedge \quad \dot{\boldsymbol{x}}_{des}(t) = \boldsymbol{x}'_{des}(t)\dot{s}(t) \qquad t \in \left[t_i, t_f\right], \tag{2}$$

where $s \in \mathbb{R}$ is the curvilinear abscissa that parametrizes the geometrical path $\boldsymbol{x}_{des}(s(t))$, $\boldsymbol{x}'_{des}(t)$ is the vector tangent to the desired path, and $\dot{s}$ constitutes the magnitude of the robot velocity. While executing the trajectory, the robot behaviour must be compliant with the safety limit imposed by the ISO/TS 15066. In particular, focusing on the SSM, it is possible to derive an upper bound of the robot velocity towards the human operator v_{rh} [5]:

$$v_{rh_{max}} = \sqrt{v_h^2 + (a_{max}T_r)^2 - 2a_{max}(C + Z_d + Z_r - S_p)} - a_{max}T_r - v_h. \tag{3}$$

$v_h \in \mathbb{R}$ is the scalar velocity of the human operator towards the robot, $a_{max} \in \mathbb{R}$ is the maximum deceleration, and $T_r \in \mathbb{R}$ is the robot reaction time. C is the intrusion distance, i.e. the distance that a part of the body can intrude into the sensing field before it is detected, while Z_d and Z_r are the position uncertainties of the human operator inside the workspace and of the robot system, respectively. Lastly, S_p represents the protective separation distance. Thus, the shared workspace is equipped with a monitoring unit that allow to track the human operator, both position and velocity, during the execution of the task.

The aim of this work is to create an architecture that allows the robot to follow the desired path accurately while utilizing redundancies to increase the distance from the human. Ensuring compliance with (3), increasing this distance could potentially reduce task completion time.

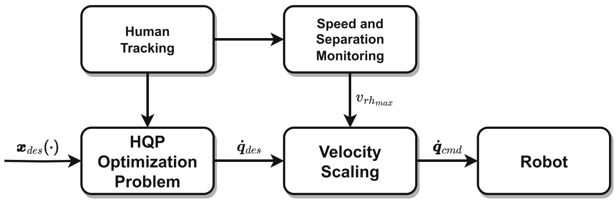


Fig. 1. Proposed architecture.

3 Architecture

The proposed architecture is illustrated in Fig. 1, where two main layers can be distinguished. The HQP Optimization Problem, which solves online an optimization problem in order to generate the reference joint velocities $\dot{\boldsymbol{q}}_{ref}$ such that the desired path is preserved and that, with the redundant DOFs, the distance with the human operator is increased. The Velocity Scaling, which modulates online the velocities $\dot{\boldsymbol{q}}_{ref}$ to ensure that the robot behaviour will be safe, i.e. the robot speed towards the human operator will be lower then (3).

The overall procedure starts with a desired reference trajectory $\boldsymbol{x}_{des}(\cdot)$, which can be both predefined or computed online exploiting dedicated strategies [8]. At this point, the framework solves the following optimization problem:

$$\begin{aligned} \min_{\dot{\boldsymbol{q}}_{ref},\boldsymbol{\delta}} \quad & \frac{1}{2}\|\dot{\boldsymbol{q}}_{ref}\|^2 + \frac{1}{2}\|\boldsymbol{\delta}\|^2 \\ s.t. \quad & \boldsymbol{J}_T(\boldsymbol{q})\dot{\boldsymbol{q}}_{ref} = \dot{\boldsymbol{x}}_{des}, \\ & \boldsymbol{J}_i(\boldsymbol{q}_i)\dot{\boldsymbol{q}}_{ref,i} = \nabla U_{rep,i}(\boldsymbol{x}_i) + \delta_i \qquad \forall i \in [1, \dots, n-1]. \end{aligned} \tag{4}$$

$\boldsymbol{\delta} \in \mathbb{R}^{n-1}$ represents necessary slack variables for maintaining optimization feasibility. $\boldsymbol{J}_T(\boldsymbol{q}) \in \mathbb{R}^{m\times n}$ is the task Jacobian, and $\boldsymbol{J}_i(\boldsymbol{q}i) \in \mathbb{R}^{3\times i}$ is the Jacobian of the linear part of i-th link. $\nabla Urep, i(\boldsymbol{x}_i) \in \mathbb{R}^3$ is the gradient of a repulsive potential field, generating a repulsive force between the i-th link and obstacles, e.g., the human operator. The first constraint ensures path consistency between desired and executed paths, while the second constraint optimizes redundant robot DoFs to increase the separation distance when the human operator approaches.

The result of the optimization problem (4) does not guarantee that the robot behavior complies with ISO/TS 15066. Thus, $\dot{\boldsymbol{q}}_{ref}$ can not be sent directly to the robot. To make sure the collaboration is safe, the following safe scaling strategy is employed [5]:

$$\begin{aligned} \max_{\alpha} \quad & \alpha, \\ s.t. \quad & \boldsymbol{n}_{rh}\boldsymbol{J}_i(\boldsymbol{q}_i)\alpha\dot{\boldsymbol{q}}_{ref} \leq v_{rh_{max,i}} \qquad \forall i \in \{1,\dots,n\}, \\ & \dot{\boldsymbol{q}}_{min} \leq \alpha\dot{\boldsymbol{q}}_{ref} \leq \dot{\boldsymbol{q}}_{max}, \\ & 0 \leq \alpha \leq 1. \end{aligned} \tag{5}$$

$\alpha \in \left[0, 1\right]$ represent a scaling factor which modulates the robot speed without changing the path. $\boldsymbol{n}_{rh} \in \mathbb{R}^3$ is the unit-vector representing the robot-human direction and $\boldsymbol{J}_i(\boldsymbol{q}_i) \in \mathbb{R}^{3\times i}$ is the linear jacobian of the $i{-}th$ link. $v_{rh_{max,i}} \in \mathbb{R}$ is the velocity limit imposed by the ISO/TS 15066. $\dot{\boldsymbol{q}}_{min} \in \mathbb{R}^n$ and $\dot{\boldsymbol{q}}_{max} \in \mathbb{R}^n$ are the joint velocity lower bounds and the joint velocity upper bounds, respectively.

Once the optimization problem (5) is solved, it is possible to compute the final robot input:

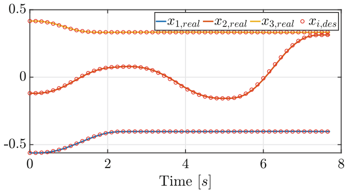
$$\dot{\boldsymbol{q}}_{cmd} = \alpha\dot{\boldsymbol{q}}_{ref}. \tag{6}$$

With the proposed approach, it is possible to ensure that the robot follows the desired path to accomplish the task, i.e. solving (4), and that the final robot behaviour is compliant with safety regulations, i.e. modulating the joints speed with (5).

4 Experiments

The proposed framework has been experimentally validated in a real scenario involving a Kuka LWR4+, a 7 DoFs collaborative robot, that has to follow a desired reference trajectory planned with SoA techniques. The task that must be executed is a linear path following, i.e. $m = 3$, which emulates all the industrial tasks that require only a linear translation, e.g. pushing of an object. In order to track the human operator seven OptiTrack Prime$^{\text{X}}$ cameras along with the Motive software have been exploited. Both optimization problems are solved online in cascade exploiting the Gurobi solver, achieving both solutions in less than 2 ms.

In the experiment, during the execution of the task the human operator hinders the robot at the elbow. Thanks to the second constraint in (4), a repulsive velocity is commanded and the robot is able to avoid to exploit the redundancies to avoid the human, efficiently completing the task. To validate the effectiveness of the framework, the same experiment has been performed with [5]. In this case, when the human operator hinders the robot the only admissible solution is to stop its motion until the human operator moves way. i.e. increasing the



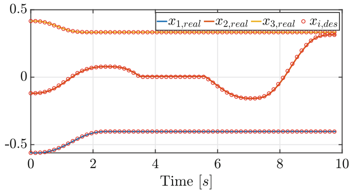


Fig. 2. Real and desired positions of the end effector with the proposed architecture (left) and with the approach proposed in [5] (right).

required time. Figure 2 and the video[1] show that, with the proposed approach, we achieved a time reduction of 21%.

5 Conclusions

This paper presents a novel two-layer framework for safe HRC. The first layer generates reference inputs, preserving the task path and enhancing human-robot distance exploiting redundancies. The second layer dynamically adjusts these inputs to meet ISO/TS 15066 standards. Future work will focus on refining the null space input generation, e.g. using CBFs instead of potential fields, or possibly replacing the first layer with a model predictive control approach.

References

1. Iso/ts 15066:2016(e). robots and robotic devices–collaborative robots, February 2016
2. Prianto, E., Kim, M., Park, J.H., Bae, J.H., Kim, J.S.: Path planning for multi-arm manipulators using deep reinforcement learning: soft actor-critic with hindsight experience replay. Sensors **20**(20), 5911 (2020)
3. Merckaert, K., Convens, B., Wu, C.J., Roncone, A., Nicotra, M.M., Vanderborght, B.: Real-time motion control of robotic manipulators for safe human–robot coexistence. Robot. Comput.-Integr. Manuf. **73**, 102223 (2022)
4. Benzi, F., Ferraguti, F., Secchi, C.: Energy tank-based control framework for satisfying the iso/ts 15066 constraint. arXiv preprint arXiv:2304.14059 (2023)
5. Pupa, A., Arrfou, M., Andreoni, G., Secchi, C.: A safety-aware kinodynamic architecture for human-robot collaboration. IEEE Robot. Autom. Lett. **6**(3), 4465–4471 (2021)
6. Ferraguti, F., Bertuletti, M., Landi, C.T., Bonfè, M., Fantuzzi, C., Secchi, C.: A control barrier function approach for maximizing performance while fulfilling to iso/ts 15066 regulations. IEEE Robot. Autom. Lett. **5**(4), 5921–5928 (2020)
7. Escande, A., Mansard, N., Wieber, P.B.: Hierarchical quadratic programming: fast online humanoid-robot motion generation. Int. J. Robot. Res. **33**(7), 1006–1028 (2014)
8. Pupa, A., Secchi, C.: A safety-aware architecture for task scheduling and execution for human-robot collaboration. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1895–1902. IEEE (2021)

[1] https://doi.org/10.5281/zenodo.10091336.

Grasp-O: A Generative System for Object-Centric 6-DoF Grasping of Unknown Objects

Kuldeep R. Barad[1,2(✉)], Andrej Orsula[1], Antoine Richard[1], Jan Dentler[2], Miguel Olivares-Mendez[1], and Carol Martinez[1]

[1] SpaceR, Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Esch-sur-Alzette, Luxembourg
kuldeep.barad@uni.lu

[2] Redwire Space Europe, 2530 Luxembourg City, Luxembourg

Abstract. Generative models are a promising avenue for learning generalizable robotic tasks from data. A fundamental task that remains a challenge to autonomous manipulation is the 6-DoF grasping of unknown objects. This work proposes Grasp-O: a simple, fast, and robust system for general-purpose vision-based 6-DoF grasping applications. Our system is built using a powerful and efficient Variational Autoencoder (VAE) that learns a distribution of $SE(3)$ grasp poses conditioned on object point clouds. The generative model is complemented by a grasp classification network that discriminates between good and bad grasp. We conduct extensive evaluations in simulation and the real world and demonstrate that our system outperforms existing VAE-based methods.

Keywords: vision-based grasping · variational autoencoder · point clouds

1 Introduction

Robotic grasping from visual observations is a key task in autonomous robotic manipulation across industry sectors. Grasping involves reasoning about a complex distribution of grasps based on geometric and physical properties of the object to find possible grasps from a set of infinite possible gripper configurations. While analytical solutions have been available for long, they rely on perfect knowledge of the object models. However, for general-purpose manipulation, object models are rarely available. Instead, a solution must rely on partial sensor observations of unknown objects. Recently, data-driven approaches [1] have produced remarkable progress. Most works focused on table-top grasping or bin picking of unknown objects using a 4-DoF grasp parameterization, using 3D translation and rotation in the plane, of a parallel-jaw gripper. On the other

Supplementary video is available at https://kuldeepbrd1.github.io/projects/grasp-o.
This work is supported by the FNR Industrial Fellowship grant (15799985).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 280–285, 2024.
https://doi.org/10.1007/978-3-031-76428-8_52

hand, 6-DoF grasping from visual observations is an active research topic [2] and applies to general manipulation settings beyond top-down grasping.

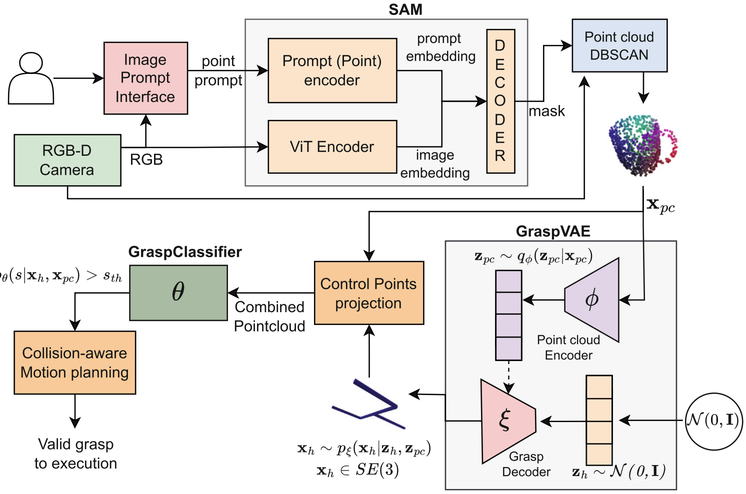


Fig. 1. Grasp-O architecture for 6-DoF grasping from an RGB-D image

In this work, we propose Grasp-O, an object-centric system for 6-DoF parallel-jaw grasping based on learned grasp synthesis and classification from a single-view RGB-D image. Figure 1 shows the overview of the architecture. The Grasp-O system's core is composed of a grasp generation stage based on a conditional Variational Autoencoder (GraspVAE) and a points-based grasp classification stage (GraspClassifier) that reasons about occlusions and collisions to aid in ranking several possible grasps. It is further facilitated by an interactive instance segmentation stage using the state-of-the-art segment-anything [3] foundation model. We train our models on 1000 training objects from 63 categories of ACRONYM [4] dataset and show that they readily transfer to the real world using noisy RGB-D observations. Further, Grasp-O with our models demonstrates a superior grasp success rate on 16 unseen objects across 80 grasp attempts than the baseline models [5].

2 Architecture

The system is composed of two core components: (1) a learned grasp sampler (GraspVAE) and (2) a points-based grasp classifier (GraspClassifier). We use an additional interactive segmentation stage to avail object-centric inputs to the grasp sampler. We describe each component and its design in detail below.

2.1 GraspVAE

GraspVAE is built as a conditional VAE [6] and is composed of the point cloud encoder (ϕ), the grasp pose encoder (ψ), and the grasp pose decoder (ξ). We use the point cloud encoder $q_\phi(\mathbf{z_{pc}}|\mathbf{x_{pc}}) : \mathbf{x}_{pc} \in \mathbb{R}^{3\times n} \mapsto \mathbb{R}^m$ that can operate on unordered point-sets of size $n \in \mathbb{N}^+$ to provide a fixed size latent ($\mathbf{z}_{pc}$) of size $m \in \mathbb{N}^+$ called the shape latent. The shape latent is used as conditioning in the grasp pose encoder $p_\psi(\mathbf{z}_h|H, \mathbf{z}_{pc})$ and the grasp pose decoder $p_\xi(H|\mathbf{z}_h, \mathbf{z_{pc}})$. $\mathbf{z}_h$ is the conditional grasp latent at the VAE bottleneck. Finally, the model is trained by jointly optimizing the parameters (ψ, θ, ξ) of the encoders and decoders to maximize the Evidence Lower Bound (ELBO) [6]:

$$\begin{aligned}\mathcal{L}_{ELBO}(\phi, \psi, \xi) = & \mathbb{E}\big[\log p_\xi(H^*|\mathbf{z}_h, \mathbf{z}_{pc}) \\ & - \lambda D_{KL} \left(q_\psi(\mathbf{z}_h|H, \mathbf{z}_{pc}) \,||\, \mathcal{N}(\mathbf{0}, \mathbf{I})\right) \big]\end{aligned} \tag{1}$$

In terms of implementation, λ is annealed linearly from 1e-7 to 0.1 until 50% of the total training steps and held constant thereafter. This is done to avoid KL vanishing and posterior collapse problems common to VAEs. Furthermore, we use efficient point-voxel convolution layers to build up the point cloud encoder, which encodes the input point cloud into a latent vector of length 128. This latent vector is used to condition the feature maps in the grasp encoder and decoder using feature-wise linear modulation [7]. We parameterize gripper pose $H \in SE(3)$ as a 6-vector comprising of 3 translation elements and 3 modified Rodriguez parameters. The network is trained on a labeled grasp set on 1000 diverse objects from 63 categories from ACRONYM [4] dataset generated in simulation.

2.2 Points-Based Grasp Classifier

The GraspVAE model may generate grasps that are in a collision, far away, or in occluded regions of the noisy point cloud from the sensor. Therefore, ranking the sampled grasps is critical for successful execution. To accomplish this, we introduce GraspClassifier, a 6-DoF grasp classification network that learns to assign a probability of success $p(s|\mathbf{x}_{pc}, \mathbf{x}_h)$ for an object-centric grasp pose. To incorporate grasp pose input into the network, we use a set of control points representing the Franka Emika Panda gripper and project them to a given pose. The gripper control points are appended to the object point cloud into a composite point cloud ($\mathbf{x}_c = \mathbf{x}_{pc} \cup \mathbf{x}_g$) is fed to the network as input. To distinguish between the points of the object and the control point of the gripper, we use a binary feature label assignment for each point. We use a 1D convolution layer and a linear layer after the PVCNN [8] base network to get prediction logits. The logits are then passed through the sigmoid function to get a value of class probability between 0 and 1. The model is trained to minimize the binary cross-entropy loss:

$$L(s, \hat{s}) = -\left(s \log(\hat{s}) + (1 - s) \log(1 - \hat{s})\right) \tag{2}$$

where s and $\hat{s}$ are the true and predicted binary success labels for the input point cloud-grasp pair.

2.3 Interactive Instance Segmentation

We use Segment-anything model (SAM) [3], a foundation model with zero-shot generalization capabilities to unfamiliar images and objects. SAM is designed to provide ambiguity-aware masks in a promptable setting, where a prompt specifies what to segment in the image. The prompts can be points, bounding boxes, regions, or natural language input. We use SAM with point-based prompting to extract the segmented point cloud of an object required for GraspVAE. A simple user interface allows the operator to select random point(s) on the object of interest in the image. The image is then fed to the image encoder, and selected point coordinates are provided to the prompt encoder of SAM. Once a predicted mask is obtained, we mask the aligned depth image from the depth stream of the RGB-D camera and de-project it to get the partial point cloud of the object. To remove the points bleeding over the edges of the objects and other outliers, we use DBSCAN clustering.

Table 1. Comparison of success rate and inference time for real-world evaluation on 80 grasps.

Model	Success Rate	Inference Time
GraspVAE + GraspClassifier (ours)	76.25 %	22 ± 5 ms 14 ± 2 ms
6DoF-Graspnet + Classifier [1]	37.5 %	56 ± 5 ms

3 Experiments and Results

The primary goal of our experiments is to evaluate the performance of our system in real-world 6-DoF grasping scenarios. We want to validate that our models trained purely from simulation data in a fully supervised manner can transfer and provide robust grasp synthesis in the presence of sensor noise and unmodeled artifacts. We use the *success rate* metric to report the performance.

We place a single object on a table in a random stable pose, and the goal of the task is to drop the object into a predefined bucket. An RGB-D camera is mounted in an eye-in-hand configuration on the robot's wrist. The RGB image is used by SAM to mask the depth map that is de-projected into a segmented point cloud of the object down-sampled to 1024 points. GraspVAE then samples 100 grasps conditioned on this point cloud, which are classified using GraspClassifier. The grasps are then sorted in decreasing order of success probability and checked for collision in the workspace. We execute the first feasible grasp without user intervention. In the post-grasp phase, the robot lifts the gripper to a safe clearance distance upwards, perpendicular to the table plane, without changing the gripper orientation. From there, the arm goes to a pre-defined gripper pose above the bucket and opens the fingers. The process ensures testing of grasp accuracy, robustness, and stability with constrained observations and motion planning. A grasp is successful if the object remains secure in the gripper until it is released into the bucket. To compute a meaningful success rate, We conduct five trials each for 16 objects in random relative poses to report the results in Table 1. We use 6DoF-Graspnet [5] as the baseline and only include their sampler and classifier for a fair comparison. We do not include their iterative refinement stage but keep the rest of the components of the Grasp-O system. Table 1 shows that the Grasp-O system with GraspVAE and GraspClassifier models significantly outperforms the baseline in our evaluation. We also report the inference time of the models on a system with Nvidia RTX3080Ti and Intel(R) Core(TM) i7-12800H, which shows that our models are comparable to the baseline, indicating promise for real-world applications (Fig. 2).

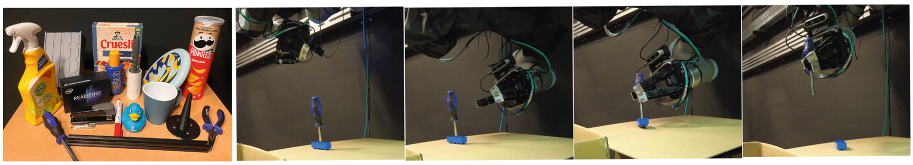


Fig. 2. Objects and grasp execution sequence for evaluation on a real robot

4 Conclusions

Learning generalizable representations for grasp generation and integrating them into practical systems is fundamental to robotic grasping. In this work, we proposed Grasp-O, a system for fast, accurate, and efficient object-centric 6-DoF grasping of unknown objects from a single RGB-D image. Our system trained on simulated data transfers directly to real robotic setup. Grasp-O provides 76.25% grasp success rate on challenging real-world evaluation with 16 unseen objects in various poses and improves significantly over the baseline.

References

1. Bohg, J., Kragic, D.: Learning grasping points with shape context. Robot. Auton. Syst. **58**(4), 362–377 (2010)
2. Newbury, R., et al.: Deep learning approaches to grasp synthesis: a review. arXiv preprint arXiv:2207.02556 (2022)
3. Kirillov, A., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
4. Eppner, C., Mousavian, A., Fox, D.: Acronym: a large-scale grasp dataset based on simulation. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 6222–6227. IEEE (2021)

5. Mousavian, A., Eppner, C., Fox, D.: 6-dof graspnet: Variational grasp generation for object manipulation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2901–2910. IEEE (2019)
6. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
7. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
8. Liu, Z., Tang, H., Lin, Y., Han, S.: Point-voxel cnn for efficient 3d deep learning. In: Advances in Neural Information Processing Systems, vol. 32 (2019)

Personalized Safety: Considering the Worker's Anthropometry in Safety Evaluation of Human-Robot Collaboration

Clara Fischer[1,2](✉), Friedrich Gregshammer[1], Martin Steiner[3], Michael Neuhold[3], and Sebastian Schlund[1]

[1] Institute of Management Science, TU Wien, Theresianumgasse 27, Vienna, Austria
clara.fischer@tuwien.ac.at

[2] Joanneum Research - Robotics, Lakeside B13b, 9020 Klagenfurt, Austria
clara.fischer@joanneum.at

[3] Tüv Austria GMBH, Deutschstraße 10, 1230 Vienna, Austria

Abstract. A risk-mitigating strategy for human-robot collaborations limits the power and force to comply with biomechanical limits in a collision, according to ISO/TS 15066:2016. Before industrial use, a safety evaluation is required to estimate potential impact forces and pressures, and the likely body regions affected. The verification depends on the chosen body region, as the standards specify physical parameters and biomechanical limits for specific regions. Depending on the worker's anthropometry, different regions can be hit with a particular robot configuration, e.g., chest, abdomen, or pelvis. The problem is that the person working with the robot and their body parameters are often unknown, and the affected body region can only be estimated. This paper analyzes the influence of the worker's anthropometry on safety evaluation of a human-robot collaboration. A study with 31 participants revealed considerable differences in affected body regions in possible impact scenarios. An exemplary biomechanical measurement illustrates these effects on the impact forces, confirming that consideration of the worker is essential.

Keywords: Industrial robot safety · human-robot collaboration (hrc) · power and force limitation · safety evaluation · worker anthropometry

1 Motivation and Problem Statement

Implementing a human-robot collaboration (HRC) for industrial use requires a risk assessment of the application. For HRCs using power and force limiting (PFL), possible impacts between the human and the robot must be identified and verified whether biomechanical limits are met. Biomechanical limits are maximum impact force F and pressure p, specified in ISO/TS 15066:2016 for 29 body areas for quasi-static, constrained and transient, unconstrained impacts [1]. Safety evaluation requires biomechanical measurements, using a power and

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 286–291, 2024.
https://doi.org/10.1007/978-3-031-76428-8_53

force measuring device (PFMD) to measure F, and pressure evaluation [2,3]. A PFMD models the affected body region (BR) as a mass spring damper model. An effective spring stiffness k simulates the mechanical properties of the BR, with an effective mass m_H, and a rubber damping element of specific Shore hardness (SH) simulates human tissue. Depending on the BR examined, a PFMD is built of particular parameters for k, SH, and m_H, which have a significant influence on F and p [3]. Therefore, identifying the affected BR is a crucial step in safety evaluation, and different biomechanical limits apply for each BR. However, the worker is often unknown and the affected BR can only be estimated, e.g. using standardized body percentiles of [4]. Currently, ISO 10218-2, the standard for robot systems, applications, and integration, is being revised, and updated content of ISO/TS 15066:2016 will be integrated, but with the initial biomechanical limits [5]. Some researchers criticize that the thresholds are too restricted regarding efficiency and work on new limits based on impact pain [6] or injury studies [7]. Furthermore, [5] does not consider the worker's anthropometry; therefore, we analyze this research gap and present the following contributions:

- HRC risk assessment knowledge provision, considering anthropometry.
- Creating input to the research field of PFL and biomechanical limits.

The paper is organized as follows. Sect. 2 investigates the influence of the worker's anthropometry in an impact user study and biomechanical measurements. The results are discussed in Sect. 3, and summarized with possible solutions in Sect. 4.

2 Investigating Worker's Anthropometry

2.1 Potential Impact Study

In a study with 31 participants (13 ♀ and 18 ♂) of different anthropometries (height 176 ± 19 cm), we evaluated the BRs the persons would be hit in potential impacts. We didn't perform physical impact tests, as presented in [8]; the aim of our study was only to verify the potentially hit BR. Therefore, first, the robot was moved in the impact position, and then the subjects were placed at a predefined location in front of the robot while it was stopped. In addition to examining the subjects, we evaluated the BR at which the 5th and 95th female and male percentiles (per.) were impacted using figures with standardized body measurements of [4]. A detailed description of the study can be found in [9].

Figure 1 shows the potential impacted BRs of an exemplary transient impact at which the robot collides linearly towards the human's upper body. In this scenario, the majority of the subjects would be hit on the chest, 77.4% (7 ♀, 17 ♂). This is also the BR where the 95th ♀ and ♂ per. would be hit. 9.7% (2 ♀, 1 ♂) and the 5th per. ♂ would be impacted on the neck, and 12.9% (4 ♀) and the 5th per. ♀ on the face. Similar to Fig. 1, we have created four further scenarios. In the second scenario, 61.3% (7 ♀, 12 ♂), and the 95th per. ♀ and ♂ would be impacted on the pelvis. For 35.5% (5 ♀, 6 ♂), and the 5th per. ♀ and ♂, the abdomen was the target BR, and for 3.2% (1 ♀), the chest BR would

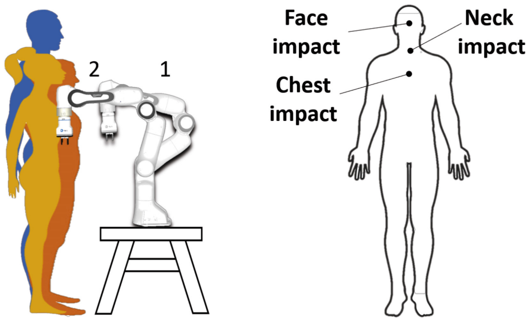


Fig. 1. Transient impact scenario and BRs hit (1 - Initial robot position, 2 - impact)

be impacted. In the third scenario, we identified two BRs: upper arms, elbows (93.6%, 13 ♀, 16 ♂, 5th & 95th per. ♀ & ♂) and lower arms, wrist BR (6.4%, 2 ♂). In the fourth scenario, the hand, finger BR (90.3%, 10 ♀, 18 ♂, 95th per. ♀ & ♂, 5th per. ♂) would be affected, or no impact will occur (9.7%, 3 ♀, 5th per. ♀). Identified BRs of the last scenario are the chest (64.5%, 9 ♀, 11 ♂, 5th & 95th per. ♀, 5th per. ♂) and the abdomen (35.5%, 4 ♀, 7 ♂, 95th per. ♂).

2.2 Consideration in an Exemplary Safety Evaluation

To verify the effect of various BRs hit based on different anthropometries, we have conducted biomechanical measurements of a transient impact for different BRs with the Franka Emika robot. In a setup similar to Fig. 1, the robot performed a linear movement towards the PFMD until a collision occurred. To simulate the human recoil after impact, the PFMD was mounted movable with an attached mass to achieve m_H. This setup was similar as described in [5], but the PFMD and the mass attached were suspended from a gallows with ropes to create a pendulum, as in [2]. The amount of additional mass resulted from m_H minus the PFMD's weight. We evaluated F with the PFMD by GTE CBSF 35/75. Based on the results of Sect. 2.1, we evaluated collisions for four different adjacent BRs likely to be impacted in a frontal transient collision, with the following parameters and transient (F_t) or quasi-static (F_s) force limit of [5]:

- **Face**: $m_H = 4.4\,\text{kg}$, $k = 75\,\text{N/mm}$, SH A70, $F_s = 65\,\text{N}$
- **Chest/Pelvis**: $m_H = 40\,\text{kg}$, $k = 25\,\text{N/mm}$, SH A70, $F_t = 280\,\text{N}/360\,\text{N}$
- **Abdomen**: $m_H = 40\,\text{kg}$, $k = 10\,\text{N/mm}$, SH A10, $F_t = 220\,\text{N}$

Chest and pelvis have the same parameters, only F_t differs. As a 10 N/mm or 25 N/mm k was unavailable, we used a higher one (35 N/mm). For the face, no F_t is provided, only F_s. Fig. 2 illustrates F over three robot velocities v (260/500/760 mm/s) for the investigated BRs. The presented F are the mean forces of three measurements, and σ represents their standard deviations, which are in the frame of the PFMD uncertainty (± 15 N). All F increase over v. F_s for the face is exceeded by all values. At $v = 500\,\text{mm/s}$, the abdomen F_t is not met anymore. At 760 mm/s, F_t for the chest is exceeded, while F_t for the pelvis is still met. We measured highest F for chest and pelvis, as these BRs share the highest k, SH, and m_H. The abdomen has the same m_H but a lower k and SH, which results in a lower F. The lowest F occurred for the face BR, caused by the lowest m_H. The correlation of F and the values of k, m_H, and SH is consistent with [3].

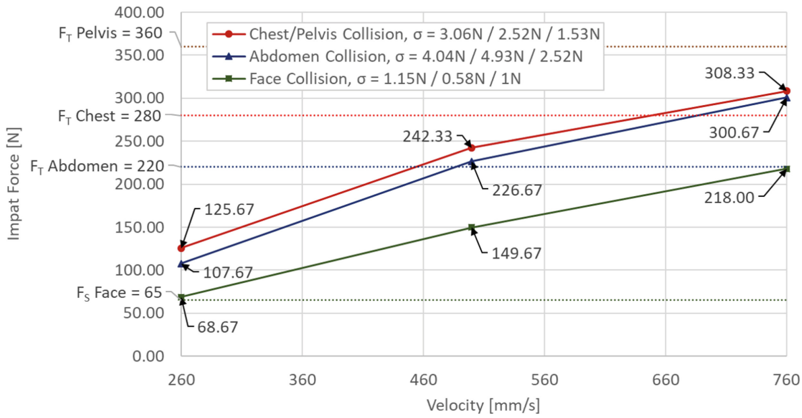


Fig. 2. Safety evaluation, F over v, for different BRs (F_T transient and F_S static limit)

3 Discussion

The results of Sect. 2 showed that worker's anthropometry influences the affected BR of a HRC collision. In our tests, the standardized body measurements of [4] present a valuable approach to identify affected BRs. Except for two people in the second and one in the fifth scenario, we identified all potentially affected BRs with the 95th and 5th per. female and male. However, a meaningful statement requires a study with more participants and several collision cases. As shown in Fig. 1, the evaluated BR is decisive for the measured F. However, the highest F is not always the most critical, as can be seen at 500 mm/s; the chest and pelvis maintain their biomechanical limits, while F_t of the abdomen is exceeded. In an industrial case, selecting only the chest and pelvis as affected BRs would lead to a false safety statement. The underestimation becomes even more critical when the face is hit, see Fig. 1. However, it should be mentioned that a scenario like in Fig. 1 won't occur in a HRC application and was only investigated for research purpose, as it is mentioned in [5] to avoid contact with the head.

Interviews with HRC safety experts, presented in [9], revealed that in most cases, the worker is unknown, and potential impacts and the affected BR must be estimated. The final workstation evaluation for the employee remains the operator's responsibility. Besides anthropometry, further factors, as higher pain sensitivity, e.g., due to pre-existing injuries or diseases, are not covered in current safety evaluation. Some experts mentioned it can be considered if pre-existing diseases or injuries exist. Others claimed that current biomechanical limits are already reduced sufficiently to consider increased pain sensitivity. In general, the experts said that current biomechanical limits are close to economic viability limits, as in [6,7]. For research on new thresholds, we recommend considering anatomic gender, as it is decisive for some BR's pain sensitivity, e.g., the chest.

4 Conclusion

This paper shows that the worker's anthropometry is decisive for the affected BR in human-robot collision. The investigated BR affects the safety statement of HRC in PFL. As the worker is often unknown in safety evaluation, testing adjacent potential hit BRs is required. To reduce the effort, the worst-case affected BR can be evaluated. In [3], a model to identify the worst-case BR in terms of the highest F and p is presented. A possible solution for future work is to determine the worker's anthropometry and adapt the robot path accordingly, as a safety function, that in a contact, the same BR is hit for each person.

Acknowledgment. This paper was supported by TÜV AUSTRIA #SafeSecLab Research Lab for Safety and Security in Industry. We thank Steffi Günther for proofreading.

References

1. ISO/TS 15066:2016, Robots and robotic devices: Collaborative robots (2017)
2. Fischer, C., Steiner, M., Neuhold, M., Papa, M., Markis, A., Schlund, S.: An investigation of the measurement of transient contacts in human-robot interaction. In: Müller, A., Brandstötter, M. (eds.) Advances in Service and Industrial Robotics, pp. 547–555. Springer International Publishing, Cham (2022)
3. Fischer, C., Neuhold, M., Steiner, M., Haspl, T., Rathmair, M., Schlund, S.: Collision tests in human-robot collaboration: experiments on the influence of additional impact parameters on safety. IEEE Access **11**, 118 395–118 413 (2023)
4. DIN SPEC 33402-6, Ergonomics - human body dimensions - part 6: Guidelines on the correct selection and application of anthropometric data (2015)
5. ISO/DIS 10218-2:2020, Robotics - safety requirements for robot systems in an industrial environment – part 2: Robot systems, robot applications and robot cells integration, March 2021
6. Han, D., Park, M., Choi, J., Shin, H., Kim, D., Rhim, S.: Assessment of pain onset and maximum bearable pain thresholds in physical contact situations. Sensors **22**(8), 2996 (2022)
7. Behrens, R., Pliske, G., Piatek, S., Walcher, F., Elkmann, N.: A statistical model to predict the occurrence of blunt impact injuries on the human hand-arm system. J. Biomech. **151**, 111517 (2023)

8. Haddadin, S., Albu-Schäffer, A., Hirzinger, G.: Requirements for safe robots: measurements, analysis and new insights. Int. J. Robot. Res. **28**(11–12), 1507–1527 (2009)
9. Gregshammer, F.: Der vergleich von standardisierten körpermodellen nach din iso 33402 und der individuellen arbeitskraft im zusammenhang mit der sicherheitsbewertung einer mensch-roboter kollaboration (2023)

A Robot Fleet Management System for the Energy Industry

Lorenzo Paladini, Enrico Meloni, Deepti Dighe, Marta Fiorucci, Luigi Bono Bonacchi, Manuel Pencelli, Guido Schillaci, Andrea Politano, and Giovanni De Magistris(✉)

Artificial Intelligence Team, Baker Hughes, Florence, Italy
{lorenzo.paladini,giovanni.demagistris}@bakerhughes.com

Abstract. This paper presents a robot fleet management system designed for the energy industry. The aim of the system is to simplify and optimize the monitoring and inspection process in energy plants using a fleet of robots and sensors. The proposed architecture follows a plugin-oriented approach, where each functionality is implemented as a plugin that can be individually enabled and configured based on specific monitoring requirements. The plugins can handle different tasks including event detection, mission coordination, data collection, and analytics. They interface between the automated monitoring hardware and an analytics dashboard providing real- time information and historical data for plant monitoring and problem detection. The platform is divided into two main layers: the Development Kit and the Analytics Dashboard. The Development Kit can be used by Robotics and Software Engineers to develop, test, and deploy new plugins. A web interface with interactive elements, including 3D environmental mapping and data visualization, enhances the user experience.

Keywords: fleet management · automatic inspections · robotics

1 Introduction

The rapid evolution of robotics and artificial intelligence have pushed the research on autonomous inspection, monitoring and management of production facilities. This shift towards inspection automation gives new possibilities of continuous data acquisition of the plant assets and instruments. However, programming fleets of autonomous robots to accomplish specific tasks still requires complex software and hardware infrastructures.

Most of the robot fleet management frameworks available in the scientific and industrial communities are based on open-source solutions, such as Robotics Operating System (ROS) or similar cloud-based platforms [1,2,6]. Managing groups of robots through such tools still demands highly specialized skills and extensive customization to meet industrial requirements. Notably, these platforms mostly address specific challenges in robot fleet management, such as

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 292–296, 2024.
https://doi.org/10.1007/978-3-031-76428-8_54

path planning and communication between multiple robots - e.g. see RoboFleet framework [5] - but miss the focus on the final application, for instance how to best manage the fleet for inspection and maintenance activities.

Despite the recent advances, very few studies focus on the integration of analytics capabilities within such platforms and on optimizing the joint activity of groups of robots for inspection tasks. Robotic inspection has the potential to increase the productivity and safety of industrial sites in the energy fields, especially when the infrastructure to inspect is in extreme and harsh environments [4]. Upscaling robot solutions to a fleet of robots on site is essential to achieve comprehensive, safer, cost-effective and efficient asset inspection [3]. This paper proposes a framework for managing a fleet of heterogeneous robots that carry out autonomous inspections in the energy industry. The solution takes advantage of ROS to create a template of plugins for gathering and analyzing sensor data. Furthermore, the architecture aims at an efficient handling and storage of data, enabling users to easily visualize the fleet state and the asset/instrument telemetry history through a web interface.

2 A Robot Fleet Management System for the Energy Industry

A plugin-oriented architecture is proposed, where each of the functionalities offered by the platform is installed in the form of plugins that can be individually enabled and configured according to the needs of the plant monitoring. Each plugin can handle different aspects of the monitoring processes, ranging from detection of events of interest (e.g., oil leakage, interruption of process) to orchestration and coordination of missions (e.g., metrics-based selection of the robot that is assigned a task), including taking measurements both from smart sensors and from direct inspections by robotic monitors. Plugins are fed data extracted by automated sensors and robots through ROS, elaborate the data with AI models, and store it on a database that can be accessed by the dashboard, or stream it in real-time. An operator can visualize the status of the plant, real-time information and historical data, allowing to extract insights on the evolution of the process during time. The dashboard allows to plan missions, by specifying a list of points of interest that should be examined, and for each of them one or more plugins that should elaborate the data before adding it to the report. A 3D map of the plant can be used to visually navigate the environment and place points of interest. The proposed architecture consists of 6 components (Fig. 1):

Tech Toolkit, a set of software libraries with a high-level abstraction over ROS, enabling an easy interface from the ROS layer to the analytics and plugins. This library also comes with an AI model zoo that can be re-used for different tasks. It also contains a set of base docker images that can be used in different phases of plugin development through pre-configured environments for developing plugins and for production and testing runtime. Its main advantage is speeding up development and increasing code maintainability through standard

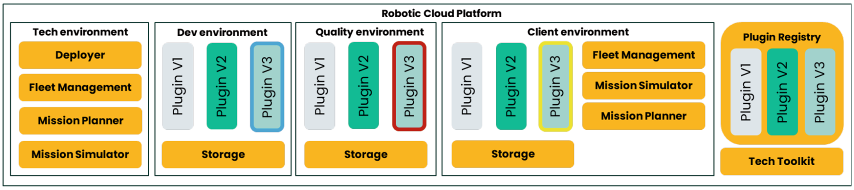


Fig. 1. The diagram shows the components of the Fleet Management System and how they are used in different environments: development, quality, and client.

practices, ensuring deployment reliability, by providing a standard configuration and guaranteeing the correct running conditions at runtime.

Plugin Deployer, a service for deploying new plugins to a cloud catalog, from which registered instances of the platform can download and use the plugin in their missions without the need of redeployment of the whole platform. In case of a plant detached from the internet for security purposes, the platform instance can be manually updated through local files provided in an external storage.

Fleet Manager, shows info about the fleet - such as device's health state, position, and capabilities - and allows live monitoring through onboard sensors.

Mission Planner, provides a graphical interface to plan a mission by declaring tasks and point of interests. The mission can be assigned manually to a particular set of robots and sensors or use cost-optimization algorithms or AI-based solutions automatically plan the mission based on robot capabilities, health and operativity status and future scheduled missions.

Mission Simulator, allows the creation of a simulated environment where to test designed missions before deploying the fleet to the actual plant.

Storage Manager, is the data storage for the plugins. It contains specific data related to the monitoring process of the plugin itself. The storage type can be relational database, non-relational database, and Amazon Web Services S3 (Fig. 2).

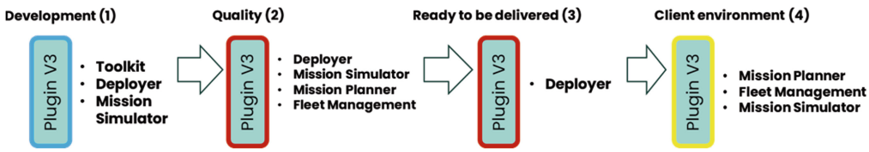


Fig. 2. The delivery process of a plugin. A plugin is first developed with the aid of the *Tech Toolkit*, the *Mission Simulator* and is then loaded by the *Deployer* into to the Quality Environment where the plugin can be tested more thoroughly on handcrafted simulations. When the plugin is ready for deployment, it is uploaded on the cloud by the *Deployer*. Then, the clients can download and install the plugin in their environment with their instance of the *Deployer*. They can use *Mission Planner*, *Fleet Management, and Mission Simulator* to plan and monitor missions on their plant.

3 Mockup and Web Interface

The Robot Fleet Manager web interface is meticulously crafted to serve as a versatile and user-friendly platform for efficient fleet management. It encompasses a rich set of features, including real-time fleet status monitoring, plugin management for individual robots, comprehensive mission inspection at both abstract and granular levels, and immersive 3D environmental mapping. This multifaceted tool is tailored to address a wide range of applications.

A dashboard overview acts as the central hub for overseeing the entire robot fleet with a quick and comprehensive summary of each robot's essential attributes.

A module for mission inspection offers an insightful summary of recent robotic missions executed by the fleet. These missions encompass diverse inspection tasks, from equipment status checks to anomaly detection (Fig. 3).

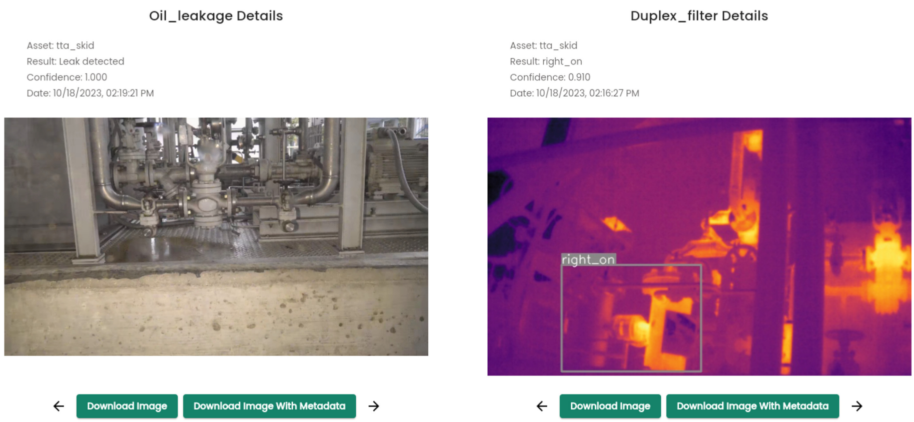


Fig. 3. Output of two plugins applied to Point of Interest. On the left, Oil Leakage visual inspection; on the right: Duplex Filter thermal inspection.

Data can be visualised through an integration with Grafana dashboard, which empowers users to gain insights through visual data representation of missions in the form of highly responsive and informative plots. Finally, a robot map dashboard leverages 3D point cloud data to provide a comprehensive visualization of the robot's operating environment. This functionality assists Mission Managers in mission planning and execution.

4 Tests on Site: Technical Training Academy

The platform has been tested on a quadruped robot at the Technical Training Academy of Baker Hughes in Florence. The tests aimed to evaluate the system's performance, functionality and compatibility with the energy industry's monitoring and inspection requirements. The testing area is a mineral lube oil console, a system designed for the capacity, filtration, and dissemination of mineral-based greasing oil in mechanical applications. The Anymal D quadruped robot from Anybotics was utilized during the tests and was programmed to perform visual inspections on the components of the gas turbine package. Several plugins were integrated into the autonomous mission including the motor pump thermal inspection, the oil leak detection, and the ventilation system inspection.

The test results at TTA played a crucial role in validating the effectiveness and practicality of the robot fleet management system, further enhancing its readiness for deployment in energy plants.

5 Conclusions

In this paper we presented the architecture of our Robot Fleet Management System. The architecture was developed with the requirements of being highly modular and customizable. The proposed solution focuses heavily on a plugin-oriented framework, allowing, contrary to most of the solutions already presented in literature, an easy integration with multiple heterogeneous inspection analytics. The system consists of 6 independent components, making the overall architecture simple yet highly flexible and powerful, and can handle multiple agents and sensors simultaneously. Finally, an easy-to-use dashboard has been developed to gather the most valuable information and present it to the operator in an intuitive way.

In future works we intend to further develop the platform. In particular, we will focus on smart cooperation between several agents with the objective of accomplishing multiple mission's targets concurrently and efficiently. In addition, more development will be done on the mission definition with the objective of eventually enabling automatic generation of the mission definition.

References

1. Alexandru, M., Dragoş, C., Bălă-Constantin, Z.: Digital twin for automated guided vehicles fleet management. Procedia Comput. Sci. **199** (2022)
2. Hazik, J., Dekan, M., Beno, P., Duchon, F.: Fleet management system for an industry environment. JRC **3**(6) (2022)
3. Jacobs, T.: What to expect when you're expecting robots. JPT **73**(08) (2021)
4. Shukla, A., Karki, H.: Application of robotics in onshore oil and gas industry-a review part i. Robot. Auton. Syst. **75**, 490–507 (2016)
5. Sikand, K.S., Zartman, L., Rabiee, S., Biswas, J.: Robofleet: open source communication and management for fleets of autonomous robots. In: IROS, pp. 406–412. IEEE (2021)
6. Singhal, A., Pallav, P., Kejriwal, N., Choudhury, S., Kumar, S., Sinha, R.: Managing a fleet of autonomous mobile robots (AMR) using cloud robotics platform. In: ECMR. IEEE (2017)

Process Orchestration and Product Traceability for Human-Robot Collaborative Remanufacturing

Angelos Christos Bavelos, Christos Gkournelos, George Michalos, and Sotiris Makris(✉)

Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Patras, Greece
makris@lms.mech.upatras.gr

Abstract. This paper explores the synergy between process orchestration and product traceability in the realm of human-robot collaborative remanufacturing, aiming to present a unified framework for heightened operational efficiency and comprehensive product traceability. Remanufacturing processes, characterized by complexity and variability, necessitate agile systems for effective workflow management. We delve into the integration of real-time, AI-driven process orchestration to dynamically allocate tasks, ensuring optimal resource utilization and workflow efficiency. Simultaneously, robust product traceability mechanisms are implemented to enhance transparency and accountability throughout the product lifecycle. Through case study, we showcase the successful integration of these components, emphasizing their collective impact on flexibility, quality control, and resilience in remanufacturing. This collaborative approach not only enhances operational agility but also establishes elevated standards for product traceability, addressing the growing demand for sustainable and transparent practices in remanufacturing industries. This research contributes a holistic perspective, providing a viable model for industry-wide adoption to advance remanufacturing processes. A remanufacturing scenario is given as paradigm for the proposed method.

Keywords: Remanufacturing · traceability · process orchestration

1 Introduction

Process orchestration has been gaining a lot of traction in manufacturing processes in the last years especially when a lot of actors are involved, such as both humans and robots in collaborative systems which greatly enhance flexibility and efficiency [1]. Remanufacturing is challenging, due to the complexity and uncertainties that the process of managing returned products introduces [2]. Decision-making tools like process orchestrators, when combined with traceability, mitigate uncertainty by tracking products throughout their lifecycle. Orchestrators adapt to environmental changes and resource positioning, as reported in [3]. In [4] the authors propose a process orchestrator in the automotive industry which dynamically allocates tasks to resources, with the

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 297–301, 2024.
https://doi.org/10.1007/978-3-031-76428-8_55

goal of creating a resilient architecture. Remanufacturing and resilience form a symbiotic bond. Remanufacturing boosts supply chain resilience through sustainability and flexibility, while resilient supply chains foster the growth of remanufacturing. Crucial for navigating modern business complexities. Product traceability enhances manufacturing with insights for quality control, complaint management, and addressing issues in damaged products and inefficiencies.. Additionally, it aids in the distribution of responsibilities within the manufacturing ecosystem [5]. The authors in [6] claim that there are five steps for initiating remanufacturing, (1) Select a product family, (2) involve actors prone to be impacted by remanufacturing, (3) iteratively identify prerequisites and assess the system performance, (4) develop a plan and industrialize remanufacturing, and (5) refine and validate the assessment in Step 3. Process orchestration and product traceability fit to step 4, since they industrialize remanufacturing by introducing efficiency, quality control, and adaptability.

2 Approach

The goal of this paper is to combine product traceability and process orchestration to produce the criteria that make remanufacturing flexible and resilient, as is shown in Fig. 1.

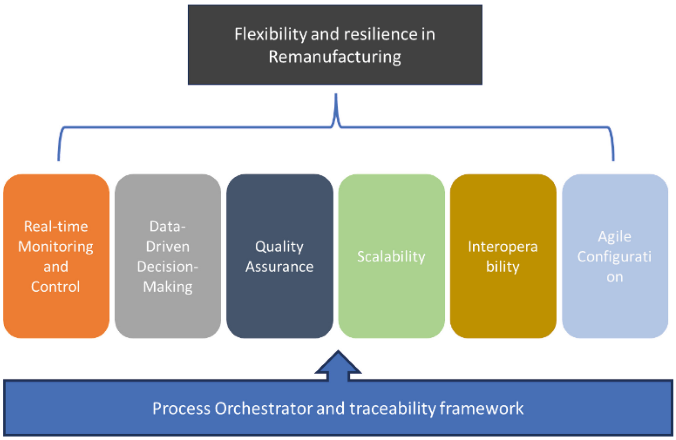


Fig. 1. Remanufacturing enhancement concept

2.1 Process Orchestrator Architecture

The AI-enhanced process orchestrator platform excels in allocating tasks and actions based on events, managing regular workflow or disruptions. Responsible for dynamic planning and scheduling, it evaluates the ongoing process, considering criteria like energy efficiency and material availability. Through a comparative analysis of hyper-heuristics and evolutionary AI, the platform optimizes task dispatching, adapting intelligently to diverse situations. Dedicated to responsive and efficient task allocation, this platform employs advanced techniques for enhanced adaptability and performance, ensuring optimal strategies tailored to unique requirements.

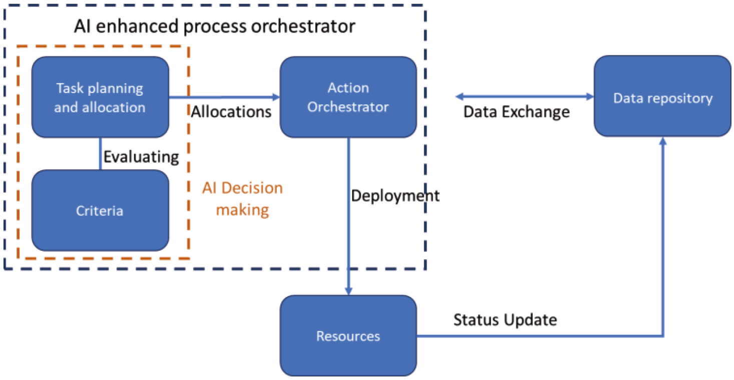


Fig. 2. AI enhanced orchestrator concept

2.2 Product Traceability

The integration of product traceability involves amalgamating data from diverse sources, including sensor data generated by remanufacturing equipment, historical information derived from prior remanufacturing processes, and feedback from the industrial actors. This concerted effort aims to enhance the accuracy and comprehensiveness of both product and process data. Achieving this objective necessitates the development of standardized data formats and protocols, establishing a cohesive framework for data representation and communication. Concurrently, robust data governance and security measures must be implemented to safeguard the privacy and confidentiality of sensitive information. These measures align closely with the industrial requirements, ensuring that product traceability not only improves data quality but also adheres to stringent standards of data protection and privacy.

2.3 Proposed Framework Advantages

The proposed framework is the combination of product traceability with the process orchestrator as shown in Figs. 1 and 2, which has the following advantages.

Traceability guides process orchestration, optimizing workflows with insights into configurations and historical data. This synergy allows agile adjustments to manufacturing lines based on traceability data, ensuring adaptability to evolving product specifications. Informed by traceability, process orchestration dynamically allocates resources, enhancing scalability for varying production volumes. Traceability's comprehensive view of the product lifecycle ensures seamless integration with diverse technologies, fostering interoperability. Real-time traceability data empowers process orchestration to monitor, control, and promptly resolve issues, optimizing processes for enhanced efficiency. Traceability's historical perspective guides process orchestration in implementing robust quality assurance measures, ensuring consistent high-quality remanufactured products.

The framework of the orchestrator communicates with other modules via Robot Operating System (ROS) communications [7], such as topics, services and actions, using existing ROS messages or if need be custom ones. In case of integration with non-ROS modules, ROSbridge [8] is used as a generic JSON API to ROS functionality.

Traceability relies on standardized data, protocols, and robust governance for privacy and security. RAMI 4.0, a framework within Industry 4.0, standardizes industrial processes, integrating communication, data exchange, and system architecture, covering the entire life cycle. It ensures seamless interconnection of assets, services, and processes, adhering to open standards. This comprehensive approach aligns with industry requirements, creating a unified foundation for traceability in industrial processes and systems.

3 Use Case

The proposed method is meant to be deployed in human-robot collaborative remanufacturing processes. In these use cases, efficiency, scalability, flexibility and resilience are at the forefront. Human-robot collaborative (HRC) processes have several challenges, such as safety concerns, sensor integration, human-robot interaction and correct allocation of tasks to the resources, real-time decision making, adaptation to interruption in the process, and many others. In Fig. 3, a paradigm of bicycle remanufacturing is shown, where the tasks to be completed are disassembly, inspection, cleaning, minor repairs, sanding and painting, and assembly.

The orchestrator dynamically allocates the tasks based on real-time data to the human and the robot, optimizing efficiency and sustainability. Traceability tracks the product lifecycle, integrating diverse data sources for a comprehensive understanding. Enhanced by traceability's historical data, the orchestration covers real-time decision-making and sensor data integration, ensuring interoperability. It adeptly handles interruptions, optimizes human-robot interactions with a focus on safety, ensuring resilience. For example, if inspection finds an issue with the wheel, that needs minor repairs, the orchestrator smartly allocates it to the human who has deft hands. In case of an issue with the painting, this task can be given to the robot. Traceability helps the inspection by providing a comprehensive record of every component's history, enhancing the orchestrator's allocation capabilities. The final assembly can be allocated to both human and robot, with the orchestrator handling the safety of the interaction.

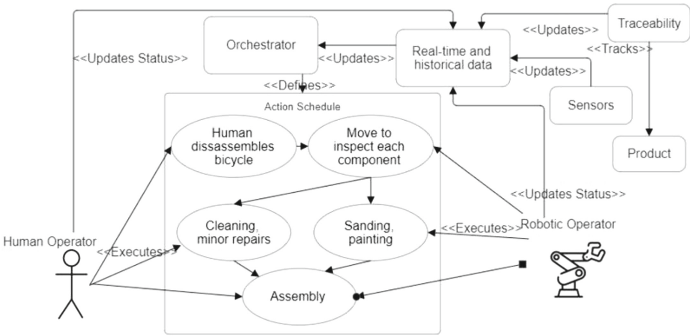


Fig. 3. Framework deployment in bicycle remanufacturing

4 Conclusion

The aim of this research is to examine the effect that process orchestration and traceability have in remanufacturing when combined. From the challenges and requirements of HRC remanufacturing use cases it can overcome, and the paradigm presented, it is concluded that the proposed method streamlines task execution and enhances data accuracy, creating a robust foundation for continuous improvement initiatives.

As future work, it is intended for the proposed method to be used in a real deployment of the described use case, to validate its usefulness in real industrial environments.

Acknowledgement. This research has been supported by the "RENÉE: Flexible remanufacturing using AI and advanced Robotics for circular value chains in EU industry" (Grant Agreement: 101138415) and "ASSISTANT: Learning and robust decision Support systems for agile manufacturing environments" (Grant Agreement: 101000165).

References

1. Makris, S.: Cooperating Robots for Flexible Manufacturing. Springer International Publishing, Cham (2021)
2. Rizova, M.I., Wong, T.C., Ijomah, W.: A systematic review of decision-making in remanufacturing. Comput. Ind. Eng. **147**, 106681 (2020)
3. Evangelou, G., Dimitropoulos, N., Michalos, G., Makris, S.: An approach for task and action planning in human–robot collaborative cells using AI. Procedia Cirp **97**, 476–481 (2021)
4. Schindewolf, M., Grimm, D., Lingor, C., Sax, E.: Toward a resilient automotive service-oriented architecture by using dynamic orchestration. In: 2022 IEEE 1st International Conference on Cognitive Mobility (CogMob), pp. 000147–000154. IEEE (2022)
5. Schuitemaker, R., Xu, X.: Product traceability in manufacturing: a technical review. Procedia CIRP **93**, 700–705 (2020)
6. Vogt Duberg, J., Kurilova-Palisaitiene, J., Sundin, E.: 5-step approach for initiating remanufacturing (5AFIR). Business Strategy and the Environment (2023)
7. URL ROS https://www.ros.org/
8. URL ROSbridge https://wiki.ros.org/rosbridge_suite

AI-Powered Human-Centred Robot Interactions: Challenges in Human-Robot Collaboration Across Diverse Industrial Scenarios

Francisco Fraile[1(✉)] and Sharath Chandra Akkaladevi[2]

[1] Universitat Politècnica de València, Valencia, Spain
ffraile@cigip.upv.es
[2] Profactor GmbH, Im Stadtgut D1, 4407 Steyr-Gleink, Austria

Abstract. This paper explores the AI-Powered Human-Centred Robot Interactions for Smart Manufacturing (AI-PRISM) project concept, through its various pilot scenarios. Focusing on human centered challenges, we discuss how these interactions can be optimized for better productivity and personnel well-being.

Keywords: Human-Robot Collaboration · AI-Powered Robotics · Smart Manufacturing · Programming by Demonstration · Industrial Automation

1 Introduction

Advanced robotics open up new possibilities for increasing efficiency and productivity in various sectors. The integration of robotics into human-centric work environments has become a pivotal aspect of technological advancement. However, this integration is not without its challenges. Key issues include ensuring safety, facilitating seamless communication, and maintaining a balance between automation and human expertise [3]. This paper delves into the AI-Powered Human-Centred Robot Interactions for Smart Manufacturing (AI-PRISM) project [1], focusing on diverse industrial scenarios. The exploration of pilot scenarios in industries like furniture, semiconductor, food and beverages, consumer electronics assembly, and discrete manufacturing further exemplifies the practical applications of AI-PRISM, showcasing its potential to address real-world challenges and contribute to the continuous improvement of Human robot Collaboration (HRC) efficiency and productivity optimization. It aims to provide insights into the multifaceted aspects of HRC, including user acceptance, ergonomic design, and the psychological impact on workers [2].

This work is funded under the EU Project AI-PRISM (EU grant agreement No 101058589).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 302–307, 2024.
https://doi.org/10.1007/978-3-031-76428-8_56

2 AI-PRISM Concept

AI-PRISM focuses on the adoption of HRC in industrial environments. One of the primary goal is to overcome the complexity associated with collaborative robots, especially for Small and Medium-sized Enterprises (SMEs), through concepts like Programming by Demonstration (PbD). Figure 1 shows the overarching concept of AI-PRISM.

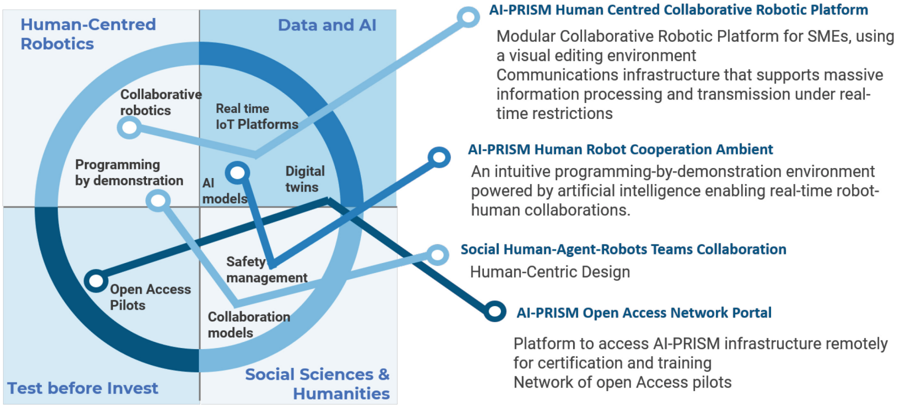


Fig. 1. AI-PRISM Concept

The key features and objectives of AI-PRISM can be summarized as follows:

- **Situational Awareness and Learning:** to gain comprehensive awareness of human activities in an industrial setting, where robots learn complex processes by observing and mimicking human operators [4].
- **Programming Simplification:** Overcoming the programming complexity barrier by providing an affordable solution for SMEs by developing intuitive programming interfaces that do not require advanced programming skills.
- **Human-Centric Collaborative Platform:** Development of a collaborative platform that integrates a ROS-based robotic system capable of digitalizing the collaboration environment, including humans.
- **AI Enhancement Toolset:** Integration of AI-based tools to enhance reasoning, perception, and coordination abilities of the robotic system.
- **Scientific Base in Social Sciences:** Building a solid scientific foundation rooted in social sciences to guide the design and implementation of AI-PRISM solutions.
- **User Pilots:** Conducting demonstrations in five user pilots across different industrial sectors, involving diverse and challenging-to-automate processes.
- **Open Access Network Portal:** Establishing a network of open access pilots engaging academic and industrial actors. Using AI-PRISM simulation services for demonstration, training, certification, and development.

In essence, AI-PRISM seeks to revolutionize collaborative robotics by combining advanced technologies, intuitive programming, and continuous learning, ultimately aiming to make industrial processes more efficient, flexible, and less stressful for human workers.

3 Use Case Scenarios

3.1 Furniture Sector

The furniture use case scenario in the AI-PRISM project focuses on the specific challenges of the furniture manufacturing industry. It emphasizes the integration of robotics in processes such as painting, and sanding, aiming to improve efficiency and reduce the physical strain on human workers. The key objectives include enhancing worker safety, optimizing production efficiency, and implementing ergonomic solutions to reduce work-related strain. In this use case scenario, programming by demonstration is used to capture expert knowledge of furniture craftsmanship, especially in challenging tasks like sanding. This approach allows skilled operators to share their expertise directly with the robots, ensuring quality and precision in work, empowering operators as they shift to higher added value tasks. The integration of robotics requires advanced scheduling to effectively balance the workloads of these skilled operators, who are also responsible for training the robots.

3.2 Semiconductor Sector

In this use case, the main challenge is to design an innovative AI robotic solution designed to assist workers in two highly demanding tasks: gluing and quality inspection of semiconductor components. These tasks are particularly stressing for human workers, causing both mental and physical stress due to the high precision and concentration required. The complexity of these tasks also presents a significant challenge for robotic systems, which must match or exceed human levels of dexterity and accuracy in a micro-scale. The AI-PRISM response to this challenge involves the integration of artificial vision industrial precision equipment. The core of this system is the control of a robotized precision positioning table, used for the accurate placement and inspection of micro-electronic components. The required precision is achieved by learning by example from an experienced worker's expertise. The worker's actions and decisions are captured through images taken from a microscope and data captured from the robotized precision table, which are seamlessly integrated into AI-PRISM. Based on this input data, The AI model learns to drive the robotized positioning table to apply the correct positioning learned for each specific component.

3.3 Food and Beverages Sector

The AI-PRISM use case in the Food and Beverages Sector focuses on three crucial scenarios in the brewing process: Filtration Powder Preparation, Sorting

Return Bottles from the Market, and Packaging and Palletizing Custom Orders. In Filtration Powder Preparation, manual handling of heavy sacks poses physical strain and potential exposure to hazardous substances for human workers. The Sorting Return Bottles scenario involves a complex automation system requiring human intervention for bottle categorization, while Packaging and Palletizing Custom Orders demand dynamic customization and manual lifting, leading to physical exertion. Shared challenges include repetitive strain injuries, health risks, and the need for dynamic task adaptation, aiming to enhance both physical and psychological comfort. To address these challenges, AI-PRISM introduces innovative solutions, contributing to the broader transformation of the brewing industry toward Industry 5.0. Specifically, in Filtration Powder Preparation, a collaborative robot handles sack transportation, reducing physical strain on human operators. In the Sorting Return Bottles scenario, a collaborative robot with a machine vision system assists human workers in bottle sorting, alleviating their workload. For Packaging and Palletizing Custom Orders, an Autonomous Mobile Robot (AMR) minimizes manual heavy lifting, allowing workers to place products on the robot for streamlined transportation. These implementations enhance safety, reduce physical exertion, and optimize workflow, addressing the complexities of dynamic industrial tasks.

3.4 Consumer Electronics Assembly

The Consumer Electronics Assembly use case in AI-PRISM deals with the assembly of range hoods. The assembly consists of 12 steps and requires visual checks, grounding tests, functional tests, cleaning, bagging, labeling, and more, and is currently performed manually. The manual assembly and quality control is quite challenging as it creates physical and mental strain on operators leading to increased ergonomic and health problems. This could further lead to reduced productivity, safety concerns, and decreased job satisfaction. To address these challenges, AI-PRISM envisions a collaborative setup between humans and robots in the assembly of range hoods. AI-PRISM proposes eliminating overlapping manual steps, introducing human-robot collaboration to enhance physical safety (e.g. grounding testing performed by cobots), and increasing cognitive stimulation for operators. The envisioned process incorporates robots for visual control and functional tests. The integration of AI-PRISM includes intelligent and adaptable control stations, with robots conducting manual tasks either independently or collaboratively with humans. The use of artificial intelligence-based models and object recognition algorithms enables robots to perform tasks such as object recognition, real-time tracking, and error detection on metal and glass surfaces. The system aims to increase resource efficiency, reduce production costs, enhance production volume, and improve worker safety. The AI-PRISM solution addresses the high-level objectives of decreasing mental fatigue, increasing cognitive engagement, and enhancing satisfaction and motivation for operators in the consumer electronics assembly process.

3.5 Discrete Manufacturing Sector

The Discrete Manufacturing Sector use case in AI-PRISM focuses on the production process of Printed Circuit Boards (PCBs). The challenges involve the manual handling and testing of PCBs in two scenarios. The first, PCB handling and testing, requires a human operator to manually load PCBs into a testing fixture after configuring the work cell with the correct test adapter. The second scenario involves configuring the work cell for new PCB types, requiring the installation of the required testing adapters and handling new PCBs accordingly. Challenges include repetitive manual tasks, quick adaptation to new PCB types, and potential mental fatigue due to the fast-paced nature of the process. AI-PRISM proposes the introduction of collaborative robotic assistance in the PCB testing process. In the first scenario, a static robotic manipulator is planned to handle and test PCBs, recognizing various types and placing them precisely into the testing adapter. This aims to reduce the physical and mental strain on human operators, enhance efficiency, and provide consistent and precise testing. In the second scenario, the work-cell configuration for new PCB types involves human-robot collaboration, where a mobile robot follows the human, carrying necessary parts and facilitating the assembly of testing adapters. This setup aims to streamline the process, improve adaptability to new PCB types, and reduce the manual burden on human operators. The integration of easy-to-use interfaces, such as voice-centric, graphical, and kinaesthetic teaching, enhances the collaboration between humans and robots, addressing the high-level objectives of decreasing mental fatigue, and increasing cognitive engagement.

4 Conclusion

In conclusion, this paper provides a comprehensive overview of challenges and potential strategies for effective human-robot collaboration within the AI-PRISM project. By examining pilot scenarios in industries like furniture, semiconductor, food and beverages, consumer electronics assembly, and discrete manufacturing, the paper contributes to the ongoing discourse on robotics integration in the workforce. Future steps should involve further refinement of AI-PRISM solutions based on user feedback from the diverse industry pilots. The successful application will set a precedent for the project's potential impact on various industries, suggesting that similar strategies could be implemented to optimize HRC across diverse industrial scenarios in the future.

References

1. AI-PRISM: Ai-powered human-centred robot interactions for smart manufacturing. https://aiprism.eu/ (2023), Accessed 12 Jan 2024
2. Faccio, M., et al.: Human factors in cobot era: a review of modern production systems features. J. Intell. Manuf. **34**(1), 85–106 (2023)

3. Hayes, B., Scassellati, B.: Challenges in shared-environment human-robot collaboration. Learning **8**(9) (2013)
4. Mukherjee, D., Gupta, K., Chang, L.H., Najjaran, H.: A survey of robot learning strategies for human-robot collaboration in industrial settings. Robot. Comput.-Integr. Manuf. **73**, 102231 (2022)

Towards Enabling Intuitive Interaction and Control of Mobile Robots Utilising Augmented Reality Techniques

Dimosthenis Dimosthenopoulos, George Mountzouridis, George Michalos, and Sotiris Makris(✉)

Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Patras, Greece
makris@lms.mech.upatras.gr

Abstract. Significant effort has been allocated recently by the research community for enabling seamless co-existence and communication between human operators and robots. In that context, the adoption of Augmented Reality (AR) has gained major attention towards supporting the operator's interaction with the robots. This paper aims towards enabling human-robot interaction (HRI) and robot control of mobile manipulators through the development of an AR application providing simple, customisable, and user-friendly interfaces. Robot data such as real-time status and planned pathways are projected to the operator's augmented environment enhancing their safety awareness. Additionally, interaction tools designed to enable robot control functionalities and holographic indications are tailored to provide real-time information to the operators, following User Experience (UX) design principles. The designed augmented environment is easily interacted with, utilising voice commands and virtual buttons. The implemented functionalities are realised through online communication and data exchange between the AR application and the ROS framework. The proposed application is validated through an industrial scenario from the elevators' manufacturing industry focusing on the elevator's electrical panel assembly.

Keywords: Augmented Reality · Human Robot Interaction · mobile robots · UX design

1 Introduction

In recent years, Human Robot Interaction (HRI) has gained popularity in the industry [1]. While automation is essential for high-volume production, market demands for product variability necessitate flexible and easily reconfigurable factories [2]. The HRI approach blends robot and human skills to enhance behaviour and efficiency of production systems assigning robots to handle repetitive and ergonomically stressful tasks, while human operators focus on tasks requiring dexterity and cognition [3].

The introduction of mobile robot workers able to autonomously navigate through the shopfloor executing various tasks and supporting human operators gain significant

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 308–312, 2024.
https://doi.org/10.1007/978-3-031-76428-8_57

attention among the research community [4, 5]. Nevertheless, intuitive interfaces on the human side are essential to facilitate smooth integration and communication between human operators and robot workers. At the same time, technologies such as the Augmented Reality (AR) have evolved and can provide assistance to mutable, human-related manufacturing tasks [6]. The see-through displays, that AR headsets use, does not cut out the real-world environment from the user allowing the operator to interact directly with their environment and the physical robot in an intuitive way [7].

This paper focuses on the deployment of an AR app towards enabling efficient HRI allowing the operators to be aware of the robot behaviour and directly influence it. Significant effort has also been directed towards the human-centric aspect adopting UX principles to strengthen the usability and acceptance of the app.

2 Approach

Moving forward of the discussed state-of-the-art, the main goal of this work is to contribute towards the realization of seamless human-robot collaborative production systems by utilising AR-based methods (Fig. 1), integrated into the Hololens2 suite.

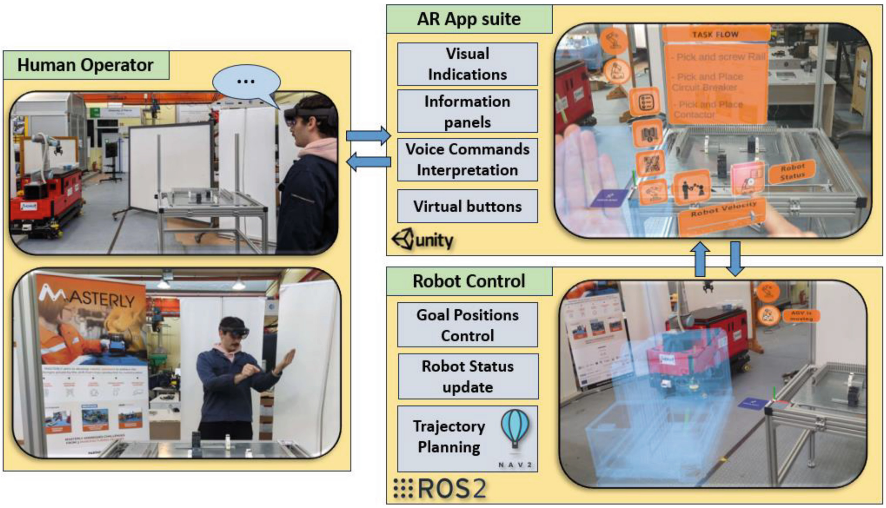


Fig. 1. AR based Human-Robot collaborative system

More specifically, the provided tools can simplify the interaction between the human operator and the robots as well as providing insight to them regarding the robot actions increasing their safety awareness and acceptance towards the robotic actors.

2.1 Voice Commands

The incorporation of voice commands handling into the AR interface enhances the simplicity of controlling and interacting with the robot and its augmented environment.

Users can seamlessly issue commands to the robot without disrupting their ongoing activities. These voice commands are translated by the MRTK library's voice input speech interpreter module integrated in the AR app into dedicated robot actions before passed to the robot controllers for execution. To minimize the risk of misunderstanding and unintentional communication, pre-determined words are introduced for the user to trigger and terminate the voice commands interpretation module.

2.2 Robot Path Visualisation and Status Update

Hologram-based robot trajectory visualisation techniques are also used to inform the operator about the path planned to be followed by the robot during the execution of a provided navigation goal enhancing the safety awareness of the operator. By integrating the AR suit with the Robot Operating System (ROS) framework and the robot controllers, the planned collision-free trajectory calculated using real-time data from the robot safety sensors is published as specific trajectory data to the AR app in the form of transform positions of the midway points. This data is utilised to generate 3D meshes indicating the planned robot's route until it reaches its destination. Moreover, the real-time status of the robot ("Idle", "Moving", "Malfunction" etc.) is communicated to the operator, through dedicated indications to their field of view.

2.3 User Interface and Visual Elements

The User Interface (UI) stands as the bridge between the operators and the capabilities of the system allowing them to interact with the provided functionalities in an intuitive and human-centric manner. The design of the application's UI has been tailored to be adaptable to the operator's needs through customisable panels and notifications.

An on-hand menu has been designed encapsulating interactive buttons and panels through which the operator can (de)activate notifications, initiate functionalities related to the robot control and receive visual indications or alerts related to the robot status and planned paths, as described above. A gaze-based approach is used to activate the menu, enabling its appearance when the user looks at the palm of their left hand, eliminating unnecessary obstructions for the operators. Finally, UX methods were considered during the design of the menu to ensure optimised placement of the various visual elements to the operator field of view without compromising their comfort. Thorough testing was conducted by multiple users assumed the operator role and the obtained feedback was utilised in further enhancing the ease of use of the application.

3 Implementation

The app is developed in Unity, a cross-platform game engine widely used for AR applications. Unity offers built-in support for various platform-specific features and a flexible VR/AR API allowing the creation of AR apps across diverse software systems. This versatility facilitates communication between different devices involved in HRI and the delivery of digitized information for operator support. On top of that, various libraries are utilised for the deployment of the implemented functionalities like MRTK, providing assets like interactive buttons, sliders and panels, User Input profiles for Voice Commands and Hand Tracking functions to build upon.

To enable the visualisation of the robot information through the AR UI, a connection between the AR interface and the robot has been established through the ROS2 framework (Fig. 2). To achieve this, a bridging protocol has been employed, utilising the "rosbridge_server" package and a dedicated ROS# library as client from the AR app side. This protocol offers a JSON API to ROS functionalities for non-ROS-based programs. Through this server-client channel, stable communication has been established between the app and various deployed ROS2 nodes. Real-time positional data and planned movement goals are transmitted as ROS messages to the Unity scene subscriber. These messages are then translated into transform position data and assigned to 3D holographic objects representing the mobile robot's current position and trajectory. In the same fashion, robot targets published from Unity include transform position data utilised by the motion planners to plan the needed trajectory. To minimize time delays, the User Datagram Protocol (UDP) is used to publish ROS messages in packets and thus increasing reception. In addition, the Robot status received through the ROS interface, is sent to the ROSBridge Server in the form of ROS Topics and is then published as ROS-Unity messages to trigger indicative visual elements.

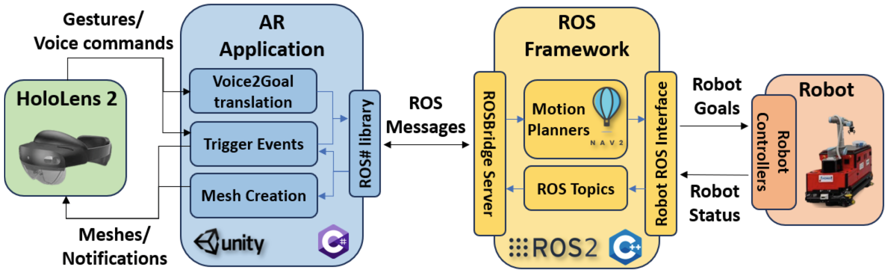


Fig. 2. Overall system architecture

4 Industrial Scenario

The framework described above is integrated in a scenario inspired from the elevator's manufacturing industry focusing on the assembly of an electrical panel. In current practice, the assembly is a manual process having an operator in each workstation and additional operators moving parts from the storage area to the assembly stations.

In the proposed scenario, a mobile platform equipped with a collaborative robotic arm takes over the logistic operations, while also being able to support the operator during the assembly of the panel (Fig. 3). Through the AR interface, the operator has overview of the robot behaviour through virtual representations of the planned trajectories and indications related to its status and possible malfunctions while also being able to instruct the robot to execute additional tasks through voice commands (bring consumables from the warehouse, perform a specific assembly task, etc.).

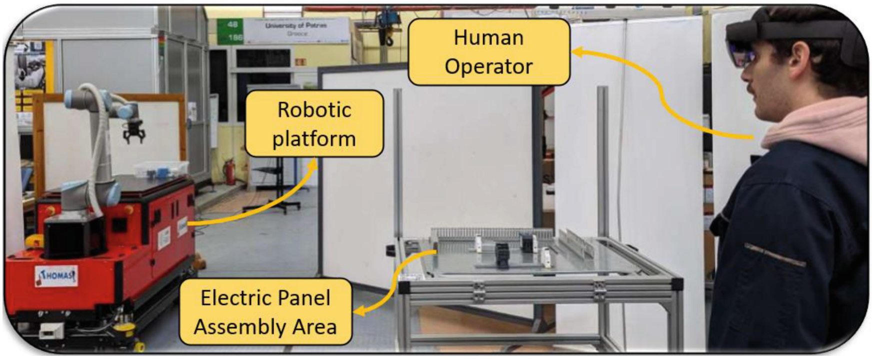


Fig. 3. Proposed industrial scenario setup

5 Discussion and Future Work

The aim of this work is to develop an AR-based framework helping human operators working in hybrid, human-robot collaborative industrial environments. Intuitive and human-centric interfaces are provided for interacting with the robots while keeping the operators aware about the robot actions and thus enhancing their "safety feeling".

Further research should focus on the expansion of the provided functionalities to incorporate more means for the operator to intervene and alter the robot behaviour. In addition, AI methods will be examined to allow the operator to communicate with the robot in natural language making their collaboration much more seamless and robust.

Acknowledgement. This research has been partially funded by the MASTERLY project, which received funding from the European Union's research and innovation programme Horizon Europe under the grant agreement 101091800.

References

1. Makris, S.: Cooperating Robots for Flexible Manufacturing. Springer International Publishing, Cham (2021)
2. Chryssolouris, G.: Manufacturing Systems: Theory and Practice, 2nd edn (2006)
3. Michalos, G., Makris, S., Spiliotopoulos, J., Misios, I., Tsarouchi, P., Chryssolouris, G.: Robo-partner: seamless human-robot cooperation for intelligent, flexible and safe operations in the assembly factories of the future. Procedia CIRP **23**, 71–76 (2014)
4. Green, S.A., Chen, X.Q., Billinghurst, M., Chase, J.G.: Collaborating with a mobile robot: an augmented reality multimodal interface. IFAC Proc. Vol. **41**, 15595–15600 (2008)
5. Kousi, N., Gkournelos, C., Aivaliotis, S., Giannoulis, C., Michalos, G., Makris, S.: Digital twin for adaptation of robots' behavior in flexible robotic assembly lines. Procedia Manuf. **28**, 121–126 (2019)
6. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. CIRP Ann. **61**, 657–679 (2012)
7. Lotsaris, K., Gkournelos, C., Fousekis, N., Kousi, N., Makris, S.: AR based robot programming using teaching by demonstration techniques. Procedia CIRP **97**, 459–463 (2021)

An AI-Based Decision-Making Framework with Task Planning and Dynamic Reconfiguration Capabilities

Apostolis Papavasileiou[1], Sotiris Aivaliotis[1], Christos Glykos[1], Spyros Koukas[2], and Sotiris Makris[1](✉)

[1] Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Patras, Greece
makris@lms.mech.upatras.gr

[2] Netcompany-Intrasoft, 2B Rue Nicolas Bové, 1253 Luxembourg City, Luxembourg

Abstract. This paper proposes a decision-making framework based on Artificial Intelligence (AI) functionalities for the dynamic planning and control of human-robot collaborative (HRC) scenarios both on the design and execution phase. Allocation of tasks among resources is being investigated based on user-defined metrics prioritized according to importance. Integration with a digital simulation tool is presented in order to take into account alternative scenarios and define the optimal solution. The proposed framework is applied in tandem with a central orchestrator in order keep track of the real execution status and dynamically reconfigure the provided plans upon request. The solution is validated in a case study derived from white goods industry where the operator is working under a collaborative environment with a robot in order to perform a defined assembly.

Keywords: AI Task Planning · Reconfiguration · Human-Robot Collaboration

1 Introduction

During the last years, focus has been given on the research of digital technologies towards the dynamic reconfiguration of production processes under HRC workstations [1]. This is an outcome of the continuously rising need for customized products combined with the existence of unexpected events that often may lead to loss of time and costs in production [2]. Given these challenges, the need for effective planning and dynamic adaptation of production schedules is evident in order to remain competitive.

Researchers presented different solutions regarding the task planning under human-robot collaborative workspaces. Such a solution is presented in [3], where a sequence planner allocates the different tasks between the robot and the human towards minimizing the disassembly time for a hard disk component. Another solution has been investigated in [4], where a task sequencing system was employed in order to dynamically adapt the task plans based on the human behavior. In addition, an interesting tool has been presented in [5] for assisting designers in deploying workstations for human-robot coexistence and allocate tasks based on their active skills, reducing the complexity.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 313–318, 2024.
https://doi.org/10.1007/978-3-031-76428-8_58

Investigating the existing literature, it was interesting to identify that most of the planning solutions are focusing on distributing the tasks in the identified resources in the design phase. However, during the phase of execution, multiple unexpected changes may occur in the real environment and need to be handled accordingly. Under this scope, this paper proposes an AI-based decision-making framework interconnecting modules for Task Planning, Digital Simulation and Central Orchestration of the production, providing thus robust planning and dynamic reconfiguration capabilities.

2 Approach

The proposed solution can automatically analyse the production tasks and rearrange the available production resources in real-time. The assembly operation is broken down to several tasks. These assembly tasks are inserted in the proposed framework for HRC scenario's modelling and added in the world model including the ontology-based resource and workload hierarchical modelling. The resource hierarchical model consists of information about resource's maximum lifting payload, walking and handling ability. The workload hierarchical model consists of data regarding which resource can execute each task and how these tasks should be executed per each assigned resource (Fig. 1).

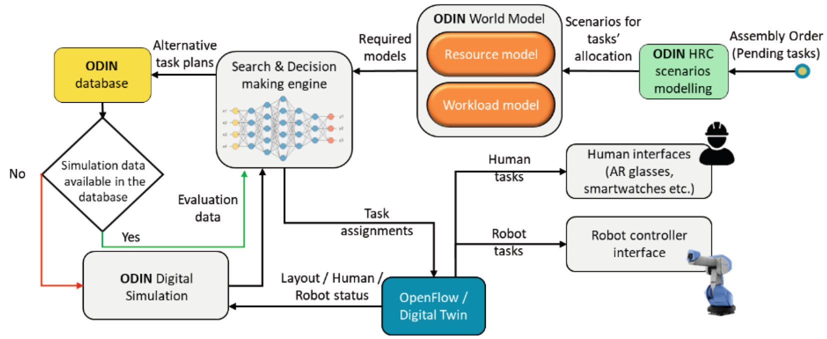


Fig. 1. AI based Decision Making framework methodology

The search engine of the AI based decision making framework generates alternative task plans based on tasks' and resources' models. The core of the scheduler consists of an Artificial Intelligence algorithm, which is based on custom AI-based heuristic and iterative deepening search functions. This algorithm intelligently explores a part of the solution space and provide a result of high utility value, without exhaustively examining all possible combinations of task-resource assignments and sequences based on:

- *DH – Decision Horizon*: The number being examined in each searching step.
- *SR – Sampling Rate*: This parameter depicts the number of samples that will be defined for each of the decision points of the decision horizon.
- *MNA – Maximum Number of Alternatives*: This parameter defines the maximum number of alternatives that will be formed at each step of the algorithm.

The generated plans are stored in the framework's database for evaluation through a Digital Simulation tool. Additionally, information for simulated execution of each task plan is stored to be used by the decision making algorithm of the planner. Digital Simulation layout initialization is based on data exported from the Digital Twin and the OpenFlow [6] modules. OpenFlow is an integration and orchestration framework for flexible, responsive and robust HRC manufacturing systems and handles the data exchange between the integrated modules. The task planner check if tasks' evaluation data are already available in the database and after the simulated execution, the proposed framework retrieves data about each task plans' execution from the database for evaluation. Based on the evaluation criteria's weight, the best alternative task plan is visualized on UI and sent for physical execution through OpenFlow.

The generated task plans evaluation is based on a set of evaluation criteria as follows:

1. *Flowtime* – Total time duration of a generated task plan execution.
2. *Human busy-time* – Total time duration that a human execute tasks.
3. *Distance Covered* – Distance covered by each resource.
4. *Ergonomics* – How much stressful are the human tasks for the operator.
5. *Non-Adding values activities time* – Time duration that no tasks are executed.
6. *Utilization of Resources* – Resources Utilization during task plan execution.
7. *Safety HR distance* – Avoid plans that do not conform to the safety concept.

The differences of the proposed framework when running in online mode is that the task plans generated by the intelligent scheduler include the tasks which have not been executed until the time that the re-scheduling request sent to the planner. OpenFlow monitors tasks' execution and excludes the executed tasks from the planning procedure.

To calculate the evaluation data in a more accurate way, the initialization of the simulation layout is based on data comes from the Digital Twin and the OpenFlow. Human model's location in the simulation as well as robot arm's pose are provided as input for the initialization of the simulation scene to replicate more accurately the physical layout of the work cell as it was at the moment the re-planning request sent by the OpenFlow.

3 Implementation

Production managers are able to interact through the developed User Interface (UI). The UI provides information regarding: a) Resource and Workload hierarchical models, b) definition of the research parameters values for task plans' generation, c) definition of the evaluation criteria and d) generated task plans' evaluation results (Fig. 2).

Digital simulation is used for each alternative's validation and the evaluation through human and robot motions execution. During task plans' simulated execution, data regarding the execution of each task, their duration and the distance covered are stored in database. Flowtime and distance covered criteria calculation is based on these data.

The proposed framework has been designed based on Robotic Operating System (ROS) utilizing all its communication and robot controlling libraries. Robotic resources motions are based on the MoveIt ROS library. The developed simulated tool utilized the GAZEBO simulation engine of ROS and a set of robot ROS controllers for robotic tasks

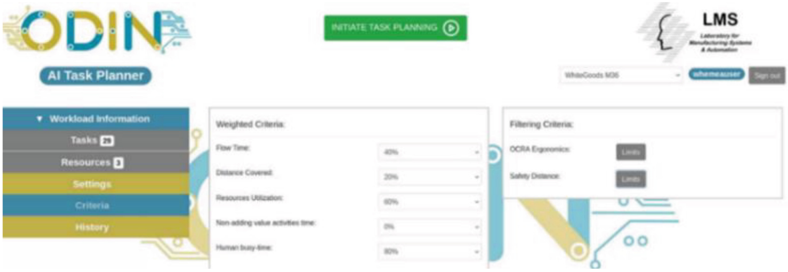


Fig. 2. Evaluation criteria definition through the developed UI

execution. Several ROS Action servers have been defined and are integrated with the proposed framework for the simulated execution of all human and robotic tasks (Fig. 3).

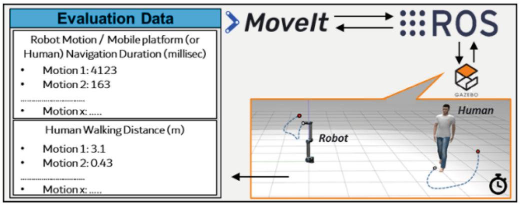


Fig. 3. Simulation engine for evaluation data calculation

4 Case Study

The proposed framework has been validated in a use case from the White Goods industry focused on different parts manipulation in terms of geometrical characteristics and weight. A human operator collaborates with a UR10 robot for the installation of three types of cooktops and knobs in gas cooktops and for transformers' installation in ovens. Pick, place and robot tool change tasks are modelled. The robot picks the required parts from two kitting tables placed on its left and the right side. While making different assignments to the available resources, several task plans are generated by the planner and sent to the simulation tool for validation and evaluation data calculation. The framework finds the best alternative and sends it for execution in the physical environment (Fig. 4).

The testing of the proposed framework at the initial phase of the production was successful. The assembly tasks were executed by the assigned resources successfully. During the execution of the assembly process, the operator needs to enter either the right or the left kitting area of the robotic cell for empty kitting carts' replacement. This task results in the violation of the safety concept and thus the replanning capabilities of the framework have been utilized. Instead of stopping robot's motion, new task plans are generated allocating tasks to the robot in non-interrupted areas. When the robot

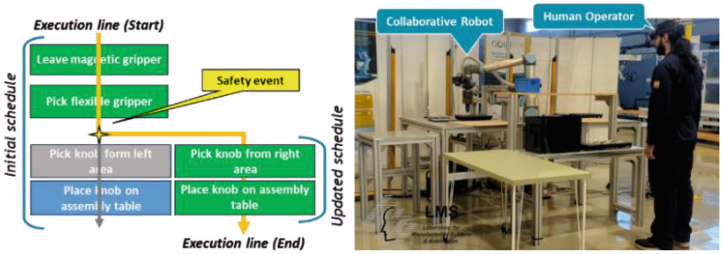


Fig. 4. Validation of AI-based decision making framework in the selected use case

is functional again, the developed framework generates a new task plan and equally distributes the tasks to the available resources.

At the initial state of the production, carts' replacement is a time-consuming process requiring production stoppage by the operator performing the replacement process. After cart's replacement, the operator recovers the production using a reset button installed on the back area of the cell. This process results in 21 s idle time for the resources executing the assembly tasks. Thanks to the developed framework, the robot is able to continue its tasks despite the human's presence reducing the idle time of the resources to 12 s which is required for the task replanning process execution.

5 Conclusions and Future Work

The proposed framework is used for production plans generation and evaluation either at the initial phase of the production design or on-the-fly during execution. It has been validated in a use case from the White Goods industry. After the proposed framework's integration, the resources are able to continue the assembly process regardless the presence of human operators inside the work cell.

Future work includes the connection of the framework with other simulation tools for validation and evaluation purposes with higher accuracy. The validation in other industrial pilots will also be investigated.

Acknowledgement. This work has been partially funded by the EC research project "ODIN – Open-Digital-Industrial and Networking pilot lines using modular components for scalable production" (Grant Agreement: 101017141) (www.odin-h2020.eu).

References

1. Kousi, N., Dimosthenopoulos, D., Aivaliotis, S., Michalos, G., Makris, S.: Task Allocation: Contemporary Methods for Assigning Human-Robot Roles, The 21st Century Industrial Robot: When Tools Become Collaborators, pp. 215–233. Springer (2022)
2. Chryssolouris, G.: Manufacturing Systems: Theory and Practice, 2nd edn. Springer, New York (2006)
3. Lee, M.-L., Behdad, S., Liang, X., Zheng, M.: Task allocation and planning for product disassembly with human–robot collaboration. Robot. Comput.-Integrated Manuf. **76**, 102306 (2022)

4. Cesta, A., Orlandini, A., Umbrico, A.: Fostering robust human-robot collaboration through AI task planning. Procedia CIRP **72**, 1045–1050 (2018)
5. Salunkhe, O., Stahre, J., Romero, D., Li, D., Johansson, B.: Specifying task allocation in automotive wire harness assembly stations for Human-Robot Collaboration. Comput. Ind. Eng. **184**, 109572 (2023)
6. Koukas, S., Kousi, N., Aivaliotis, S., Michalos, G., Bröchler, R., Makris, S.: ODIN architecture enabling reconfigurable human–robot based production lines. Procedia CIRP **107**, 1403–1408 (2022)

Symbiotic Human-Robot Collaboration: The FELICE Approach in Smart Assembly Lines

Dimitrios Kalogeras[1(✉)], Maria Pateraki[1,2], Sharath Chandra Akkaladevi[3], and Bartlomiej Stanczyk[4]

[1] Institute of Communication and Computer Systems, Athens, Greece
dkalo@noc.ntua.gr

[2] Laboratory of Photogrammetry, School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens, Athens, Greece

[3] Profactor GmbH, Im Stadtgut D1, 4407 Steyr-Gleink, Austria

[4] Accrea Engineering, Im Stadtgut D1, 4407 Steyr-Gleink, Austria

Abstract. This paper outlines a novel collaborative robot system designed for smart manufacturing, emphasizing coordinated interaction and skill integration between humans and robots on the assembly floor. FELICE proposes a hierarchical architecture for enhanced system manufacturing performance and improved worker ergonomics via the combination of physical sensing, digital twin modeling, and vision analytics.

Keywords: Collaborative Robots · Digital-Twin · Smart Manufacturing

1 Introduction

Today, car manufacturing [3] is the most automated sector of modern Industry, involving a broad variety of static and mobile manipulators, automated conveyor lines, and other robotized equipment. Advances in lidars, cameras and sensor technology made possible human-robot interaction (HRI) systems. Despite the increased interest in collaborative robots, operational applications in manufacturing are still not widespread. Specifically for assembly, despite the consensus that it has the highest potential for human-robot collaboration, most existing applications dubbed as being collaborative to date are either of the coexistence or sequential (or synchronized) type. In symbiotic human-robot collaboration, a coordinated, synchronous, goal-oriented activity is assumed from the human and the robot, sharing different capabilities, competences and resources. The paper introduces the FELICE system which aims to address the challenge of coordinated interaction and combination of human and robot skills in manufacturing

This work is funded under the EU Project FELICE (EU grant agreement No. 101017151).

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 319–324, 2024.
https://doi.org/10.1007/978-3-031-76428-8_59

assembly floor. FELICE envisions robots to operate safely and ergonomically alongside humans, sharing and reallocating tasks between them, allowing the reconfiguration of an assembly production process in an efficient and flexible manner.

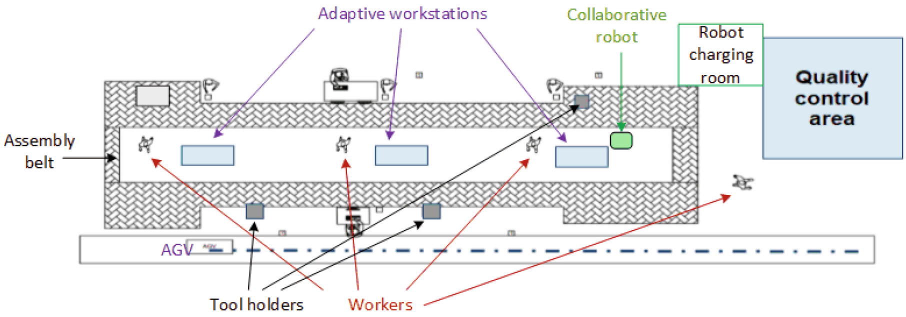


Fig. 1. Typical industrial belt type operating shop floor

2 High Level Architecture

The FELICE system's approach involves defining a workplan for each adaptive workstation (AWS) and assigning the robot to collaborate with the worker. A high-level controller manages robot perception, action planning, and communication with external devices, including the orchestrator, adaptive workstation, and shop floor sensors (Fig. 1). The hierarchical scheme (Fig. 2) encompasses the physical layer, receiving input from sensors on the robot, shop floor, multiple workstations, and human operators with wearable sensors. System actuators, such as the manipulator, gripper, mobile platform, and adaptive workstation actuators, are situated at this level. The *Local layer utilizes physical sensing for environment perception, including scene mapping, object detection, and localization, as well as human behavior modeling and cooperation. The* Global layer analyzes and organizes local layer information through digital twin modeling, assembly orchestration, optimization, and analytics.

Information flows from sensors on the robot, worker, and adaptive workstation through specific functional blocks. The Orchestrator evaluates workstation characteristics and optimizes the distribution of assembly tasks between the human worker and the robot while considering fatigue minimization and workflow duration time. If the worker explicitly requests robot support, the Orchestrator decides whether the worker should take a break or be assisted by a robot, directing the robot accordingly to a new AWS.

2.1 User Requirements and Functional Design

In the early stages of FELICE system development, use case scenarios are analyzed, leading to the decomposition of robot and adaptive workstation systems into a functional architecture with their interfaces. This process produced the physical architecture design coupled with environmental and regulatory constraints as well as with functional design. The requirements analysis identified key functional components, including perception components (e.g., image processing, scene mapping), interaction components for communication with workers, decision-making components in the global orchestrator, robot action components for task planning, execution and monitoring. These are shown in Fig. 2 and are described afterwards.

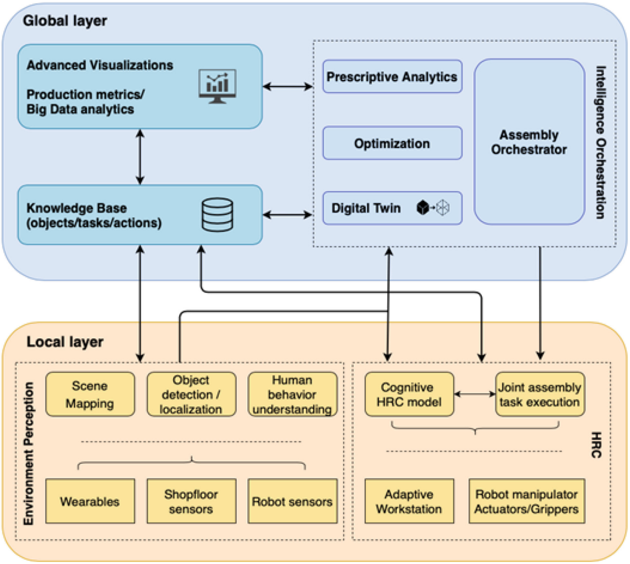


Fig. 2. FELICE Hierarchical scheme comprising of a Global and Local layer

Workflow optimizer performs optimization based on several KPIs (such as total duration for workflow execution or ergonomic penalty) thus resulting in a selection of Pareto-optimal, fine-tuned workflows. These workflows are consequently consumed by the orchestrator module thus enabling the creation of "optimal" schedules based on different demands.

Orchestrator + Intelligence ensures the fulfilment of the productivity targets via a planning model with an associated solution execution runtime engine.

Robot action planning and execution validates the planning and execution of robot tasks and verifies the partial, short-term goals which have been accomplished. It mainly consists of three components that undertake the navigation of the robotic platform, the movement of the robotic arm to reach a tool or object, and the grasping and manipulation of a tool or object.

Object detection/localization identifies and estimates the location of tools and objects present in a scene and at the same time, estimates with high precision in real time the object 6D pose, which is essential for efficient grasping. The implementation was originally based on the EPOS method [6] with training data synthetically rendered [9] using OpenGL's shader pipeline against a black (or white) background from different camera viewpoints, thus densely sampling different 6D poses of the object. Deep machine learning optimization with ONNX and TensorRT PTQ (Post training Quantization) [5] minimize the inference latency on the Jetson AGX of the cobot.

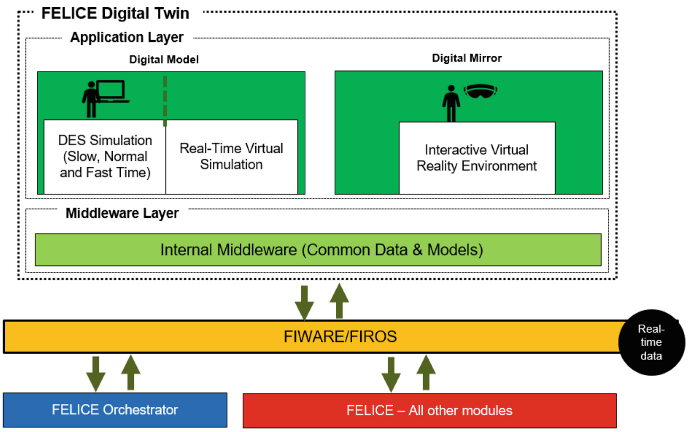


Fig. 3. FELICE-DT module

Scene perception continuously estimates the location and orientation of the cobot relying on an RGB-D stream from the navigation camera with a custom visual Simultaneous Localization and Mapping (vSLAM) algorithm [4]. The developed vSLAM estimates a 6DoF pose coupled with a Poser [7].
HRI Decision Maker resolves the short-term human-robot interaction issues that may arise during the execution of collaborative activities.
Speech synthesis performs text-to-speech synthesis to be used by the robot as an additional output communication modality. This module receives the command recognized by the speech & gesture analysis module and reproduces it with the speaker to provide the user with acoustic feedback.
Human behavior understanding visually detects the location and 3D skeleton-based, of 25 body joints, representation of a worker using the OpenPose [1] deep learning-based method. Subsequently, MURI posture scores are evaluated. Any deviations in terms of ergonomics trigger orchestration actions.
Analytics provides predictive models for cognitive and physical state of human workers as well as the productivity of the assembly line from data residing in the knowledge base.
Digital Twin provides a simulation environment (i.e. Digital Model) of the Orchestrator and an interactive virtual reality environment (called Digital Mirror) that can be used for reasoning and monitoring purposes. The Digital Model as shown in Fig. 3 includes a discrete event simulation model based on ema for

the evaluation of fast time simulations (what-if analysis). The real-time virtual simulation based on NVIDIA Isaac Sim as in [8] is a validation environment for the Orchestrator due to simulated physics and interaction.
Communication infrastructure utilizes ROS for in cobot communication and FIWARE [2] between distributed components.
Knowledge Base (KB) serves as a repository for mining and managing all the different types of information produced by the various components.

3 Conclusion

In summary, collaborative robots alleviate strenuous tasks through integrated technologies from traditional robotics to advanced speech analysis and optimization with modern sensors. The FELICE system lays the groundwork for collaborative robotics advancements, emphasizing ongoing refinement, expanded applications, and enhanced adaptability in diverse manufacturing environments. Continuous optimization, machine learning integration, and emerging technologies remain pivotal for further enhancing FELICE's capabilities.

Future research will delve into more sophisticated HRI techniques, including natural language processing, and context-aware communication. These advancements would contribute to more intuitive and seamless collaboration on the assembly floor. The FELICE system will adapt and optimize its performance based on real-time feedback, changing work conditions, and evolving manufacturing requirements. Future research will also focus on enhancing the robustness and adaptability of the collaborative robot to handle unforeseen events, dynamic task allocation, and seamless transitions between different assembly processes.

References

1. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y.: Openpose: realtime multi-person 2d pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. **43**(1), 172–186 (2019)
2. Cirillo, F., Solmaz, G., Berz, E.L., Bauer, M., Cheng, B., Kovacs, E.: A standard-based open source IoT platform: FIWARE. IEEE Internet Things Mag. **2**(3), 12–18 (2019)
3. Du, G., Wang, K., Lian, S., Zhao, K.: Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. Artif. Intell. Rev. 1–58 (2020)
4. Gao, X., Zhang, T.: Introduction to Visual SLAM - From Theory to Practice. Springer, Singapore (2021)
5. Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M.W., Keutzer, K.: A survey of quantization methods for efficient neural network inference (2021)
6. Hodan, T., Barath, D., Matas, J.: EPOS: estimating 6D pose of objects with symmetries. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
7. Lowry, S., et al.: Visual place recognition: a survey. IEEE Trans. Rob. **32**(1), 1–19 (2016)

8. Monteiro, F.F., Vieira, A.L.B., Teixeira, J.M.X.N., Teichrieb, V.: Simulating real robots in virtual environments using nvidia's isaac sdk, Anais Estendidos do Simpósio de Realidade Virtual e Aumentada (SVR) (2019)
9. Pateraki, M., Sapoutzoglou, P., Lourakis, M.I.: Crane spreader pose estimation from a single view. In: Proceedings of the 18th International Joint Conference on Computer Vision, vol. 5, pp. 796–805. SCITEPRESS (2023)

Preliminary Evaluation of an Embedded FBG-Based Force Sensor for In-Hand Grasp Monitoring

Jawad Masood(✉), Abel F. Alonso, Joaquín A. Muruzabal, and Tania G. González

Aimen Technology Center, 36418 O Porriño, Spain
{jawad.masood,abel.feijoo,joaquin.ascorbe,tania.grandal}@aimen.es

Abstract. During the object grasping process, variations in actual and calculated contact point location can occasionally occur. These variations are attributed to overestimations, manipulator dynamics, calculation errors, environmental factors, or the shape of the object. To mitigate this issue and facilitate continuous in-hand grasp monitoring an embedded Fiber Bragg Grating (FBG)-based force sensor was developed with low latency and high sensitivity. This paper presents preliminary findings on the material selection and the correlation between applied contact force with the sensor response. This sensor has been designed for monitoring continuous grasp quality across multi-finger grasps. A systematic set of experiments was conducted with the assistance of a robotic manipulator. The experiment involved applying a constant travel distance of the probe relative to the gripper finger to record the corresponding wavelength response. Our results indicate the sensor's behavior with two different embedded materials with Shore A hardness of 28 and 55. The harder material exhibited greater sensitivity and coverage. Additionally, an excellent correlation and linearity between the forces applied on the sensor and the corresponding sensor wavelength was discovered. Future research will aim at the grasping force localization relative to the sensor.

Keywords: Fiber Bragg Grating (FBG)-based force sensor · continuous grasp quality monitoring · multi-finger gripper

1 Introduction

Continuous monitoring of gripper and object contact force is crucial for effective grasping and manipulation. It assists in managing variations in real grasp points, accounting for overestimation, manipulator dynamics, calculation errors, environmental factors, or the object's shape. Two significant characteristics of grasp restraint are force closure and form closure. They are vital for maintaining object equilibrium and controlling the position and orientation of the grasped object relative to the gripper fingers. Quality monitoring plays a key role in

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 325–330, 2024.
https://doi.org/10.1007/978-3-031-76428-8_60

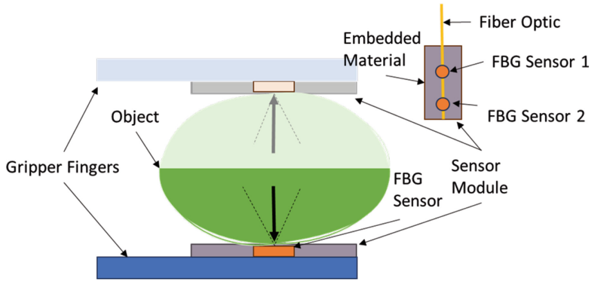


Fig. 1. The illustration of the sensor module

achieving stable grasping that satisfies friction cone constraints while ensuring appropriate contact forces without causing damage to the object [1].

Quality monitoring of grasp can optimize through two methods: dynamic sliding from initial to final grasp configuration, and in-hand manipulation of the grasped object by adapting the gripper configuration. This research aims to target the former by integrating force and torque sensors into industrial grippers.

In literature, considerable effort has been expended in developing force and torque continuous monitoring based on FBG-based force sensing technology [2–4]. Some of these studies have concentrated on robotic joints and multidimensional force and torque monitoring [2], while others have focused on robotic applications such as collaboration and manipulation [4]. Recent advancements have seen the integration of FBG sensors within robotic grippers, enabling the detection of the motion of gripping and sliding objects during manipulation [3], where, the FBG sensing primarily relied on the deformation of the gripper finger itself, which was integrated along the length of the finger instead of a localized sensor module. To identify potential slippages, this presented paper is geared towards developing modular and portable sensors with high sensitivity and low latency acquisition abilities. These sensors utilize embedded FBG, offering the flexibility to adapt to any type of gripper finger.

The structure of the remaining sections of this paper is as follows: Sect. 2 delves into the methodology employed, including the concept of the gripper, sensor integration, the experimental setup, testing procedures, and data processing. Section 3 then presents and analyzes the design of experiment results, and the calibration of force sensors. Finally, Sect. 4 concludes the paper by summarizing the findings.

2 Methodology

2.1 Conceptual Gripper Finger and FBG Sensor Integration

The sensor module with the given FBG sensor is sandwiched between the two fingers and the object as shown in Fig. 1. The target objects will be the plastic panels, small C-shaped rings, hubs, a spindle, and a wrench. The black arrow shows the normal force at the contact point with the dashed lines mentioning the friction cone. The force closure limits were determined based on the weight of the objects, ranging between 13N and 55N. The lower half symmetrical part shows the gripper being considered was an industrial multi-finger force closure gripper, which operates on a friction-based principle and features modular finger configurations. The sensor module was embedded with multiple FBG sensors. The considered ones were the sensor 1 and 2 as mentioned in Fig. 1. The sensors were embedded into the silicon with Shore A hardness of 28 and 55. The sensors were equally distanced from the center of the sensor module. The fabrication procedure of the sensor module is outside the scope of this paper.

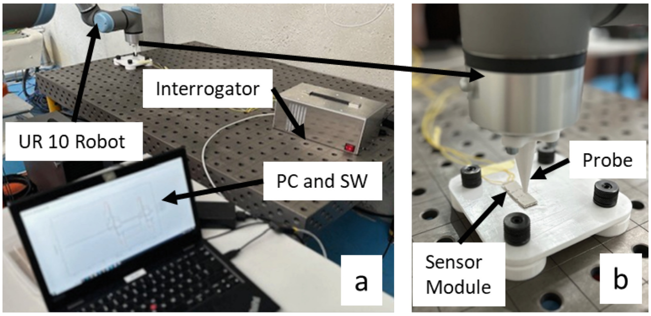


Fig. 2. Setup for data collection. The sensor is connected to the interrogator and PC

2.2 Experimental Setup and Testing

Figure 2 illustrates the experimental setup for data collection. It utilizes a prototype sensor currently under development at AIMEN facilities. The sensor was linked to an interrogator (Model Hyperion, with a hardware acquisition rate of 1000 Hz and wavelength tracking of 400 pm per acquisition) and a personal computer. External forces were applied using a universal robot UR 10, with the force readings obtained through an internal force sensor. The sensor module was secured to a table using a custom-designed jig. The interrogator collected the sensor signals, which were then processed by the manufacturer's software. The experiment involved the systematic application of forces between the 13N and 55N, oriented perpendicular to the sensor module. The sensor module was segmented into 5mm intervals. The application of each force was linked to the travel distance of the probe attached to the robot TCP, which moved along the sensor module's top surface. The force was maintained for 5 s before moving to the next location on the sensor module after a pause of 10 s.

2.3 Data Processing

The software ENLIGHT V.1.14.0.0 was used to process data from a single-channel interrogator. Initial noise was introduced to distinguish between sensor readings. As illustrated in Fig. 3, the starting wavelength for both sensors is below zero, and they differ from each other. The time series data at a frequency of 1 kHz were recorded. Additionally, the applied force data were collected from the robot software as a time series at 200 Hz. To calculate the linear regression, five data points were used corresponding to the sensor 1 and 2. This involved taking the mean values for the applied forces and their corresponding wavelengths, considering only instances when the force was applied perpendicular to the sensor.

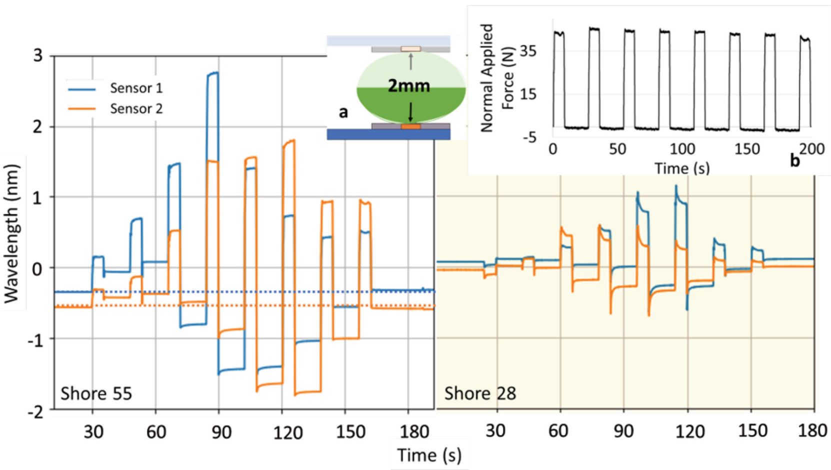


Fig. 3. The wavelength response for the sensors. a. illustrates the 2 mm travel between the object and the sensor module, b. illustrates the eight cycles of the applied force

3 Results

The design of experiments yielded three primary outcomes: the impact of silicon hardness on FBG sensor robustness, the force distribution on the sensor surface, and the calibration parameter calculation for normal force measurement.

Shore 55 material showed superior resolution compared to Shore 28 as shown in Fig. 3, maintaining a constant force profile for eight cycles, and causing similar settling noise when the object touched the sensor. Sensor performance remained consistent throughout trials, with uniform recovery time and similar patterns regardless of the force applied.

Force distribution across the two sensors was observed as shown in Fig. 3, with simultaneous detection across sensors when force was applied near one. Combined readings from all embedded sensors revealed the force distribution throughout the sensing module.

Calibration findings showed a strong linear relationship between the normalized applied force and wavelengths, with high linearity for both sensors as shown in Fig. 4. Both sensors display linearity above 0.96. Furthermore, the correlation between the applied force and the corresponding wavelength for both sensors is 0.98. Similar calculations can be extended for other sensors on the module.

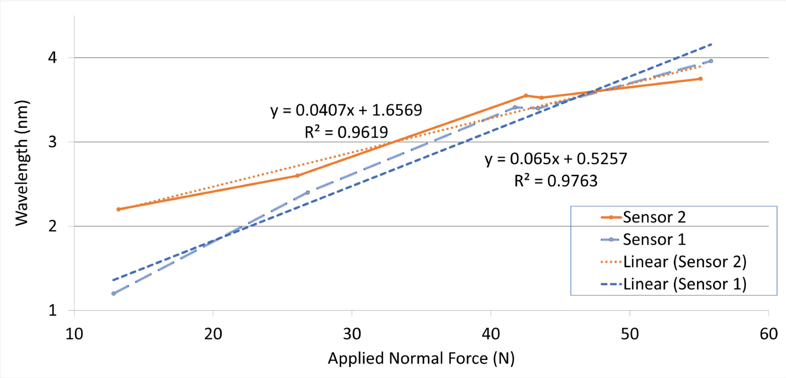


Fig. 4. Linear regression of sensor 1 and sensor 2

4 Conclusions

This paper delineates a systematic experimental design to measure the response of the FBG-embedded sensor for monitoring gripper-object contact quality. The findings suggest that material hardness significantly impacts the sensor module's sensitivity and coverage. The sensors demonstrate the capability to detect forces both locally and globally. Excellent correlation and linearity between the applied forces and the corresponding sensor wavelength were observed. Future research will focus on addressing the localization problem through global sensor module investigation. Additionally, efforts will be directed towards mitigating sensor cost and temperature effects, which are significant technological challenges.

Acknowledgements. This research was partly funded by the HARTU project, which received funding from the European Union's research and innovation programme Horizon Europe under the grant agreement No. 101092100.

References

1. Roa, M.A., Suárez, R.: Grasp quality measures: review and performance. Auton. Robots **38**, 65–88 (2015). https://doi.org/10.1007/s10514-014-9402-3

2. Xiong, L., Guo, Y., Jiang, G., Zhou, X., Jiang, L., Liu, H.: Six-dimensional force/torque sensor based on fiber Bragg gratings with low coupling. IEEE Trans. on Ind. Electronics **68**(5), 4079–4089 (2020). https://doi.org/10.1109/TIE.2020.2982107
3. Tripicchio, P., D'Avella, S., Avizzano, C.A., Velha, P.: Towards robust grasping: an analysis of in-hand object motion with FBG optical fibers as force sensing technology. Mechatronics **93**, 102990 (2023). https://doi.org/10.1016/j.mechatronics.2023.102990
4. Long, J., Liang, Q., Sun, W., Wang, Y., Zhang, D.: Ultrathin three-axis FBG wrist force sensor for collaborative robots. IEEE Trans. Instrum. Meas. **70**, 1–15 (2021). https://doi.org/10.1109/TIM.2021.3086898

Robotic Grasping Decision Making Assisted by AI and Simulation

Jon Ander Ruiz(✉), Ander Iriondo, Andoni Rivera, Ander Ansuategi, and Iñaki Maurtua

Department of Autonomous and Intelligent Systems, Tekniker - Basque Research and Technology Alliance (BRTA), Iñaki Goenaga 5, 20600 Eibar, Gipuzkoa, Spain
{jonander.ruiz,ander.iriondo,andoni.rivera, ander.ansuategi,inaki.maurtua}@tekniker.es

Abstract. Artificial intelligence (AI) and simulation technologies are currently experiencing a rise within robotic manipulation. This work introduces the autonomous grasping point detection (GPD) system proposed in the HARTU project, which is assisted by simulation and AI. We propose a two stage pipeline: In the first one grasping points are extracted with deep learning (DL) techniques and analytical sampling methods in isolated objects, and tested in simulation. Then, we develop a deep reinforcement learning (DRL) decision module, trained in simulation, to select the best grasping points in complex scenes. The whole pipeline is integrated within the ROS2 ecosystem.

Keywords: Grasping · simulation · manipulation · artificial intelligence

1 Introduction

The grasping point detection problem involves the complex task of enabling robots to autonomously identify suitable locations on diverse objects for effective grasping. The challenge arises from the variability in object shapes, sizes, and materials, as well as the inherent noise and uncertainty in sensor data. Ambiguity in object poses, real-time processing constraints, cluttered environments, and the need for generalisation across objects further contribute to the difficulty.

Current lines of research seek to improve the resilience of robotic manipulation systems using DL techniques to obtain the grasping points in objects [1]. In addition, another prominent line of research is the use of reinforcement learning (RL) to train robotic arms [2]. This training phase is, in a good amount, done in simulations, as they provide a safe space where all the parameters can be modified. Those simulation technologies have been experiencing a rise on popularity. For instance, Gazebo, PyBullet, Nvidia Isaac and MuJoCo are now frequently used to test such machine learning algorithms [3]. This work introduces the HARTU[1] project's GPD system, employing AI for decision-making and simulation for grasp validation and DRL agent training.

[1] https://www.hartu-project.eu/.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 331–335, 2024.
https://doi.org/10.1007/978-3-031-76428-8_61

2 System Overview

This work addresses the mono-reference random bin-picking problem. Assuming the availability of a computer-aided design (CAD) model of the part, the system employs a two-stage process, involving a local grasp planner for isolated items and a global planner for complex scenes. The goal is to provide a tool to automatically extract grasping points that are actually valid for the 5 HARTU industrial use cases, and for any provided CAD. All the system is integrated into the Robot Operating System 2 (ROS2) ecosystem and each module has been implemented as a ROS2 node.

2.1 Simulation

In the context of this project, the simulation is used with three main tasks in mind; first to aid computer vision and segmentation algorithms by providing realistic synthetic data to train them, second to validate the grasping candidates for isolated objects, and third to train an agent to learn the object picking order in a cluttered environment. Thus, we consider two main aspects when developing a simulation environment: Visual and physics realism.

Due to the difficulty on acquiring a large, real image dataset, the visual simulation, developed in Unity[2], allows to increase the set of training images for the segmentation algorithms. It provides the DL algorithms with realistic synthetic images, along with segmented labels, that are used to train a generic segmentation model. To close the gap between real images and computer generated images, Physically Based Rendering textures are used. The simulation provides domain randomisation options in order to improve the quality of the data.

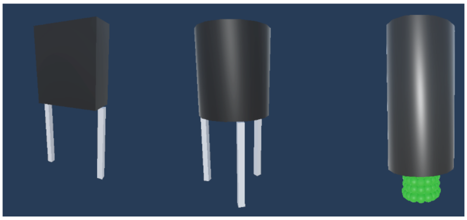


Fig. 1. Grippers. From left to right: Two finger, three finger and suction gripper.

The physics simulation in HARTU is developed in Unity, paired with the MuJoCo physics engine thanks to the Unity-MuJoCo plug-in[3] developed by DeepMind, and integrated within the ROS2 ecosystem thanks to our Unity-ROS2-Control software. This combination has been chosen as MuJoCo is state-of-the-art on contact simulation, open source and offers a Unity plug-in while Unity provides great visual realism and tools to connect to ROS2.

[2] https://unity.com.
[3] https://mujoco.readthedocs.io/en/stable/unity.html.

Among MuJoCo's best features is the extensive set of parameters that allow the precise modelling of the behaviour of two objects in contact. Within the context of this project, a set of grippers have been modeled, as seen in Fig. 1.

2.2 Grasp Planning

The local grasp planner is an automatic system that receives as input both the CAD model of the element to be picked and the specifications of the end effector, and outputs the valid grasping points in XML format. The complete pipeline is depicted in Fig. 2.

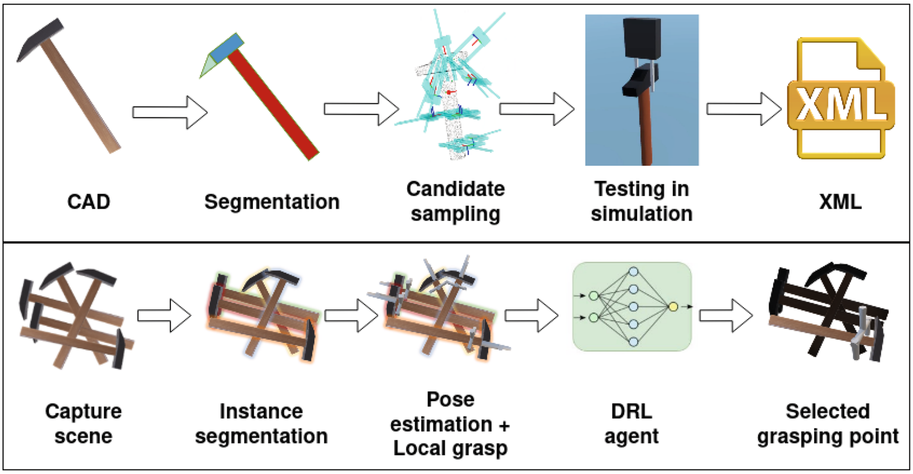


Fig. 2. From top to bottom: Local and global grasping pipelines.

The process is as follows: The planner first converts the CAD into a point cloud and segments it using SAM3D [4]. Then, sampling algorithms are used to get potentially good grasping points just considering the geometry of both the item and the gripper. The segmentation helps us to uniformly sample candidates in all the regions of interest of the object. To that end two main sampling algorithms are used: For two parallel finger grippers we use the method proposed in [5], and we also extended it for three finger grippers. This method looks for gripper positions where the fingers do not collide with the CAD and at least a portion of it falls inside the fingers. For suction/magnetic end effectors we use the algorithm proposed in [6], which finds flat surfaces in the CAD where the end effector can successfully actuate. Then, each candidate is assessed in the realistic physics simulator to get only the ones that lead to a robust grasp. Finally, the selected grasping points are stored in the XML file. This process is executed just once per each item/end effector pair.

Although at this point it is known how each object can be grasped, in random scenes with multiple instances of the object, a policy is required to decide which grasping point is the best one in each situation. In HARTU we propose to use DRL to model an agent that decides which potential grasping point is the best one in each iteration of the picking process.

The DRL agent is trained in simulation. The procedure is as follows: First, a random bin-picking scene is generated with multiple instances of the CAD model. After estimating the pose of all that are not occluded, i.e. are visible, all the potential grasping candidates are extracted using the local grasp plan. In that moment comes into play the global planner. Having as input all the candidates and a set of features describing the scene, it decides which one is the most appropriate one for that scene. This grasping point is executed and assessed in simulation, and the feedback is used to reward or penalise the DRL agent. This process is repeated iteratively until no objects are left inside the bin.

3 Preliminary Results

The main work has been focused on the development and validation of the local grasp planning pipeline for several end effectors and objects. The first approach to simulate suction and grippers using Unity and MuJoCo, as seen on Fig. 3, shows promising results. In both cases the goal has been to assess the grasping candidates generated by the DL and analytical sampling algorithms, finally to select the most robust ones.

(a) Suction end effector. (b) Two finger gripper.

Fig. 3. Grasping candidate testing in simulation.

In the case of the suction, the ability of the suction cup to adapt to the shape of the surface it is in contact with, and using sensors to determine if contact is made, regulates the "suction" force. If no contact is detected on some point of the suction cup the simulation does not generate force, as it is considered that no vacuum is achieved. Regarding the gripper, both opening and closing grasps have been considered. In both cases, the gripper closes/opens the fingers with a predefined force until it contacts the object.

Both for suction and gripper, the grasp robustness is measured tracking the pose of the object after the grasp has been executed. The smaller the oscillation of the pose, the more robust the grasp is.

4 Conclusion and Future Work

Current results are promising, though further developments and testing is needed to make the full GPD pipeline work. More specifically, the local grasp planner has been successfully tested with multiple objects and has demonstrated the ability to accurately estimate grasping points in isolated objects. The main current work focuses on the development and extensive testing of the global planner. The main difficulty at present is the correct modelling of scenes with a large number of contacts. It has been shown that the correct modelling of the grasping objects is essential both in physical parameters and correctly defining the collision mesh itself.

Future work includes the optimisation of the simulation, improving contact realism and testing it by comparing with a real ground truth in a variety of different contact scenarios, besides to integrating the DRL agent and simulation.

Acknowledgement. This work has been partially funded by HARTU project that has received funding from the European Union's research and innovation programme Horizon Europe under the grant agreement No. 101092100, and the project "5R- Red Cervera de Tecnologías robóticas en fabricación inteligente", contract number CER-20211007, under "Centros Tecnológicos de Excelencia Cervera" programme funded by "The Centre for the Development of Industrial Technology (CDTI)".

References

1. Tian, H., Song, K., Li, S., Ma, S., Jing, X., Yan, Y.: Data-driven robotic visual grasping detection for unknown objects: a problem-oriented review. Expert Syst. Appl. **211**, 118624 (2023)
2. Imtiaz, M.B., Qiao, Y., Lee, B.: Prehensile and non-prehensile robotic pick-and-place of objects in clutter using deep reinforcement learning. Sensors **23**(3), 1513 (2023)
3. Collins, J., Chand, S., Vanderkop, A., Howard, D.: A review of physics simulators for robotic applications. IEEE Access **9**, 51416–51431 (2021)
4. Yang, Y., Wu, X., He, T., Zhao, H., Liu, X.: Sam3d: segment anything in 3d scenes. *arXiv preprint*arXiv:2306.03908 (2023)
5. Ten Pas, A., Gualtieri, M., Saenko, K., Platt, R.: Grasp pose detection in point clouds. Int. J. Robot. Res. **36**(13–14), 1455–1473 (2017)
6. Wan, W., Harada, K., Kanehiro, F.: Planning grasps with suction cups and parallel grippers using superimposed segmentation of object meshes. IEEE Trans. Rob. **37**(1), 166–184 (2020)

An Approach of Automated Assembly Evaluation Using AI - Based Computer Vision Methods for Human – Robot Collaboration

Konstantinos Katsampiris-Salgado, Nikos Dimitropoulos, Alexandros Kanakis, George Michalos, and Sotiris Makris(✉)

Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Patras, Greece
makris@lms.mech.upatras.gr

Abstract. Artificial Intelligence (AI) and Machine Vision (MV) methods are increasingly applied in manufacturing systems as this enables automated processes that lead to both cost and time efficient solutions, minimizing errors and maintaining the quality of the products in high level. This paper discusses an AI – based vision system that is responsible for the evaluation of an assembly performed in the context of a hybrid manufacturing system. Different object detection models are evaluated for the detection of the parts, utilizing both real-life and synthetic datasets, while a fully integrated implementation using ROS is presented. Finally, the solution is applied and evaluated in a Human – Robot Collaboration (HRC) based assembly scenario.

Keywords: Artificial Intelligence · Machine Vision · Object Detection · Assembly Evaluation · Human-Robot Collaboration

1 Introduction

The requirement for advanced manufacturing systems is highly motivated from the need to cope with complex manufacturing processes, dealing with the demand for highly reconfigurable production systems that are capable of producing assorted products in various scales in terms of production size [1]. In the context of flexible and reconfigurable manufacturing, cooperating robots have acquired great research focus, either by the aspect of robot-to-robot or robot-to-human cooperation [2].To enhance robot capabilities and specifically their surroundings perception, machine vision is extensively used in industrial applications [3], with reference examples in vision-based inspection systems [4], process control, parts detection and identification, vision guided robotic systems [5] and safety optical systems for human operators' protection [6], especially in cases of HRC. While collaborative robots perform assembly operations precisely, they lack the knowledge to evaluate the assembly validity of the final product, letting humans the responsible resource to validate the assembly [7]. Perceptually engaging tasks are an important factor that offers high potential for human error, while in addition, operator's fatigue [8], is also an important predictor of error, and it highly depends on the mental

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 336–340, 2024.
https://doi.org/10.1007/978-3-031-76428-8_62

demands of the task. To maintain assembly quality high, it is highly suggested to develop tools that can automate the assessment, and the utilization of AI based vision methods may facilitate this task, as per the existing literature, thus improving both the quality and the efficiency of manual assembly [9, 10]. Though, to the authors' knowledge, in the current literature no assembly verification methods exist, that can perform the evaluation in an HRC scenario, for different product variants that may be differently placed in the assembly workbench.

Considering the above, the authors propose a method that combines different state of the art technologies such as machine learning, computer vision algorithms and artificial dataset generation, in order to detect correct part placement in the assembly process and aid the human – robot collaboration. The only equipment required for the implementation of the proposed method is a standard RGB camera, expanding the applicability of the solution. This method is also fully integrated with a Robot Operating System (ROS) based framework in a real industrial HRC application.

2 Approach

The proposed method evaluates product assembly using images from an RGB sensor, assessing the correct count, position, and orientation of parts. It employs traditional computer vision algorithms and AI-based models like SSD and YOLO, trained on custom datasets featuring varied angles, positions, orientations, and lighting conditions. Both synthetic and real-life datasets are used, with synthetic datasets being particularly beneficial for handling lighting variations and enhancing model evaluation. Image augmentations such as shearing, rotations, and HSV transforms are applied to increase dataset diversity and reduce overtraining and false positives [11].

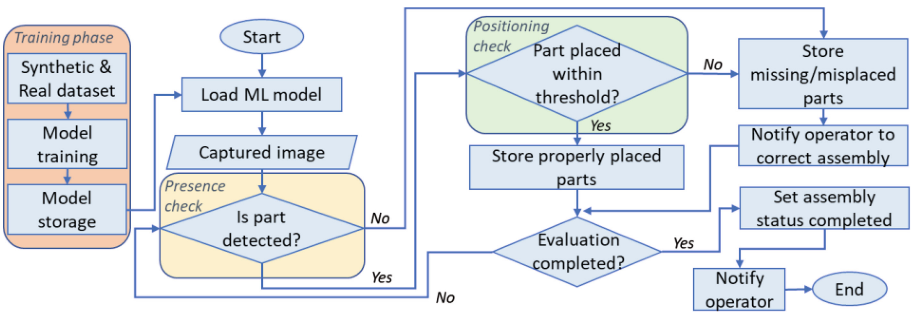


Fig. 1. Assembly evaluation flowchart

The method requires a complete assembly image to ascertain expected component positions and assesses part orientation, with some parts being undetectable if misoriented. Communication between the system and the operator is managed through a TCP/IP server-client setup, facilitating bidirectional communication. The system begins evaluation after assembly completion and informs the operator of any verification issues, allowing for repeated assessments if necessary. The approach is illustrated in Fig. 1.

3 Implementation

3.1 Dataset Creation and Augmentation

The development of a capable dataset for object detection model training necessitates a large volume of accurately labeled images, a challenge intensified by varying lighting conditions and the reflections from metallic parts. To address this, we utilized Blender, a 3D modeling tool, for synthetic dataset creation, using the realistic RGB image renders it can create. This approach offers significant advantages: it eliminates the need for manual image capture and labeling, allows easy manipulation of lighting, background, and part orientations, and supports scripting for automatic changes in scene settings and labeling. The precision of Blender in annotating CAD files ensures high accuracy and consistency, surpassing manual labeling efforts. In parallel, a real-life dataset was assembled using manually captured and labeled images. These images were taken in conditions mirroring the deployment environment, including varying light and angles. Unlabeled, irrelevant parts in these images served as negative examples, aiding in reducing false positives. Additionally, both synthetic and real-life datasets benefited from image augmentations like HSV transforms, shearing, and flipping. These techniques not only expanded the dataset size but also prevented overfitting. By applying these augmentations to already labeled images, the dataset's size was effectively doubled without the need for additional labelling.

3.2 Object Detection and Position Estimation

The authors chose a detection model based on two criteria: detection speed and performance, opting for YOLOv5 (medium model) [12] over SSDv2 [12] due to its higher recall and mean average precision on the available case specific datasets. Detection models with complex, multi-stage pipelines were avoided, to prevent delayed detection execution. Higher resolution images were utilized as this facilitated the detection of smaller parts on the image, compromising though detection speed.

Once the detection model is satisfactorily trained, the next step is position evaluation, carried out as follows: A reference image delineating the correct final assembly positions of parts is used. Parts are expected to appear within specified rectangular areas in this image, with the corners of each area clearly defined. The actual detected parts' bounding boxes are compared with these predefined areas. A part is considered correctly placed if the difference between the bounding box and the area is within a set tolerance threshold. A smaller threshold implies a need for more precise placement. Incorrectly positioned or undetected parts in the designated areas signify an assembly error, and the names of such parts are recorded in the verification results. This method also helps in identifying orientation errors. Non-symmetrical parts placed in the wrong orientation will have mismatched bounding box proportions compared to the expected area, leading to their detection as missing or incorrectly placed. The system, integrating a Process Orchestrator, Assembly Evaluator, and Sensor Layer, uses ROS and TCP/IP communications to evaluate assembly operations with a machine learning model, and communicates any errors or adjustments needed to the operator to take action.

4 Case Study

The industrial use case involves the transformation of a manual kitchen hob assembly process derived from the white goods domain to an human robot collaborative system involving a robotic worker supporting a human operator. The current process, primarily manual, includes tasks like preparation of base plates, cable assembly, and screwing, for two product variants. The future vision aims to introduce a dual-arm humanoid robot to collaborate with human operators. This robot will assist in demanding tasks like cable handling and screwing, improving assembly efficiency and ergonomics.

An Intel RealSense camera, acting as an RGB sensor, monitors the panel for component detection. During the assembly execution the assembly evaluation system, monitors component's placement on the line, preventing at the time any misplaced or forgotten component. In the end the system's orchestrator requests a final evaluation, receiving back a message indicating success or detailing any missing or misplaced parts.

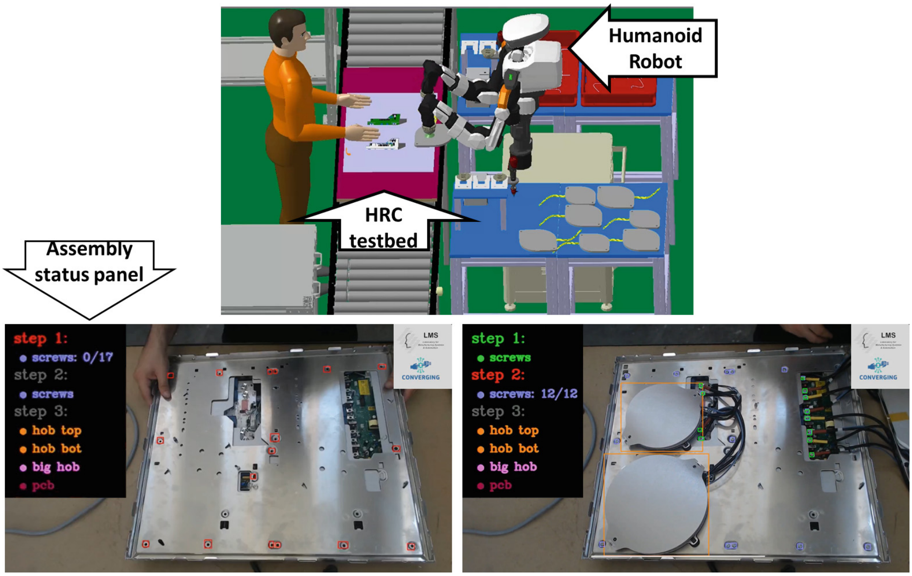


Fig. 2. Conceptual HCR testbed (upper) and different assembly evaluations stages (lower).

In Fig. 2, different captures of the assembly evaluation are presented along with the use case testbed. In the preliminary testbed, the solution seems to offer promising results, accurately detecting assembly mistakes. The solution will be further validated in the end user premises where extensive tests will be conducted.

5 Conclusions and Future Work

The experiments conducted demonstrated that the proposed evaluation method may accurately and consistently evaluate the assembly of both panels tested. The evaluation may also be conducted at any step of the assembly process, as this facilitates the evaluation in cases where overlapping parts should be previously checked.

This implementation aims to reduce cycle time and possible human errors, in both the assembly and the evaluation stages. It is also important to point out that it is a low-cost technique as it may run with just a standard RGB sensor. Moreover, the utilization of ROS for the implementation makes it easy for integrating most ROS based solutions.

In future work, the idea of synthetic datasets may be examined more thoroughly. Also, different object detection models, with different training times and detection speeds, could be explored according to each use case.

Acknowledgements. This study was funded by the HEU Project "CONVERGING Social industrial collaborative environments integrating AI, Big Data and Robotics for smart manufacturing" (Grant Agreement: 101058521) (www.converging-project.eu), funded by the EC.

References

1. Chryssolouris, G.: Manufacturing Systems: Theory and Practice. Springer New York, New York, NY (1992)
2. Makris, S.: Cooperating Robots for Flexible Manufacturing. Springer (2021)
3. Golnabi, H., Asadpour, A.: Design and application of industrial machine vision systems. Robot. Comput.-Integrated Manuf. **23**(6), 630–637 (2007)
4. Magalhaes, P., Ferreira, N.: Inspection application in an industrial environment with collaborative robots. Automation **3**(2), 258–268 (2022)
5. Bai, Q., Li, S., Yang, J., Song, Q., Li, Z., Zhang, X.: Object detection recognition and robot grasping based on machine learning: a survey. IEEE Access **8**, 181855–181879 (2020)
6. Michalos, G., Makris, S., Eytan, A., Matthaiakis, S., Chryssolouris, G.: Robot path correction using stereo vision system. Procedia CIRP **3**, 352–357 (2012)
7. Radinger, T., Stuja, K., Wolfel, W., Markl, E.: Functional safety concept for a handling robot built on optical systems. In: Katalinic, B. (ed.) Proceedings of the 28th International DAAAM Symposium 2017, pp. 0168–0172. DAAAM International Vienna (2017). https://doi.org/10.2507/28th.daaam.proceedings.022
8. Torres, Y., Nadeau, S., Landau, K.: Classification and quantification of human error in manufacturing: a case study in complex manual assembly. Appl. Sci. **11**(2), 749 (2021)
9. Chen, C., et al.: Monitoring of assembly process using deep learning technology. Sensors **20**(15), 4208 (2020)
10. Andrianakos, G., Dimitropoulos, N., Michalos, G., Makris, S.: An approach for monitoring the execution of human based assembly operations using machine learning. Procedia CIRP **86**, 198–203 (2019)
11. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 1–48 (2019)
12. Kaur, R., Singh, S.: A comprehensive review of object detection with deep learning. Digital Signal Process. **132**, 103812 (2023)

Integrating Cyber-Physical Systems in Non-rigid Assemblies: A Composites Manufacturing Case Study

Dionisis Andronas, Konstantinos Kavvathas, Nikolaos Theodoropoulos, Emmanouil Kampourakis, Panagiotis Stylianos Kotsaris, and Sotiris Makris(✉)

Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Patras, Greece
makris@lms.mech.upatras.gr

Abstract. Despite the advances in robot control and Internet of Things (IoT) services, manufacturing operations involving the handling of non-rigid products or workpieces largely remain manual. The dynamic distortion of flexible objects underlines limitations in cognitive robot dexterity as well as the necessity for digital twin platforms where modules for (co-) manipulation planning are applied. These planning modules cannot demonstrate their full potential in terms of performance and resilience without an accurate reconstruction of the physical workstation. This need becomes more apparent in unstructured environments where human intervention and improvisation may occur, or when the assembled product exhibits stochastic behavior. This paper discusses a digital-twin framework that can address the challenges associated with non-rigid assemblies and fenceless human-robot co-existence. A use case coming from the automotive industry is used for validating the framework, by enhancing aspects of safety, interaction, perception, cognition, orchestration and performance of a hybrid cell for composites manufacturing.

Keywords: Deformable Object Modelling · Cyber-physical Systems · Digital Twin · Perception · Extended Reality

1 Introduction

Over the years, manufacturing companies have increasingly automated their production processes for a competitive advantage. With the advent of Industry 4.0, a significant shift has occurred towards factory digitalization, integrating the Internet of Things (IoT), Big Data, Artificial Intelligence (AI), and autonomous robots as core technologies [1, 2]. This digital transformation, primarily realized through cyber-physical systems, is critical in dealing with uncertainties in manufacturing, especially in workstations handling non-rigid workpieces and those requiring collaborative, fenceless operations [3]. The efficacy of these systems hinges on the precision of their digital models and the reliability and promptness of data retrieval. As such, there has been a considerable focus on developing and implementing solutions in various domains, including modeling [4], perception (both environment [5] and human [6]), planning [7], and interaction [8]. These

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 341–345, 2024.
https://doi.org/10.1007/978-3-031-76428-8_63

areas are fundamental in addressing the challenges of flexible material manipulation and ensuring seamless integration of human operators and robotic agents in a shared workspace. This paper presents a digital twin framework aiming to address challenges associated with robotic manipulation of flexible materials and human-robot coexistence in fenceless environments. The subsequent section will discuss the methodology behind this framework and explain how distinct sub-modules offer an integrated solution for composites manufacturing.

2 Framework Development and Implementation

The proposed framework (Fig. 1) offers a unified system representation, grounded in CAD models that using semantics link the capabilities and characteristics of workstation entities. Geometric data defines spatial relationships, supported by uniform data structures for communication and synthesis. Data collection encompasses robot controller logs and sensors for environment, human, and mechatronics perception. This information enables processing packages for computation or decision making Continuing, the operation of mechatronics and robots is managed by their respective hardware controllers, which are in turn driven by the system's reasoning tools. ROS is chosen as the primary middleware for efficient data exchange and scalability. Key areas where the digital twin platform is elemental include:

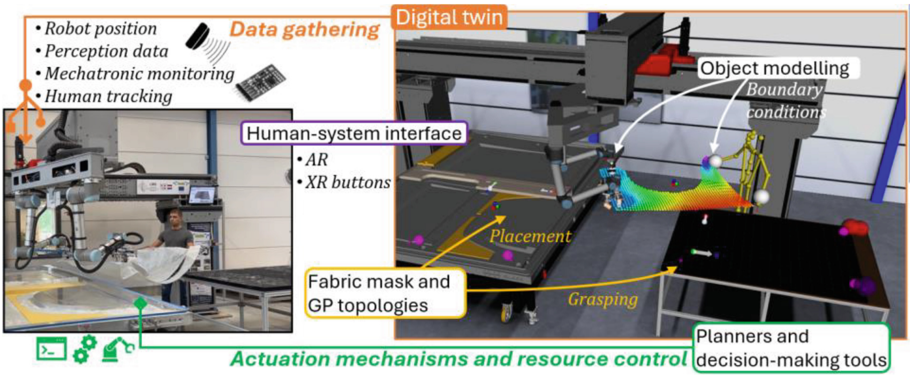


Fig. 1. Framework and technologies for data gathering, modelling, control and interaction.

- **Environment perception:** In the discussed framework, the digital scene is updated with shop floor data. A notable example is the on-robot perception system for localizing unfolded fabrics. Before grasping, the robot positions itself using the fabric's mask-defined coordinates. The perception system then calculates an offset to realign the fabric's model position in the digital twin. Subsequently, the robot adjusts its grasp based on these updated coordinates.
- **Human perception:** Within a fenceless workstation, operator tracking is essential for ensuring safety, while recognizing human activity recognition is beneficial for facilitating interaction and collaboration. Focusing on the latter, the human perception

module, utilizes the input given by a stereolabs ZED 2 stereo camera and, after sequential filtering, communicates the operator's body position skeleton coordinates. The resulting skeleton data are intended for: a) defining boundary conditions for the fabric's model, b) providing input for model-based co-manipulation, and c) functioning as a medium of interaction with extended reality buttons.

- **Modelling:** The model for simulating non-rigid parts employs Provot's mass-spring model, capable of reconstructing fabrics with diverse shapes and properties in various handling scenarios [9]. Accurate physical representation is ensured through initialization and runtime updating mechanisms. A "mask" initially defines the fabric's shape and grasping points topology. The mask's placement in the digital twin instructs operators on fabric unfolding, while robots are guided to actual grasping points. Post-detection, the model position is adjusted to compensate manual misplacement. Sensor data (robot and human palm coordinates) during operation serve as boundary conditions, maintaining alignment with the fabric's real-time spatial arrangement. To enhance the performance of the digital twin and ensure its capability to run in real time, nvidia CUDA was utilized.
- **Cognitive model-based co-manipulation:** An extension of the simulation model was implemented to enable seamless synchronization of the co-manipulation actions of the cooperative agents [9]. This extension introduces a model-based co-manipulation planner that considers: a) fabric deformations, b) operator handling inputs, and c) the specific characteristics of robot agents; all aimed at generating supportive robot trajectories that avoid fabric defects or environment collisions. By effectuating robot control, this model-based planner is one of the main actuation mechanisms of the proposed digital twin framework (Fig. 2).

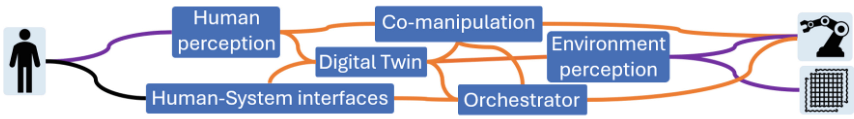


Fig. 2. Overall system architecture

- **Human-system interaction:** A primary goal of the system is to maintain high operator awareness. This is achieved by utilizing the Unity C# API to implement software that augments data from perception devices, like real-time video, with information from the digital twin. This integration creates screen-based intuitive interfaces, overlaying notifications, instructions, and digital content onto the physical environment. Additionally, extended reality buttons, designed as interactive digital spheres, are placed within the workstation and operators can interact with these by placing their palms inside for a set time, without requiring wearables. A loading graphic indicates the interaction state. These buttons' function both as acknowledgment tools or grasping interfaces.

- **Runtime orchestration:** The digital twin facilitates runtime orchestration, acting as a moderator between the orchestrator and the shop floor. For instance, by positioning each fabric's mask at the picking and placement areas of the workstation, the orchestrator enables seamless interpolation of the grasping point coordinates. Subsequently, the manipulator is directed where to locate, grasp, and release the fabric, while the operators are guided on where to unfold, grasp, and "drive" the fabric during the co-manipulation process. The co-manipulation phase concludes when the operator "hits" the placement coordinates with the fabric's centroid. This position is indicated by an augmented arrow pointing towards the placement target.

3 Case Study

The proposed framework was evaluated in an industrial environment coming from the automotive sector. The manufacturing process involves laying up fiberglass composite panels via resin infusion. This process presents ergonomic challenges for operators due to high repeatability, poor accessibility, and limited yet actual exposure to chemicals. The proposed system (Fig. 3) semi-automates this scenario while maintaining human intervention for operations requiring high dexterity and improvisation.

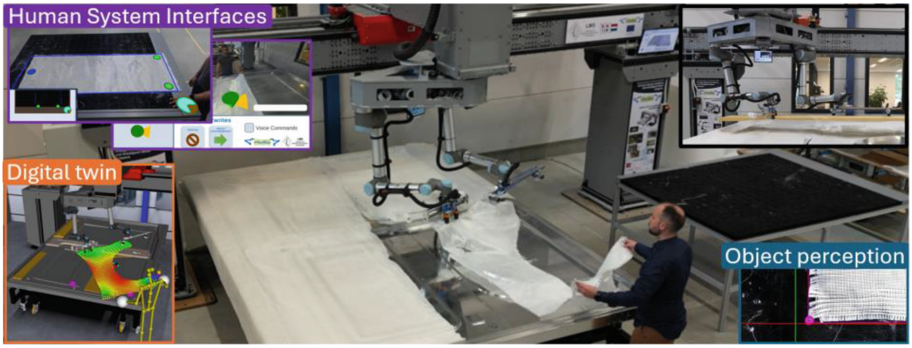


Fig. 3. Hybrid cell for composites manufacturing and validation in industrial premises.

The proposed digital twin framework is the backbone of the implemented system. All sensors, hardware, and software modules, as well as processing, planning, and orchestration packages, can seamlessly exchange data and information. The implementation of the simulation model, along with the model-based planner, allows for accurate and intelligent handling of non-rigid objects. Overall, the system has automated operations that were not feasible using conventional solutions (e.g., point to point programming). Concurrently, the implementation of high-fidelity perception systems, in combination with the digital twin, provided a solid foundation for the release of extended reality interfaces. These significantly improved the system's acceptance and usability by offering intuitive operator support content and a medium for human-to-system interaction without the need for physical buttons or wearable devices. The integration of real-time sensor data into the digital twin is reflected in the extended reality interfaces, enhancing operator situational awareness. This improved awareness allows for more informed decision-making and quicker response to potential hazards, resulting in a safer environment.

4 Conclusions and Future Work

This paper introduced a digital twin framework for cyber-physical systems, addressing challenges in non-rigid assembly and fenceless human-robot coexistence. The framework's effectiveness was validated in an automotive industry use case, demonstrating improvements in safety, interaction, and operational efficiency. User feedback indicates the existence of ergonomic improvements and reduced chemical exposure. Further operator training is expected to enhance overall performance, mental workload and user experience scores, which were promising (68.8%, 60.3%, and 65.6% respectively). Future work will focus on implementing this framework in additional industrial settings to further assess its adaptability and impact. Emphasis will be placed on industries with analogous challenges, aiming to refine the framework's capabilities in diverse environments and enhance human-robot collaboration in complex manufacturing tasks.

Acknowledgements. This work has been partially supported by the Project "MERGING, funded by the European Commission under the Horizon 2020 Research and Innovation Programme (H2020-DT-FoF-12-2019) with grant agreement ID: 869963.

References

1. Chryssolouris, G., Alexopoulos, K., Arkouli, Z.: Artificial intelligence in manufacturing systems. Stud. Syst. Dec. Control. **436**, 79–135 (2023)
2. Makris, S., Dietrich, F., Kellens, K., Hu, S.J.: Automated assembly of non-rigid objects. CIRP Annals **72**(2), 513–539 (2023)
3. Kousi, N., et al.: Digital twin for designing and reconfiguring human-robot collaborative assembly lines. Appl. Sci. **11**, 4620 (2021)
4. Theodoropoulos, N., Kampourakis, E., Andronas, D., Makris, S.: Cyber-physical systems in non-rigid assemblies: a methodology for the calibration of deformable object reconstruction models. J. Manuf. Syst. **70**, 525–537 (2023)
5. Zhou, X., et al.: Computer vision enabled building digital twin using building information model. IEEE Trans Industr Inform. **19**(3), 2684–2692 (2023)
6. Guidolin, M., Tagliapietra, L., Menegatti, E., Reggiani, M.: Hi-ROS: open-source multi-camera sensor fusion for real-time people tracking. Comput. Vis. Image Underst. **232**, 103694 (2023)
7. Kokotinis, G., Michalos, G., Arkouli, Z., Makris, S.: A behavior trees-based architecture towards operation planning in hybrid manufacturing. Int. J. Comput. Integrated Manuf. **37**(3), 324–349 (2023)
8. Andronas, D., Apostolopoulos, G., Fourtakas, N., Makris, S.: Multi-modal interfaces for natural Human-Robot Interaction. Procedia Manuf. **54**, 197–202 (2021)
9. Makris, S., Kampourakis, E., Andronas, D.: On deformable object handling: Model-based motion planning for human-robot co-manipulation. CIRP Ann. **71**(1), 29–32 (2022)

Understanding the Antiquities Market Through an AI-Driven Approach

Riccardo Giovanelli(✉) and Arianna Traviglia

Center for Cultural Heritage Technology, Istituto Italiano di Tecnologia, 30172 Venice, Italy
{riccardo.giovanelli,arianna.traviglia}@iit.it

Abstract. This paper presents AIKoGAM, an innovative AI-driven Knowledge Graph of the Antiquities Market, developed to decipher the complexities of the global antiquities trade, a realm where legal and illicit activities often intertwine. AIKoGAM is designed to tackle the dual challenges of artifact authenticity and legality in a market marked by legislative inconsistencies across countries. The system is structured into four primary modules: Information Retrieval, Data Mapping and Transformation, Event Extraction, and Knowledge Graph Construction. These modules work in tandem to collect, standardize, process, and visualize data from diverse sources like auction houses and galleries. The resultant Knowledge Graph Database (KGDB) forms an intricate network of nodes and relationships, shedding light on the connections between objects, dealers, collectors, and auction houses. This analysis offers unprecedented insights into the antiquities market's dynamics, highlighting the potential pathways of illicit activities and providing valuable information for market regulators and cultural heritage professionals. AIKoGAM shows the power of combining AI and network science in cultural heritage studies, paving the way for more informed policy-making and ethical practices in the antiquities trade.

Keywords: Archaeology · Illicit trade · Artificial Intelligence

1 Introduction: Understanding the Antiquities Market Through AIKoGAMA Subsection Sample

The transnational market for antiquities, a complex arena of both legitimate and illegitimate activities, is a domain where ambiguity and legality often intersect. This intricate market is influenced by a spectrum of entities with diverse roles that blur the lines between legal and illegal operations. Legislative ambiguities in various countries create environments conducive to transactions that, while criminalized in some parts of the world, are permissible in others [1–3].

The authenticity and legality of antiquities pose dual challenges, with artifacts ranging from fully legitimate to entirely illegitimate. The market is subject to an "Akerlofian lemons problem", where issues of legality compound concerns regarding authenticity. This complexity necessitates a comprehensive approach, one that can unravel the market's intricacies and provide insights into its dynamics.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 346–350, 2024.
https://doi.org/10.1007/978-3-031-76428-8_64

The AIKoGAM (AI-driven Knowledge Graph of the Antiquities Market) system is introduced as a tool designed to navigate this complex field. Developed through the integration of advanced computational methods and domain-specific knowledge, AIKoGAM offers a novel approach to understanding and potentially disrupting illicit networks within the antiquities trade.

2 Methodology

As highlighted by Fig. 1 depicting the architecture of AIKoGAM, the system was subdivided in four principal modules.

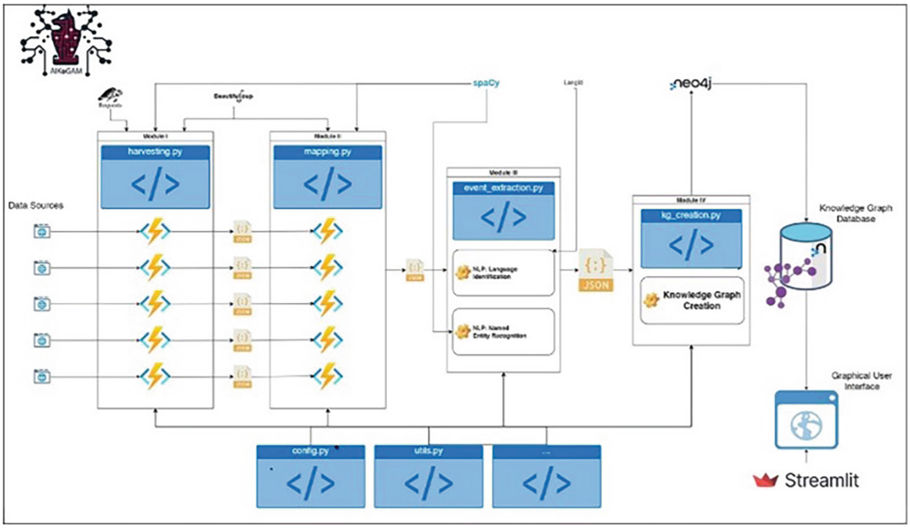


Fig. 1. AIKoGAM Architecture Diagram.

In Module I, Information Retrieval, AIKoGAM employs advanced data scraping techniques to navigate and extract information from digital platforms of auction houses and galleries. This module is characterized by its capability to handle diverse website architectures and data formats. The module employs a combination of HTML parsing, regular expressions, and conditional logic to isolate and retrieve relevant data, such as provenance information, auction results, and item descriptions. This data forms the foundation for AIKoGAM's analysis, underscoring the importance of accurate and comprehensive data collection.

Module II, Data Mapping and Transformation, focuses on the normalization and standardization of the collected data. Given the diverse origins of the data, this module implements a series of transformation processes to align the dataset into a uniform and coherent structure. This includes the normalization of date formats, currency conversions, and the standardization of terminologies. The key challenge here is to maintain the integrity and context of the original data while ensuring it fits into a standardized schema.

The architecture of Module III, Event Extraction, integrates Natural Language Processing (NLP) and Machine Learning (ML) to process and categorize the provenance texts. The primary objective is to dissect these texts into identifiable entities, such as dates, names, and locations. This module uses entity recognition algorithms to classify these elements [4], addressing the challenge of extracting meaningful information from often unstructured and linguistically diverse provenance narratives. Additionally, language detection algorithms [5] are employed to handle multilingual texts, ensuring accurate and contextually relevant entity extraction.

Module IV, Knowledge Graph Construction, is centered around the construction of a Knowledge Graph (KG) in [6] This involves the transformation of structured data from Module III into a graph format, where data points are represented as nodes and relationships. The challenge in this module is to accurately represent the complex relationships inherent in the provenance data, such as ownership transfers and geographical movements of artifacts. Neo4j's graph database capabilities are leveraged to create a dynamic and interactive representation of these relationships.

3 Results

The results gained from testing the system against three auction houses' websites and two antiquities galleries unveil the potential of the system in understanding the multilayered nature of the antiquities market through the lens of network analysis. The construction of the Knowledge Graph Database (KGDB) was instrumental in achieving this (see Fig. 2). The KGDB, a culmination of meticulous data aggregation from diverse sources, comprised an intricate network of 250,000 nodes and 300,000 relationships, representing a significant advancement in mapping the relationships between objects and actors, including dealers, collectors, and auction houses.

A critical aspect was the handling of challenges in the Knowledge Graph's accuracy and consistency. Entity recognition accuracy, a cornerstone of the project, was first confronted. The SpaCy NER models, while robust, showed limitations, with an accuracy rate of about 96.49%. Misclassifications were common, particularly in differentiating between organizations, people, or geographical locations. To counter this, we developed a custom mapping dictionary, which played a pivotal role in refining entity identification. This innovative approach significantly mitigated the NER models' shortcomings, enhancing the accuracy and reliability of the dataset.

Concurrently, during the testing phase the challenge of duplicate entries, a consequence of inconsistencies in naming conventions across different languages and sources, was addressed. Our approach, blending automated similarity detection with manual validation, was effective in identifying and merging 1,959 duplicate nodes. This process was vital in streamlining the KGDB, significantly reducing the total number of actor nodes and enhancing the integrity of the graph structure.

Transitioning from data collection to analysis, the AIKoGAM Knowledge Graph emerged as a powerful tool. Transforming raw data into a structured KGDB, the project facilitated a detailed exploration of the antiquities market's dynamics. This functionality was pivotal in exploring the ego networks [7] of objects and actors, revealing complex connections and transaction patterns.

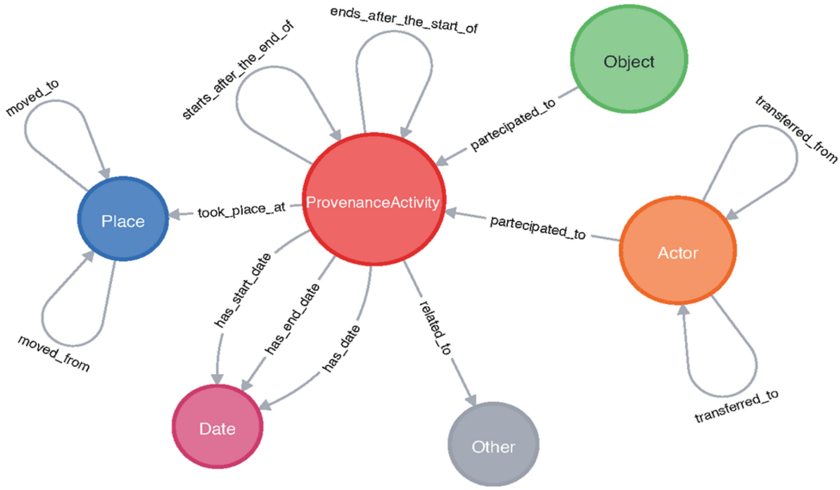


Fig. 2. Schema of the KG

The bipartite Actors-Objects projection illuminated the market's interconnected nature. Analysis of this projection revealed a network characterized by moderate connectedness and decentralization, suggesting a market landscape where actors are not dominated by a small number of entities. This projection was particularly insightful for studying smaller subgraphs and ego networks, highlighting the roles and influence of key actors.

Further, the monopartite Actors-Actors projections offered a deeper dive into the market. The projection based on mutual objects suggested a moderately interconnected market with a diverse distribution of influence, as indicated by the high modularity score. This pointed to the presence of specialized trading groups and a fragmented market structure.

In contrast, the transaction-based projection revealed a different facet of the market, characterized by selective engagement and expansive reach. This network's small-world nature [8], with a high level of modularity, suggested specialized sub-networks, likely delineated by artifact types or geographical regions. The interconnectedness of the market was evident, yet the presence of well-defined communities underlined the market's fragmented structure.

The insights gleaned from these projections are invaluable for stakeholders in the antiquities market, from cultural heritage professionals to market regulators. The analysis has uncovered key transactional nodes and potential pathways indicative of illicit activity, highlighting the complexities and nuances of the antiquities market. The results underscore the importance of network science in unraveling the intricate dynamics of cultural artifact trade and point towards the need for robust policy-making and ethical trade practices in the field.

4 Conclusions

AIKoGAM is conceptualized as more than a mere data aggregation tool. It stands as a dynamic platform offering insights into the market's intricate web of transactions, relationships, and entities. The development of AIKoGAM involved comprehensive data collection, normalization, and the establishment of an extensive Knowledge Graph (KG), utilizing sophisticated analytical techniques to decode complex relationships and transactions within the antiquities market.

AIKoGAM's key contributions lie in its ability to trace artifact provenance from open-source data and to delineate networks of crucial market actors. Its network analysis capability has proven effective in identifying potential illicit transactions, confirming its findings through pre-existing investigations and unveiling previously unnoticed connections. This has been pivotal in understanding the market's complexities, particularly in distinguishing legitimate from illicit activities.

Despite its advancements, AIKoGAM encounters challenges, notably the lack of standardized data formats in public antiquities market data and the limitations of general-purpose Named Entity Recognition models. These highlight areas for ongoing improvement and development.

The future of AIKoGAM involves enhancing its capabilities, particularly in predictive Deep Learning techniques, and refining the system's user interface. Collaborations with international cultural and law enforcement agencies are vital for integrating a broader range of data sources, essential for a deeper understanding of the market's dynamics.

References

1. Bichler, G., Bush, S., Malm, A.: Bad actors and faulty props: unlocking legal and illicit art trade. Global Crime **14**(4), 359–385 (2013)
2. Passas, N.: Globalization, criminogenic asymmetries and economic crime. Eur. J. Law Reform **1**(4), 399–423 (1999)
3. Tijhuis, A.J.G.: Transnational crime and the interface between legal and illegal actors: The case of the illicit art and antiquities trade. Universitet Leiden
4. Montani, I., et al.: Explosion/spacy: V3.0.0rc: transformer-based pipelines, new training system, project templates, custom models, improved component api, type hints & lots more (Version v3.0.0rc1) (2020)
5. Heafield, K., Kshirsagar, R., Barona, S: Language identification and modeling in specialized hardware. In: Proceedings of the 53rd Annual Meeting of the Association for Compling and 7th International Joint Conference on Natural Language Process. Association for Computer Linguistic (2015)
6. Neo4j: Neo4j desktop (2023). https://neo4j.com/docs/. Retrieved 19 June 2023
7. Perry, B.L., Pescosolido, B.A., Borgatti, S.P.: Egocentric Network Analysis: Foundations, Methods, and Models. Cambridge University Press (2018). https://doi.org/10.1017/9781316443255
8. Barabási, A.-L., Pósfai, M.: Network Science. Cambridge University Press (2016)

A Multilevel Approach to Monitor the Archaeological Park of Pompeii

Gabriel Zuchtriegel, Alessandra Zambrano(✉), and Vincenzo Calvanese

Ministero della cultura, Parco archeologico di Pompei, 80045 Pompei, Italy
alessandra.zambrano@cultura.gov.it

Abstract. Monitoring activities are indispensable in archaeological sites to implement a strategic plan for maintenance. In Pompeii this need is very challenging because an effective monitoring and inspection system requires that it be implemented on more than 1,221 buildings, in 10,400 rooms, many of which are plastered, frescoed, and without roofs, and across 8,690 m^2 of streets and squares of the ancient city. The actual monitoring activities in Pompeii is part of a strategy that allows us to evaluate the extent and severity of damage, material decay and anomalies identified on the structures and green areas, and moreover to assess the stability of the sloped edge of excavation. This information provides support for the decision-making process necessary for the maintenance program. The complexity of the cultural heritage in Pompeii has led to the development of a multileveled approach and that means that the monitoring activities are organised on multiple operational levels, characterised by different levels of detail corresponding to different scales of investigation and from different views of the town. The multilevel approach integrates the results of satellite interferometry, for monitoring deformations and slow movements of the ground and structures; the results of images recorded by drones; the results of inspection campaigns carried out by a multidisciplinary team of experts; the results of inspection campaigns obtained by robots; and Structural Health Monitoring (SHM) with devices implemented in specific critical cases. This approach allows the management of human and instrumental resources, aiming at the optimisation of the budget and time required in order to obtain an efficient, resilient and accurate procedure that yields an effective maintenance program.

Keywords: Monitoring approach · Structural Health Monitoring · Cultural Heritage

1 Decision Making Process

1.1 Multilevel Monitoring Approach

Archaeological structures are irremediably subject to decay for several reasons related to their intrinsic vulnerability and the threats posed by their environments, as generally, they are ruined structures in an advanced state of decay and directly exposed to atmospheric agents [1, 2]. In order to avoid the loss of archaeological artefacts and to extend their

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 351–356, 2024.
https://doi.org/10.1007/978-3-031-76428-8_65

life with a continuous and sustainable maintenance, the process of monitoring and the evaluation of the collected data are the fundamental steps in a strategy that allows the conservation of the archaeological sites [3]. An effective monitoring and inspection system is urgently needed to maintain the archaeological site of Pompeii, which has more than 10 thousand rooms, many of which are decorated with wall plaster and frescoes, as well as fragile mosaics on the floor. Furthermore, a large number of them are without roofs.

The knowledge of the evolution decay of the cultural heritage, possibly in real time, is essential to plan every activity.

Based on a substantial experimental phase developed in recent years at the Archaeological Park of Pompeii, an effective strategy has been implemented to support the decision-making process. The procedure is aimed at combining different types of surveys at different scales in order to plan and design the maintenance works on the structures, frescoes, mosaics, and furniture of the ancient city of Pompeii.

The monitoring project, launched by the Archaeological Park of Pompeii in 2021, has aimed to develop a strategy (Fig. 1) that can guarantee an effective process that accounts for the heavy-duty need of a reliable awareness of the state of conservation of structures and the effective resources (human and budget) of the Park.

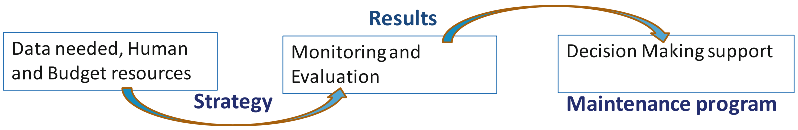


Fig. 1. Maintenance strategy.

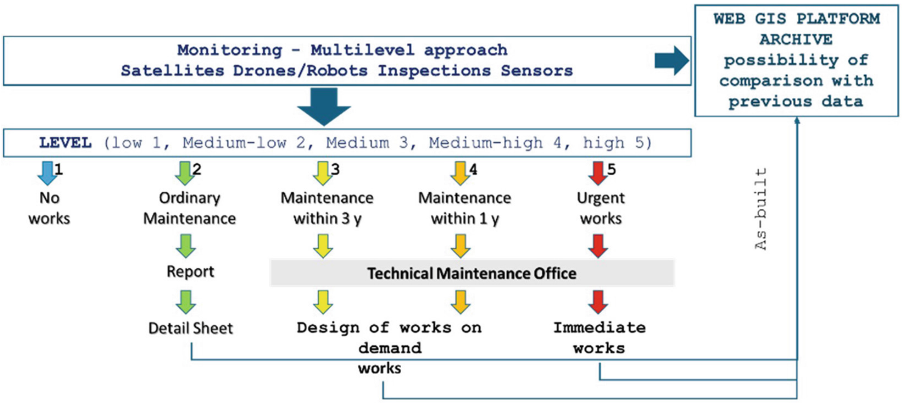


Fig. 2. Decision making process.

The decision-making process (Fig. 2) necessary for programming an efficient maintenance can be divided into three steps:

1. Monitoring by adopting a multilevel approach, characterised by different levels of detail corresponding to different scales of investigation: from the broad territorial scale to detailed focus on a single area of damage to an ancient wall or fresco.

   The multilevel monitoring approach is performed using satellite images, photos recorded with drones, inspections made by multidisciplinary teams comprised of experts, and finally, instrumental monitoring with sensors installed to observe a specific critical decay or deformation.

2. Evaluation of the data collected in order to identify the damage/decay and to assess its extension and its level of severity. For this step a scale of severity is defined in order to classify the registered anomaly.
3. Decision-making process based on the scale of severity of the damage/anomaly/decay that then informs the type and schedule of intervention required for each individual case. For example, routine maintenance, supplementary maintenance in 3 years or 1 year, or immediate works to ensure safety.

The large-scale monitoring activity is obtained by interpreting the satellite images, acquired by Cosmo-SkyMed satellite [4, 5]. Satellite radar interferometry enables the measurement of deformations and subtle movements of the ground, buildings, and infrastructure, due to subsidence and landslides, volcanic and seismic phenomena, and movements of conspicuous points or structures.

Reducing the scale, the evaluation of the images recorded by drones enables the monitoring of the roofs, tops of the walls, streets, green areas etc. The area of the entire Park is now monitored on a monthly basis with the aid of the fixed-wing drones. The flights, which can also be conducted immediately following extreme events (strong storms, etc.), provide a photographic survey of Pompeii as recorded from 80 m above the city. The orthophotos are elaborated and georeferenced. The resolution and accuracy of the orthophotos is particularly high and by superimposing images obtained from different flights, the evolution of the potential decay can be assessed. The recognition process is currently carried out by experts and the results are stored on the Archaeological Park of Pompeii's GIS Information System. The drone survey enables the identification of critical situations where a punctual inspection is needed.

A more detailed analysis can be performed on walls, frescoes, horizontal surfaces (floors and roofs), furnishings, and isolated structures by local inspection and compilation of a specific form containing options to record all possible types of damage/anomaly/degradation. The form has been installed on an informatics platform that enables the collection, storage and analysis of monitoring data recorded directly in the field. The platform consists of a web app developed with VueJS framework, a back-end application developed with Laravel, and MySQL database, an apk application for Android systems, that can be used in the field in an offline mode.

In particular, the aforementioned survey allows the description of the evolution of the main decay phenomena through an appropriate codification in terms of extension and intensity. In this way, taking into account the metric extension of the generic element investigated, it is also possible to estimate directly the costs of a routine maintenance process. The intention is to build a predictive model of decay rates, on the basis of which it will be possible to plan maintenance activities with a proactive approach.

The standardisation of the survey forms makes it possible to classify the assets according to absolute or relative degradation indices. In this way, the platform represents graphically (Fig. 3) in an immediate manner, the state of conservation of the specific element under analysis on the WEB GIS environment according to a defined deterioration scale with five levels: low, medium-low, medium, medium-high, high.

The forms have been designed to be filled out in the field using an IT platform accessible using tablets. The user, having identified the element through the topographical reference or the automatic geopositioning on the map, can fill in the form, describing any degradation/damage present with the pre-formatted notation within the form, and can attach one or more photos.

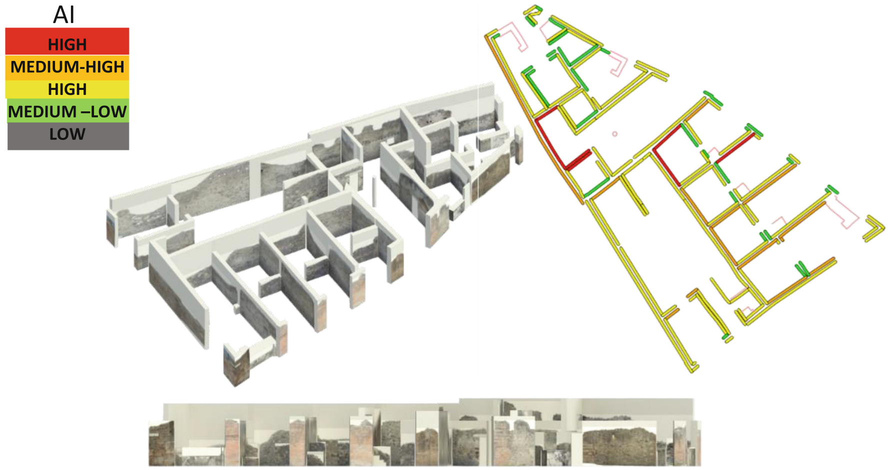


Fig. 3. Results of the inspection: vulnerable elements marked in red.

The platform enables the viewing of any previous monitoring form and the orthophotos taken monthly by drones for a quick and useful comparison.

These detailed inspections draw attention to specific cases where it is necessary to implement instrumental monitoring. In particular, when real-time knowledge of the evolution of particular conditions of degradation or potential critical issues is deemed necessary, the installation of sensors is envisaged. For example, the evolution of a pattern of cracks, a warning of a potential detachment of plaster, i.e. the triggering of a process of damage to an element, could be monitored by means of crack meters, high resolution multispectral cameras, laser-clinometric cameras, etc.

In this case, where the instrumental monitoring is needed, the data collected with the devices are processed and followed in real-time on a central panel located in the Archaeological Park of Pompeii. In such cases it is also possible to set alert thresholds for reporting conditions or for immediate action.

1.2 From Inspections to the Robotic Survey

The aim of this research, conducted in collaboration with the Istituto Italiano di Tecnologia, is to carry out inspections in an autonomous, reliable, repeatable, and effective manner. For this reason, the use of a robot for detecting anomalies on the structural features has been tested, instead of using human resources for inspections.

The anomalies are found by comparing images of successive robotic inspections, analysing them with an image-processing algorithm, and finally, thanks to an approach of deep learning, identifying the type of defect, damage or decay.

In the first inspection campaign in Pompeii the robot, called RINGHIO has monitored the structures along the roads to investigate its adaptability to the Pompeian paving. Then, during its second inspection, the rover has monitored an entire quarter recording the images of all the walls and frescoes.

In the coming months, with a Deep Learning Approach, the machine will be trained to identify the major defects present in the structures and recognize the defect/criticality or anomaly on the wall surface or fresco as classified in the inspection form. This approach will require a large number of images to teach the robot how to recognise and evaluate the anomaly, i.e. a difference found between consecutive recorded images. The Park is involved in this knowledge transfer to the machine with the objective of an automatic monitoring of the damage evolution and its identification.

2 Conclusions

Given the large expanse of the Archaeological Park of Pompeii, the proposed monitoring process has been implemented with the aim of reaching a greater degree of automation, adaptability, accuracy and efficiency, as well as increasing the reliability of results considering the importance of cultural heritage conservation at stake.

A critical aspect is related to the necessity of an expert judgement of the identified anomalies. The evaluation of the defect/damage/anomaly is a new and big challenge. To solve this critical issue the research will focus on the possibility to add a sustainable system of sensors that refine the in-depth analysis of anomalies and assess its level of danger.

References

1. Gasparoli, P.: Prevenzione e manutenzione nelle aree archeologiche, "LANX" 19, pp. 168–188 (2014). (In Italian)
2. Monitoraggio e Manutenzione dei Parchi Archeologici Cambiamenti climatici, dissesto idrogeologico, degrado chimico ambientale, edited by Russo A., Della Giovampaola I. L'Erma di Bretschneider (2020). ISBN: 9788891319470 (In Italian)
3. Rinaudo, F., Chiabrando, F., Lingua, A., Spanò, A.: Archaeological site monitoring: UAV Photogrammetry can be an answer, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B5, 2012 XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia

4. Carlucc,i R., Di Iorio, A.: Monitoraggio di siti archeologici da satellite un metodo ibrido tramite remote sensing, techniche GIS e un algoritmo di rilevamento della forma utilizzabile su imagine SAR. V. 1 N. 4 (2010): Archeomatica 4-2010. (In Italian)
5. Chen, F., Lasaponara, R., Masini, N.: An overview of satellite synthetic aperture radar remote sensing in archaeology: from site detection to monitoring. J. Cultural Heritage **23**, 5–11 (2017). https://doi.org/10.1016/j.culher.2015.05.003

Swarm Robotics and Archaeology: A Concepts Paper

Tuna Kalaycı[1](✉), Ali Emre Turgut[2], Scott Branting[3], Uluç Saranlı[4], and Mine Cüneyitoğlu[5]

[1] Faculty of Archaeology, Leiden University, Leiden, The Netherlands
t.kalayci@arch.leidenuniv.nl
[2] Department of Mechanical Engineering, Middle East Technical University, Ankara, Turkey
[3] Department of Anthropology, University of Central Florida, Florida, USA
[4] Department of Computer Engineering, Middle East Technical University, Ankara, Turkey
[5] Teknolus, Ankara, Turkey

Abstract. This concepts paper advocates for integrating swarm robotics into archaeological surveying practices. Conventional archaeological surveys, conducted on foot, often entail labour-intensive, time-consuming, and potentially hazardous tasks in vast and challenging terrains. Drawing inspiration from successful robotic missions, a swarm can assist the archaeologist in the field. The Archēbot project's primary objective centres on creating a flexible system that efficiently pinpoints areas of archaeological potential within expansive landscapes. Ultimately, the proposed hexapod robot by Teknolus emerges as a promising platform, equipped to navigate challenging terrains and withstand environmental adversities.

Keywords: archaeological survey · swarm robotics · autonomous documentation

1 Introduction

A fundamental task of archaeology is to find, identify, and record near- or above-surface past remains. Field archaeologists conduct surveys in large, heterogeneous terrains that are often difficult to traverse. Therefore, mapping features can be labour-intensive, time-consuming, and sometimes dangerous. While technologies, such as aerial imaging, offer support for how archaeologists survey, the abovementioned issues remain. Since the onset of scientific archaeology, teams have been scanning the landscapes opportunistically or in systematic transects, but almost always on foot. Significant technological progress has yet to address this most basic form of archaeological practice.

Robotic missions have become a reality, especially in exploring uncharted lands using autonomous units. One of the most well-known examples is NASA's Perseverance Rover. Yet, the current status of the Chinese Zhurong Rover is also telling: the rover was put in hibernation to protect against the Martian winter. However, researchers cannot "wake" the rover [1]-suggesting resourceful but single rover systems are too complex, and missions are prone to unexpected failures. A different approach is possible; if they

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 357–361, 2024.
https://doi.org/10.1007/978-3-031-76428-8_66

cooperate, simple and much cheaper robotic units can still accomplish complex tasks. Swarm Robotics tackles these questions [2].

The collaboration between robotics and archaeology has immense potential. Such partnerships started in the early 70s [3]. This was followed by pioneering work in the 90s, including the Upuaut [4] and Djedi [5] Projects. Current advancements in computing technologies, especially artificial intelligence (AI), have recently changed how one can use machinery in research. In contemporary field practice, the main direction is using remotely piloted aerial systems (RPAS) or drones [e.g., 6, 7]. However, there are several issues. First, they are not best suited to wide-area exploration due to short flight times and restrictions in adverse weather conditions. Second, these systems have weight constraints, reducing operational capacities. Third, they can't interact directly with objects on the surface or partially buried in the ground, limiting their ability to identify what those objects are. Therefore, ground swarm units appear to be the ideal system to be deployed as the archaeologist's assistant.

2 Rationale and Objectives

A swarm robotics project for archaeology should aim for a scalable field system that effectively identifies areas with archaeological potential within large landscapes. This follows a two-phase plan for robotic exploration of unknown terrain. The first phase is to identify an area of interest autonomously. The second is to make sensor measurements within that area to achieve a particular task. If the terrain is enormous and contains many areas of interest, single rover systems (e.g., Perseverance Rover) are not feasible since the machine has to perform tasks consequentially.

Furthermore, an area of interest might not be that interesting after all and the machine would spend considerable time and battery power. Alternatively, swarm units can collaboratively generate a 'potential field' [8] (or an archaeological 'area of interest') and perform mapping in separate groups. Archaeology can significantly benefit from this type of flexibility and scalability. In return, archaeology can offer the perfect analogue mission; roboticists can test their autonomous designs before sending them to areas with limited intervention capacity, such as Mars.

3 An Innovative Platform

The use of swarm robotics in archaeology has implications, from conceptualizing new types of low-cost archaeology sensors for these small and "expendable" robotics platforms, as opposed to high-end research and development kits on the market, to shifts in archaeologists' labour. Using a robot in archaeology is not entirely ground-breaking, but approaching robotics research in archaeology from a landscape perspective is a game changer.

Archēbot project deploys a hexapod robot built by Teknolus [9]. The platform has its origins in RHex [10]. It is designed to navigate challenging terrains, particularly in areas where either human life is threatened, or human fatigue becomes a limiting factor.

The platform was designed with a generic framework to ensure versatility. Its software and hardware architecture are strategically tailored for seamless integration of additional sensors, actuators, and payloads. In 2022, it secured the Turkish research grant under the project 'Autonomous Advancement in Constricted Environments with Legged Robotics' (Project # 3220580). The project primarily focuses on developing semi-autonomous walking and discovery capabilities with minimal user input. In case of poor reception, the robot continues autonomous exploration of the area. Navigating through uneven terrains and finding ways through narrow passages is a key aspect.

While achieving precise mapping might pose a challenge and very high processing power, the platform maximises the utilisation of surrounding data to the best of its ability with readily available sensors. The variety of sensors collecting environmental data makes the swarm framework even more applicable. Considerable effort has been devoted to developing its networking structure, API, and user interface, providing numerous options for remote commanding. The system operates on Linux and is compatible with ROS.

The platform demonstrates substantial progress on slopes exceeding a 60% side slope (Fig. 1a). If it accidentally flips, it can operate normally, even upside-down. Apart from its standard walking capability, it offers specialised tuneable modes, including a climbing mode. With its specially designed tuneable stair mode, the platform is equipped to handle various types of stair steps, even if they are broken, missing, or uneven (Fig. 1b). The leg morphology of the platform enables progress on thick grass and bushes gravel and small to large rocks, handling an average of up to 30 cm (Fig. 2).

In addition to its capability to traverse rough terrains, the legged robot is also designed to withstand various environmental challenges. It exhibits durability against heavy impact (Fig. 3), dust and sand while being water-resistant and operable on muddy surfaces, overcoming puddles of up to 10 cm. Moreover, it demonstrates resilience to extreme temperatures and prolonged exposure to heavy sunshine. This final property is of utmost importance in the field of archaeology.

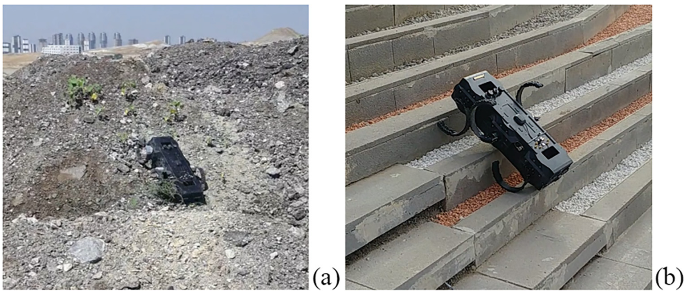


Fig. 1. a) In this rubble pile, the slope averages 75–80%. Slippage may occur, resulting in occasional robot falls or flips. Nevertheless, the climbing mode is capable of gradually navigating such steep inclines. b) A sample of uneven and irregular steps.

(a) (b) (c)

Fig. 2. The morphology of the platform enables progress on thick grass and bushes (from left to right: inside thick grass; getting out and traversing; gravel and small to large rocks).

Fig. 3. The robot faces tough, random impacts when traversing through gravel and rocks.

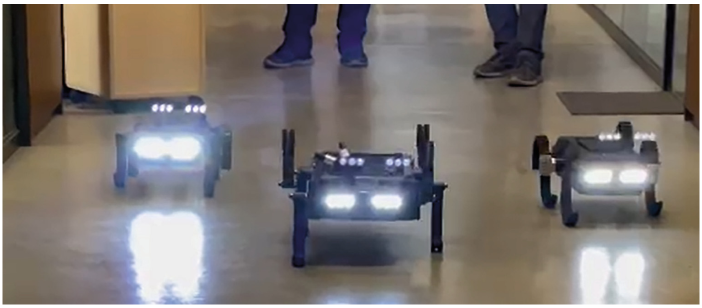


Fig. 4. The ROS compatible platform offers substantial potential for swarm applications.

The ongoing efforts of researchers in enhancing its environmental awareness speak to the platform's commitment to continuous improvement and adaptability in various settings. Its infrastructure facilitates the construction of robot swarms and enables seamless integration with a range of sophisticated sensors and systems (Fig. 4). The platform exhibits a strong potential for diverse applications, including those in demanding terrains.

4 Conclusion

Swarm robotics can alter archaeological practice and can open new sub-domains of robotics research in archaeology, such as picking relevant material (e.g., seeds, bones, etc.) with robotic "fingers" (e.g., [11]) or mapping inaccessible cave shafts and chambers (e.g., [12]). It also offers the robotics field the possibilities of cross-hybridization with "historical sciences", a combination that can generate unexpected research paths, as seen in the late 1940s and early 1950s development of carbon-14 dating [13].

The Archēbot project can also offer a unique and fruitful combination of archaeology and robotic research. Archaeology can learn from state-of-the-art engineering practices, and robotics research can gain an unexpected disciplinary ally through which autonomous systems can be put into meaningful real-life challenges. Finally, the contributions of this fruitful collaboration hold promise to advance practices in other fields and disciplines, such as agricultural sciences [14] or planetary research [15].

References

1. Mallapaty, S.: What's happened to China's first Mars rover? Nature-News (2023). https://www.nature.com/articles/d41586-023-00111-3
2. Brambilla, M., Ferrante, E., Birattari, M., Dorigo, M.: Swarm robotics: a review from the swarm engineering perspective. Swarm Intell. **7**, 1–41 (2013)
3. Pergamon Project: ArchaoMat, Super Mk II DAI, Istanbul (1973)
4. Bauval, R.G.: The Upuaut Project: New findings in the southern Shaft of the Queen's Chamber in Cheops Pyramid. Discuss. Egyptol. **27**, 5–7 (1993)
5. Richardson, R., et al.: The "Djedi" Robot Exploration of the Southern Shaft of the Queen's Chamber in the Great Pyramid of Giza, Egypt. J. Field Robot. **30**(3), 323–348 (2013)
6. Orengo, H.A., Garcia-Molsosa, A.: A brave new world for archaeological survey: automated machine learning-based potsherd detection using high-resolution drone imagery. J. Archaeol. Sci. **112**, 105013 (2019)
7. Agapiou, A., Vionis, A., Papantoniou, G.: Detection of archaeological surface ceramics using deep learning image-based methods and very high-resolution UAV imageries. Land **10**(12), 1365 (2021)
8. Howard, A., Matarić, M.J., Sukhatme, G.S.: Mobile sensor network deployment using potential fields: a distributed, scalable solution to the area coverage problem. Distrib. Auton. Robot. Syst. **5**, 299–308 (2002)
9. Teknolus Homepage: http://teknolus.com/en. Last accessed 15 Jan 2024
10. Saranlı, U., Buehler, M., Koditschek, D.E.: RHex: a simple and highly mobile hexapod robot. Int. J. Robot. Res. **20**(7), 616–631 (2001)
11. Alizadehyazdi, V., Bonthron, M., Spenko, M.: An electrostatic/gecko-inspired adhesives soft robotic gripper. IEEE Robot. Autom. Lett. **5**(3), 4679–4686 (2020)
12. Li, X.A., Yue, H., Yang, D., Sun, K., Liu, H.: A large-scale inflatable robotic arm toward inspecting sensitive environments: design and performance evaluation. IEEE Trans. Ind. Electron. **70**(12), 12486–12499 (2023)
13. Kern, E.M.: Archaeology enters the 'atomic age': a short history of radiocarbon, 1946–1960. Br. J. Hist. Sci. **53**(2), 207–227 (2020)
14. Berger, G.S., et al.: Cooperative heterogeneous robots for autonomous insects trap monitoring system in a precision agriculture scenario. Agriculture **13**(2), 239 (2023)
15. Li, C., Lewis, K.: The Need for and Feasibility of Alternative Ground Robots to Traverse Sandy and Rocky Extraterrestrial Terrain. *Advanced Intelligent Systems*, 2100195 (2022)

Robotics in Archaeology: Navigating Challenges and Charting Future Courses

Arianna Traviglia(✉) and Riccardo Giovanelli

Istituto Italiano di Tecnologia, Center for Cultural Heritage Technology, 30172 Venice, Italy
{arianna.traviglia,riccardo.giovanelli}@iit.it

Abstract. This position paper explores the current state of robotics in archaeology, emphasizing the need for dedicated solutions to address the repetitive, time-consuming, and physically demanding nature of archaeological activities. Despite the advancements in various industries, robotics tailored specifically for archaeology remains underdeveloped. The paper reviews the existing applications of robotics in archaeology and discusses promising projects, shedding light on how these initiatives contribute to the field. It explores future prospects for robotics in archaeology, anticipating the inevitable development of functionalities for on-site activities, and underscoring the need for specialized discussions between roboticists and archaeologists.

Keywords: robotics · automation · archaeology · cultural heritage

1 Introduction

Archaeology, with its repetitive, time-consuming, and physically demanding tasks, stands at the cusp of a transformative era with the integration of robotics. While other industries have made significant strides, archaeology is yet to witness a comprehensive suite of robotic solutions designed specifically for its unique challenges. Despite the repetitiveness of many archaeological activities, robotics has yet to propose tailored solutions, avoiding mere adaptations from other domains. We assert that archaeology could play a pivotal role in steering the future development of robotics, offering bespoke solutions for archaeologists' unique needs. This paper explores the untapped potential of robotics to revolutionize archaeological methodologies, alleviating the burdens of fieldwork and addressing labor shortages.

2 State of the Art

The initial integration of robotics in archaeology saw the widespread use of drones, showcasing adaptability in various applications, such as remote surveying, detection, monitoring, as well as 3D reconstruction. Autonomous and

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 362–367, 2024.
https://doi.org/10.1007/978-3-031-76428-8_67

non-autonomous robots explored hazardous environments, facilitating remote investigations of inaccessible sites, such as caves, catacombs [9] and underwater environments. Underwater remotely operated vehicles (ROVs) proved crucial for underwater archaeology, with robots equipped not only with sensors able to capture images and bathymetric data, but also with manipulators, excavation tools and thrusters to carry on direct excavation [3].

Recent developments signal a shift towards robotics explicitly designed for archaeological applications. The Fraunhofer initiative for heritage digitalization and the Italian Institute of Technology (IIT) ongoing efforts on the same trajectory epitomize this shift. Fraunhofer has been working on automation of heritage digitalization since 2014, with the launch of their Competence Center Cultural Heritage Digitization and the development of CultLab3D [8], a versatile, adaptable scanning setup that incorporates autonomous and adaptable robots, alongside advanced optical scanning methods. The system includes two scanning modules that are interconnected via a tray conveyor mechanism. CultArm3D module is a color-calibrated, autonomous scanner, employing photogrammetry with high-resolution camera, diffuse lighting, a robotic arm, and an adaptable turntable. The CultArc3D module is an imaging-based scanner, consisting of a light arc with ring lights and a camera arc with ten cameras, each designed to encircle an artifact on a carrier tray from different radii for independent rotational movement.

The Center for Cultural Heritage Technology's (CCHT) in Venice, since 2019, has been developing technologies for scanning in 3D archaeological items based on accurate sensor activation, employing polarized lenses, structured light scanners, and hyperspectral cameras [5]. Collaborating with other units within the Institute, the division is swiftly emerging as a leading center for the application of robotics in the field of cultural heritage. Their involvement in initiatives like REPAIR (Reconstructing the Past: Artificial Intelligence and Robotics meet Cultural Heritage) in Pompeii, the Casa delle Tecnologie in Genova, and RINGHIO (Robot for Inspection and Navigation to Generate Heritage and Infrastructures Observations) illustrate the evolving landscape of robotics in archaeology.

The REPAIR project in Pompeii, for instance, focuses on the restoration of archaeological frescoes in fragmentary state using AI and advanced robotic technologies. The project involves three IIT research lines, (the SoftBots, the PAVIS, and the HHCM), and one centre, the CCHT, besides several European and non-European partners. REPAIR is developing an intelligent robotic system able to autonomously process, match and physically assemble fractured large artefacts. It combines Convolutional Neural Networks and Generative Adversarial methos for image segmentation tasks, puzzle solving and breaking curves algorithms for the reassembly of broken objects [1], within a robotic system which integrates advanced manipulation and perception capabilities, envisaging a dual-arm setup with anthropomorphic end-effectors for careful handling of fragile fresco pieces, able to gently grasping fragments [6], changing high-resolution cameras viewpoints for image acquisition and texture mapping. The vision system will both autonomously plan the grasping and proceed to the scanning of the fragments with both 2D and 3D perception sensors.

The RINGHIO (Robot for Inspection and Navigation to Generate Heritage and Infrastructures Observations) robot, see Fig. 1, aims not only to monitor Pompeii's tructures anomalies, but also to maximize task times, in order to have both an autonomous system and a faster than the current human based one. The acquisition system is represented by an Unmanned Ground Vehicle (UGV) [7] able to navigate on all-terrain vehicle environment and detect its features. The navigation is in charge of a fast wheeled platform, while the inspection is demanded to an HD camera and its balance system, which is deputy to align the cameras with the Pompeii' structures.

CCHT supports the team of Barbara Mazzolai (from the IIT Bioinspired Soft Robotics Laboratory) in delineating potential applications of bio-inspired robotics for monitoring the condition of walls that still bears frescoes in situ, as exemplified in locations like Pompeii. In this category, plantoids and GrowBots stand out for their potential in cultural heritage applications. Plantoids are robots that replicate the features of plant roots, drawing inspiration from the natural growth patterns of botanical root systems. The project in Pompeii involves a key collaboration with Simone Cinquemani from Politecnico of Milan. This partnership focuses on refining the mechanical design of the root-like robotic system, aiming to enhance its exploration abilities. GrowBots, known for their lightweight and compact design, have the ability to anchor themselves, manoeuvre through narrow spaces, and climb in areas where conventional climbing robots equipped with wheels, legs, or rails might get stuck, fall, or even cause environmental damage [2]. These innovative robotic systems offer transformative prospects in the monitoring and conservation of cultural heritage.

Finally, the Casa delle Tecnologie (CTE) of Genova (the Home of Technologies of Genova), a project that sees the collaboration of CCHT and the IIT Industrial Robotic Unit, aims to build a fully automated archaeological deposit utilizing cloud-based robotic scanning for incoming archaeological materials and a robotized shelving system for efficient storage management. The infrastructure is designed to digitalize artefacts held in storage facilities, facilitate remote and automated management and movement of these items, and allow virtual access to the public of usually unseen objects. Special emphasis is put on protection and on the adaptability to various cultural depository types and artifact categories. Central to the project is the innovative use of 5G technology, which enables rapid data transmission crucial for real-time remote control of robots and cloud-based data processing [4]. This enhances the efficiency and security of operations, allowing for centralized control and management of data and artifact movement. The integration of high-resolution 3D scanning on the robotic system and AI algorithms for the early detection of degradation marks another advancement, providing crucial information on the conservation status and restoration needs.

3 Prospects

Looking ahead, the inevitable development of additional functionalities for on-site activities opens new avenues for robotics in archaeology. Future robotic applications may include excavation tasks or enhanced support for reconnaissance efforts. The ongoing integration of robotics into archaeological practices presents innovative and efficient approaches to exploration and preservation.

Fig. 1. Picture of the RINGHIO robot, while transversing a ancient roman street within the Pompeii archaeological site.

The ongoing dialogue between roboticists and archaeologists, exemplified by the collaborative efforts of the CCHT, showcases the potential for transformative impact. It ensures that technological developments align with real needs and challenges of cultural heritage practice. The multidisciplinary environment opens the possibility to test and refine innovative ideas, leading to robotics solutions that are both practical and sensitive to cultural heritage field, where the requirements are distinctively intricate and specific. Unlike standard commercial applications where uniformity and scalability are often priorities, cultural heritage projects are characterized by a high degree of specificity. One of the critical aspects of applying robotics in cultural heritage is the need for gentle and minimally invasive interventions, due to the fragility and irreplaceability of artefacts and sites of historical importance, necessitating a level of softness that is not typically a concern in more commercial settings. Robotic solutions in this field need to interact with fragile materials and structures without causing damage, which is a significant departure from industrial robots that are built for strength and speed. Precision and careful handling requirements go beyond the capabilities of many standard robotic solutions: this might involve intricate manipulations, precise movements, and sensitive feedback systems that can adapt to the unstandardized and unpredictable nature of historical artefacts and environments.

Such complexity extends to the varied environmental settings in which these robots must operate. Unlike the controlled settings of industrial environments, cultural heritage sites can be diverse, ranging from underwater excavations to ancient structures with challenging terrains. This diversity requires highly adaptable and flexible robotic systems capable of working effectively in a wide range of conditions.

The application in the cultural heritage and archaeological domain requires therefore an amalgamation of specialization, robust yet flexible technology, and a gentle, non-invasive methodology applicable to real-world scenarios.

4 Conclusions

While the potential of robotics in archaeology is vast, the field is at a crucial juncture, demanding bespoke solutions. The adaptation of existing robotic technologies must evolve into the creation of specialized tools for archaeology, particularly for demanding tasks such as archaeological excavations. Collaborative efforts between roboticists and archaeologists will be pivotal in navigating the unique challenges posed by archaeological endeavors. As we stand on the cusp of this exciting intersection between technology and archaeology, preparedness for this impending collaboration is key to unlocking the full potential of robotics in archaeology. The ongoing dialogue between these two disciplines will be essential to overcoming challenges and realizing the transformative impact of robotics in the archaeological realm.

References

1. Alagrami, A., Palmieri, L., Aslan, S., Pelillo, M., Vascon, S.: Reassembling broken objects using breaking curves. In: CVPR 2023 Workshop on 3D Vision and Robotics (3DVR). Ca' Foscari University of Venice, Department of Computer Science (DAIS) (2023)
2. Dottore, E.D., Mondini, A., Mazzolai, B.: Support localization strategy for growing robots aided by light perception inspired by climbing plants. In: 2021 IEEE 4th International Conference on Soft Robotics (RoboSoft), pp. 105–110 (2021). https://doi.org/10.1109/RoboSoft51838.2021.9479220
3. Hotta, S., Mitsui, Y., Suka, M., et al.: Lightweight underwater robot developed for archaeological surveys and excavations. Robomech. J. **10**(2) (2023). https://doi.org/10.1186/s40648-023-00240-4
4. Liu, Y., et al.: Robotic communications for 5G and beyond: challenges and research opportunities. IEEE Commun. Mag. **59**(10), 92–98 (2021). https://doi.org/10.1109/MCOM.111.2001118
5. Marchello, G., Giovanelli, R., Fontana, E., Cannella, F., Traviglia, A.: Cultural heritage digital preservation through AI-driven robotics. Int. Archives Photogrammetry, Remote Sens. Spat. Inf. Sci. XLVIII-M-2-2023, 995–1000 (2023). https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-995-2023
6. Mohammadi, S.S., et al.: 3DSGrasp: 3D shape-completion for robotic grasp (2023)
7. Rodin, C., Andrade, F., Hovenburg, A., Johansen, T.: A survey of practical design considerations of optical imaging stabilization systems for small unmanned aerial systems. Sensors (Switzerland) **19**(21) (2019)

8. Santos, P., et al.: Cultlab3d - on the verge of 3d mass digitization. In: Eurographics Workshop on Graphics and Cultural Heritage. The Eurographics Association (2014). https://doi.org/10.2312/gch.20141305
9. Serafin, J., et al.: Robots for exploration, digital preservation and visualization of archaeological sites. In: Bordoni, L., Mele, F., Sorgente, A. (eds.) Artificial Intelligence for Cultural Heritage, pp. 121–140. Cambridge Scholars Publishing (2016)

Precision Robotics for Artifact Scanning: Leveraging a DeLaN-Based Control Approach for Archaeological Preservation

Marcel Lahoud[2,4](✉), Riccardo Giovanelli[1,3], Arianna Traviglia[1], and Gabriele Marchello[2]

[1] Istituto Italiano di Tecnologia, Center for Cultural Heritage Technology, 30172 Venice, Italy
{riccardo.giovanelli,arianna.traviglia}@iit.it, riccardo.giovanelli@unive.it

[2] Istituto Italiano di Tecnologia, Center for Convergent Technologies - Industrial Robotics Facility, 16163 Genoa, Italy
{marcel.lahoud,gabriele.marchello}@iit.it, marcel.lahoud@edu.unige.it

[3] Ca' Foscari University of Venice, DSU, 3246 Venice, Italy

[4] DIBRIS, University of Genoa, Genoa, Italy

Abstract. This study introduces an innovative method for accurately guiding robots to predefined acquisition positions, specifically tailored for scanning archaeological artifacts. Leveraging advanced DL-models and model-based controllers, our approach ensures precise and efficient robot movements. By integrating cutting-edge robotics, this method contributes to the optimization of artifact scanning processes. The result is an enhanced capability for detailed archaeological research and artifact preservation. This method showcases a promising avenue for the digitisation and documentation of historical artifacts, offering high levels of precision in acquiring valuable archaeological data.

1 Introduction

Digitisation of archaeological objects is a task that is getting more importance year by year thanks to the technological improvements and increased accessibility to technologies that produce satisfactory results [1]. Furthermore, utilising a robotic platform can enhance the acquisition of 3D data by reducing pose and motion errors associated with a human operator [2].

Development of highly accurate controlled robots requires estimation of its dynamical parameters. A widely used identification method, based in the Lagrangian formulation is carried-out using the Inverse Dynamic Identification Model (IDIM) [6]. However, modern solutions using Deep Learning approaches inspired in Lagrangian mechanics are a promising solution offering adaptability and improvement over time of the dynamic model of the robot [3,4].

This paper leverages a model identification method using a Deep Neural Network inspired in Lagrangian dynamics (DeLaN [4]). This method, together

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 368–372, 2024.
https://doi.org/10.1007/978-3-031-76428-8_68

with a model-based controller addresses the required trajectories to perform a Structure from Motion (SfM) task for digitising archaeological artifacts.

2 State of Art

Model-based controllers are crucial for accurate performance, relying on the equation of motion to mathematically describe the dynamic behavior of serial robots. Equation (1) captures the intricate relationships governing joint positions, velocities, and accelerations, forming a critical foundation for precise control and optimization in robotic systems. However, in order to efficiently control a robot we must identify its physical parameters.

$$\boldsymbol{\tau} = \boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{G}(\boldsymbol{q}) + \boldsymbol{\tau}_f \tag{1}$$

where $\boldsymbol{\tau}$, $\boldsymbol{M}(\boldsymbol{q})$, $\boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})$, $\boldsymbol{G}$ and $\boldsymbol{\tau}_f$ are the joint motor torques, the inertia matrix, the Coriolis matrix, the gravity torques and the friction torques, respectively. Conventional identification methods consist in finding a solution based in the linearity that exists between the dynamic parameters and the equation of the inverse dynamics resulting from the Lagrangian formalism. This methodology is well-known as the IDIM. Nonetheless, the solution of the IDIM in some configurations can lead to physically impossible parameters, *i.e.*, a not positive definite inertia matrix or negative friction coefficients. Consequently, authors had proposed solutions to this particular problem by including a Cholesky factorization [7], optimization on manifolds [8], linear matrix inequalities [9], among others.

Other identification methods consider black-box models, such as Artificial Neural Networks (ANN). However, these black-box solutions lack of physically comprehensible results and can be used only as feed-forward controllers. However, a recent solution considering Physically Inspired Neural Networks can provide physically consistent solutions. The DeLaN methodology has demonstrated its effectiveness across various serial robots, accounting for joint frictions [4] and providing comprehensive elements that describe the equation of motion. Moreover, DeLaN offers a door to perform traditional model-based controllers using a Deep Learning approach by allowing to estimate $\boldsymbol{M}(\boldsymbol{q})$, $\boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})$, $\boldsymbol{G}$ and $\boldsymbol{\tau}_f$.

3 Methodology

A robot cell composed of an anti-vibration table and a serial UR5e manipulator is considered for the archaeological artifact scanning as depicted in Fig. 1. This setup aims to minimise external disturbances, creating a controlled environment for the robotic arm to digitise archaeological relics. Using the known kinematics and workspace restrictions, we deploy several conventional Fourier series parameterised excitation trajectories [5]. In total, we generate 46 feasible excitation trajectories, each lasting 10 s and sampled at 500 Hz. These trajectories are executed on the robot using SDU robotics' ur_rtde libraries on an Ubuntu PC (Intel Core i7-11700K Processor) with a real-time kernel. Throughout the execution, joint position, velocity, and current were measured.

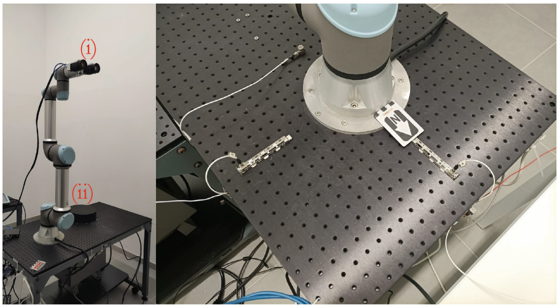


Fig. 1. To the left, the UR5e robotic system is employed for dynamic identification and 3D digitisation, incorporating both the scanner (i) and a revolving platform (ii) for digital reconstruction. On the right side, there is a close-up view illustrating how the robot is securely attached to the anti-vibration table.

3.1 Robot Dynamic Identification

By using the measured values, a DeLaN considering friction models is trained, as depicted in Fig. (2), in order to obtaining a model of the robot. The training of the DeLaN model is done in another Ubuntu PC (Intel Core i7-11700K Processor) with a RTX3090 GPU. Training is done considering the changes in validation loss and training loss in order to avoid over-fitting.

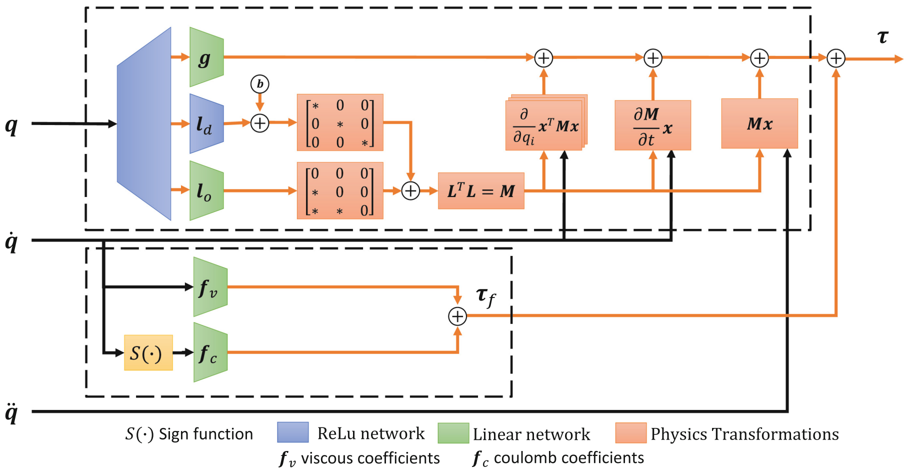


Fig. 2. Graphical representation of the DeLaN architecture used for dynamic identification of the UR5e robot.

Once our model is completely trained and validated we proceed with the application of the model-based controller for scanning archaeological artifacts.

3.2 Model-Based Controller

Model-based controllers are designed to precisely regulate the torques exerted on each robot joint, facilitating accurate execution of movements. Nevertheless, the UR5e robot employs a closed control architecture, thereby prohibiting the direct input of torque/current signals to the motors. Consequently, we propose the employment of a residual approach as the one used by Gaz *et al.* for human-robot interaction [10] depicted in Fig. (3). In this way the use of a residual approach where we input a joint torque ($\boldsymbol{\tau}_d$) based on a virtual spring-damper system from the desired trajectory and later generate a self-motion around a desired focal point based on an admittance law in Eq. (2)

$$\dot{\boldsymbol{q}}_c = \boldsymbol{J}_p^{\#}\boldsymbol{K_p}\boldsymbol{e} + (\boldsymbol{I} - \boldsymbol{J}_p^{\#}\boldsymbol{J}_p)\boldsymbol{K}_r\boldsymbol{r} \tag{2}$$

where $\boldsymbol{K_p}$ and $\boldsymbol{K}_r$ are diagonal gain matrices, $\boldsymbol{I}$ is a diagonal identity matrix, $\boldsymbol{J}_p$ is the linear Jacobian matrix $\boldsymbol{J}_p^{\#}$ is the Moore-Penrose pseudo-inverse of the linear Jacobian matrix, $\boldsymbol{e}$ is the error in the position and $\boldsymbol{r}$ is the residual. The residual is computed in Eq. (3) considering a diagonal gain matrix $\boldsymbol{G}$, the model of the robot, the desired torque ($\boldsymbol{\tau}_d$) and actual joint position, velocity and torque ($\boldsymbol{q}_v, \dot{\boldsymbol{q}}_v, \boldsymbol{\tau}_v$).

$$\boldsymbol{r} = \boldsymbol{G}\left(\hat{\boldsymbol{M}}(\boldsymbol{q_v})\dot{\boldsymbol{q_v}} - \int_{\boldsymbol{0}}^{\boldsymbol{t}} (\hat{\boldsymbol{C}}(\boldsymbol{q_v}, \dot{\boldsymbol{q}}_v) - \hat{\boldsymbol{G}}(\boldsymbol{q_v}) - \hat{\boldsymbol{\tau}}_f(\dot{\boldsymbol{q}}_v) + \boldsymbol{\tau_v} + \boldsymbol{\tau}_d + \boldsymbol{r})dt\right) \tag{3}$$

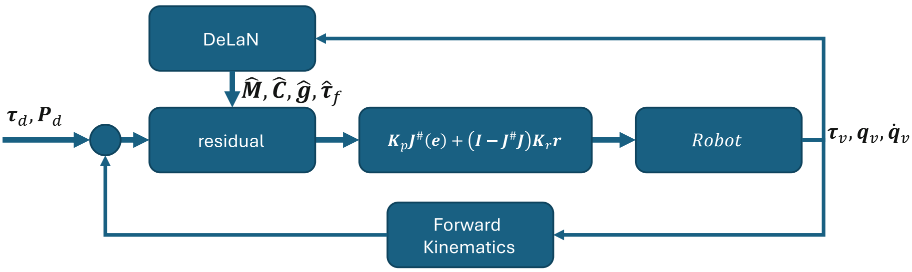


Fig. 3. Schematic representation of the control loop of the model-based approach using residuals.

Consequently, by applying the control law in Eq. (2) the residual $\boldsymbol{r}$ should provide motion in the null-space of the robot while keeping fixed the focal distance to the archaeological artifact being scanned.

4 Conclusion

A robotic arm offers a notable benefit with its capability to achieve precise movement at the millimeter level, addressing a key limitation of the SfM methodology. By adding the proposed model-based approach, this feature ensures highly accurate pose values, eliminates operator errors in the workflow, and effectively constrains the poses of all views. This, in turn, enhances the performance of the reconstruction algorithm.

Through the use of our method, the operator errors in moving the 3D scanner are removed. Hence, this leads to improved 3D data acquisition from reliable and accurate scanning trajectories. This holds particular significance in cultural heritage preservation, where accurate and dependable data is essential for making well-informed decisions regarding the conservation and protection of valuable artifacts.

Moreover, our methodology presents a viable model-based solution for controlling a robot based on Deep Lagrangian Networks applied to heritage preservation.

References

1. Ćosović, M., Maksimović, M.: Application of the digital twin concept in cultural heritage. In: Amelio, A., Montelpare, S., Ursino, D. (eds.) Proceedings of the 1 st International Virtual Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 12 September 2022 (2022)
2. Marchello, G., Giovanelli, R., Fontana, E., Cannella, F., Traviglia, A.: CULTURAL HERITAGE DIGITAL PRESERVATION THROUGH AI-DRIVEN ROBOTICS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2023)
3. Lutter, M., Ritter, C., Peters, J.: Deep Lagrangian networks: using physics as model prior for deep learning. In: 7th International Conference on Learning Representations, ICLR 2019 (2019)
4. Lutter, M., Peters, J.: Combining physics and deep learning to learn continuous-time dynamics models, ArXiv (2021)
5. Leboutet, Q., Roux, J., Janot, A., Rogelio, J., Cheng, G.: Inertial parameter identification in robotics: a survey. Appl. Sci. (2021)
6. Gautier, M., Briot, S.: New method for global identification of the joint drive gains of robots using a known payload mass (2011)
7. Gautier, M., Venture, G.: Identification of standard dynamic parameters of robots with positive definite inertia matrix (2013)
8. Traversaro, S., Brossette, S., Escande, A., Nori, F.: Identification of fully physical consistent inertial parameters using optimization on manifolds. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2016)
9. Wensing, P.M., Kim, S., Slotine, J.-J.E.: Linear matrix inequalities for physically consistent inertial parameter identification: a statistical perspective on the mass distribution. IEEE Robot. Autom. Lett. (2018)
10. Gaz, C., Magrini, E., De Luca, A.: A model-based residual approach for human-robot collaboration during manual polishing operations. Mechatronics (2018)

Robotic Infrastructures for Infrastructures Robotic Inspections

Mariapaola D'Imperio[1(✉)], Gabriele Marchello[2], and Ferdinando Cannella[2]

[1] SDA Engineering, Via Kennedy 3, 35013 Cittadella, Padova, Italy
mdimperio@sdaeng.it

[2] Italian Institute of Technology, Via Morego 30, Genova, Italy
{Gabriele.Marchello,Ferdinando.Cannella}@iit.it

Abstract. In the world, there are hundreds of automation devices and equipment on the bridges for automatic inspection and/or monitoring. These range from simple travellers to highly complex robots, all of which require infrastructure such as rails, cables, beams, and tracks to operate safely and autonomously. Moreover, if the design of inspection robots is a case-specific compromise between competing needs for sophisticated inspection sensing and for flexible locomotion in challenging field environments, that infrastructure could be even more complex. This complexity necessitates creating a contact point between the civil structure and the advanced device, bridging two domains with significantly different measurement tolerances. This detail can impact the movement accuracy of mobile systems, transforming a structured environment into an unstructured one. Consequently, this can lead to overly complicated navigation controls or incorrect positioning of the inspection probes. To address these limitations, either a human operator must remain involved, which reduces the benefits of automation, or innovative solutions must be introduced in autonomous systems. Therefore, this work aims to define the most influential parameters and provide potential solutions to improve the efficiency and accuracy of infrastructure inspection and maintenance.

Keywords: Inspection and Monitoring · Autonomous Robots · Infrastructures

1 Introduction

Considering that some of the expected outcomes of EU projects in Inspection and Maintenance (I&M) include extending the service life of civil engineering infrastructures, achieving faster and more accurate detection and analysis of maintenance and repair needs, reducing health and safety risks for workers, and saving costs in both operational and deferred terms, it is evident that current I&M technologies fall short of these expectations.

To reduce risks for operators, inspections have traditionally relied on fixed or temporary structures [1–4]. While these technologies allow access to difficult or dangerous areas, they do not eliminate the need for expensive inspection vehicles or minimize traffic flow disruptions. Consequently, mobile structures, both temporary (e.g., cable

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 373–377, 2024.
https://doi.org/10.1007/978-3-031-76428-8_69

crawlers) and permanent (e.g., travelers), were introduced, which required additional infrastructure such as rails or tracks [5, 6].

To further reduce human intervention for safety and cost reasons, the use of semi-autonomous and later fully autonomous robots was explored [7–11]. Early robotic bridge maintenance and inspection primarily relied on crawling or flying robots designed for specific tasks [12–14]. For instance, in 2000, Lorenc et al. proposed a tele-operated bridge maintenance system [15]. In 2010, Paul et al. developed an autonomous mobile manipulator for paint and rust removal on steel bridges, which included a moving platform with a manipulator, a high-performance computer, a laser scanner, and sensors [16]. In 2011, Chase and Adwards developed a lightweight, portable robotic system for bridge inspection, which operators could deploy easily and use to quickly assess bridge health [17].

Despite the high performance of robots developed between 2013 and 2022, their application in real scenarios has been limited [18]. This is not surprising, as civil infrastructures present challenging environments for autonomous machines. For example, on the San Giorgio Bridge in Genoa, Italy, a couple of fully autonomous robots are permanently installed and operate periodically [19, 20]. Various external factors (weather, traffic, operators) and intrinsic factors (geometric anomalies) pose significant challenges. That is why key questions arise: how can a robotic system be fully autonomous on infrastructure? How can it operate far from its control room? How can it make safe decisions in outdoor environments?

The answer lies in securing the robotic system to the infrastructure via tracks (referred to as "ancillary" infrastructures), which simplify the navigation environment and ensure safety in case of failure. This approach minimizes the need for constant human supervision and intervention.

This study collects data on ancillary infrastructures and suggests potential solutions to address these issues.

2 Project and Final Design Tolerances

2.1 Ancillary Infrastructures

Most permanent/mobile structures introduced in the §1 not only need the rails/track, but also weigh tons and the load on these ancillary infrastructures begings to be not negligible. Examples of these infrastructures are the motorized rolling traveller platform, which moves at a maximum rate of 60 feet/min (Fig. 1a) [6] and, the Robot Inspection (RI) and Robot Wash (RW), which are installed on the San Giorgio Bridge in Genova [19]. RI weighs over 2200kg, is equipped with 82 wheels for moving the two axes and is over 7m wide [22]. While, RW weighs about 2000kg and has 56 wheels to distribute the load on the deck edge; it is over 3.5m high, almost 8m long and is divided into two parts: one for cleaning and one for charging batteries (Fig. 1b) [23].

These structures are similar to bridge cranes, which travel on rails. If those rails become misaligned, the wheels of the crane will begin to wear out sooner than they should–and this can lead to a failure.

Fig. 1. a) (Left) Mobile Traveler [7] b) (Right) RI and RW [27].

In the traveler there is more tolerance between the actuators and the bridge structures, but, due to the huge size and weight of that platform, the misalignment of the rails could lead to wearing the lips of the edges of the bridge (Fig. 2a).

While the RI&RW (Fig. 2b) the rails misalignment could be the cause of contact between the actuators and trolleys the rails themselves. The robot structure is mechanically complex, and the high number of components increases the potential contact.

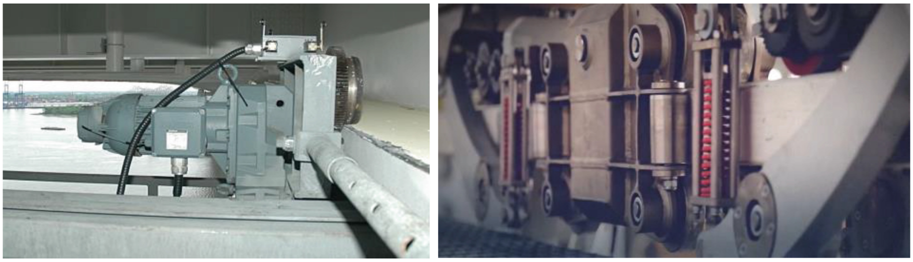


Fig. 2. a) (Left) Traveler: view of the drive motor and the lower lip of the north edge of the bridge. b) (Right) RI and RW: actuated wheels (red) and trolley wheels (still) roll on the rail.

2.2 Tolerances Between Design and Physical Structures

Because of the larger manufacturing tolerances of the civil infrastructures with respect to the mechanical ones, those already tight design gaps could not be met; they ended up being too small, creating the risk of contact/distance between robot and bridge structures. One of the crucial aspect at the design phase was the definition of the admissible tolerances for the proper operation of the robots. Mechanical tolerances and civil infrastructure tolerances can be different by several orders of magnitude.

Mechanical tolerances typically refer to the permissible deviation from specified dimensions or parameters in mechanical components such as gears, bearings, or machine parts, whose values are usually less than a millimeter. In contrast, civil infrastructure tolerances encompass a broader range of factors, including geometric, structural, and environmental considerations, whose value is more than a centimeter.

2.3 Discussion

To overcome the limitation given by the tolerances, several measures should be taken during the design phase: a) Isostatic connection between the robots and the bridge, b) 3D Collision detection analysis.

Following those results, the mechanical systems should be improved by some extra DoF and some extra control systems that take into account the potential contacts/distancing. This technological improvement is well-known in robotics. For the traveller, this extra control maintains its perpendicular alignment with the edge girders: e.g. if the left side travels a bit faster than the right side, then the overall traveller is skewed. While for the RI&RW it was applied the "Cognitive Mechatronics" [28] whose AI support the decision-making of the robots during their duties, with respect to the intrinsic and extrinsic unknowns due to the outdoor environment (§1).

3 Conclusions

The mechanical systems and the civil infrastructure have very different manufacturing tolerances, but taking into account this difference it is possible to adapt both. As shown in this work, both traveller and robots could become fully autonomous thanks to the simplified environment achieved by adding the ancillary infrastructure and the ability to adapt to the work environment and overcome the detrimental effects of the large manufacturing tolerances typical of a civil structure. The RI&RW were able to cope with these potential contacts with "Cognitive Mechatronics", which is currently working on the Genoa Bridge. Thus, the authors have demonstrated, for the first time, that it is possible to install and monitor civil infrastructures with autonomous robotic systems.

Acknowledgement. The authors would like to express their deepest gratitude to the Luca Molinari (InnSE Berardi) and Camozzi Group, for their fundamental contribution to building and installing this robotic system on the GSGB.

References

1. https://www.barin.it/products/ab-10-5-combi
2. https://www.zagro-group.com/en/produkte/hubarbeitsbuehnen/bridge-inspection
3. http://www.bridgeaccessspecialties.com/
4. https://www.harconcorp.com/bridge-inspection-equipment/topside-solutions/
5. SSS Suspended Scaffolding Solutions https://www.youtube.com/watch?app=desktop&v=R2zvxw4pUeA (Stanley Scaffolding Solutions Ltd https://find-and-update.company-information.service.gov.uk/company/11984668)
6. https://ravenelbridge.net/traveler.php
7. Sutter, B., et al.: A semi-autonomous mobile robot for bridge inspection. Autom. Constr. **91**, 111–119 (2018). https://doi.org/10.1016/j.autcon.2018.02.013
8. Nguyen, S.T., La, H.M.: A climbing robot for steel bridge inspection. J. Intell. Robot. Syst. **102**(4), 75 (2021). https://doi.org/10.1007/s10846-020-01266-1
9. La, H.M., et al.: Automated robotic monitoring and inspection of steel structures and bridges. Robotica **37**(5), 947–967 (2019). https://doi.org/10.1017/S0263574717000601

10. Tian, Y., et al.: Intelligent robotic systems for structural health monitoring: applications and future trends. Autom. Construct. **139**, 104273 (2022). https://doi.org/10.1016/j.autcon.2022.104273
11. Merkle, D., Schmitt, A., Reiterer, A.: Concept of an autonomous mobile robotic system for bridge inspection. In: Proceedings of SPIE – The International Society for Optical Engineering, 11535, art. no. 115350A (2020). https://doi.org/10.1117/12.2570633
12. Lattanzi, D., Miller, G.M.: Review of robotic infrastructure inspection systems. J. Infrastr. Syst. (2017). https://doi.org/10.1061/(ASCE)IS.1943-555X.000035
13. Bisht, R., et al.: Mobile robots for periodic maintenance and inspection of civil infrastructure: a review. s.l., International Conference on Machines and Mechanisms (2013)
14. Miyata, T., Yamada, H., Katsuchi, H., Kitagawa, M.: Full-scale measurement of Akashi-Kaikyo Bridge during typhoon. J. Wind Eng. Ind, Aerodynamics (2002)
15. Lorenc, S., et al.: Development of a robotic bridge maintenance system. Autom. Constr. **9**(3), 251–258 (2000)
16. Paul, G., et al.: A robotic system for steel bridge maintenance: Field testing. s.l., Australasian Conference on Robotics and Automation (2010)
17. Chase, S., Edwards, M.: Developing a tele-robotic platform for bridge inspection (2011)
18. Junho, A.L., et al.: Survey of robotics technologies for civil infrastructure inspection. J. Infrastruct. Intell. Resili. **2**(1), 100018 (2023). https://doi.org/10.1016/j.iintel.2022.100018
19. https://www.techsteel.net/why-the-new-bridge-in-genoa-is-among-the-worlds-most-advanced
20. https://opentalk.iit.it/robot-wash-e-robot-inspection-sul-podio-delleuropean-robotics-technology-transfer-awards-2022/
21. https://www.nord-lock.com/insights/customer-cases/2021/Rebuilding-Genoa-bridge/
22. https://www.structuremag.org/?p=25413
23. https://blog.pavementpreservation.org/2021/12/14/a-conversation-with-ferdinando-cannella/
24. https://www.gohispeed.com/keeping-it-on-the-rails-the-importance-of-overhead-crane-runway-rail-inspections/
25. Karagiannis, P., et al.: Cognitive mechatronic devices for reconfigurable production of complex parts. Appl. Sci. **11**(11), 5034 (2021). https://doi.org/10.3390/app11115034

Force-Balanced 2 Degree of Freedom Robot Manipulator Based on Four Bar Linkages

Yash Vyas[1], Marco Tognon[2], and Silvio Cocuzza[1](✉)

[1] University of Padova, Padova, PD 35121, Italy
yashjanardhan.vyas@phd.unipd.it, silvio.cocuzza@unipd.it
[2] Univ Rennes, CNRS, Inria, IRISA - UMR 6074, 35000 Rennes, France
marco.tognon@inria.fr

Abstract. We design a force-balanced 2 degree of freedom planar manipulator, which has minimized ground reaction forces, reaction torques and joint actuation torques. This manipulator is highly suitable for mobile robotics applications, in particular for aerial manipulation, as it enables greater precision and stability. The design synthesizes two four bar linkage mechanisms together in a planar configuration. Each mechanism is optimally force-balanced through an algorithm that extends the links, aggregates additional components, and applies counter-masses. We compare dynamic simulation results for a trajectory between the unbalanced and balanced designs to validate that the latter has 59% reduction in reaction torques and a constant reaction force vector caused by gravity on the base, as well as lower joint actuation torques.

Keywords: dynamic balancing · four bar linkage · robot manipulation

1 Introduction

Mobile robot manipulation offers several advantages in completing industrial tasks, such as safety, cost, and efficiency. However, manipulators designed for these applications are designed as serial or spatial closed-chain manipulators [3]. This results in significant reaction torques and forces caused by motion of the manipulator center of mass (CoM) and joint torques during manipulation.

Proposed solutions to this include differential inverse kinematic approaches [4,6], or mechanical design to balance the CoM [1]. In this research, we explore the idea of force balancing [7] to design a force-balanced manipulator suitable for mobile (particularly aerial) manipulation applications [3]. The CoM of the mechanism is static, resulting in a constant reaction force on the mechanism. Some

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no 101034319 and from the European Union - NextGenerationEU.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 378–383, 2024.
https://doi.org/10.1007/978-3-031-76428-8_70

existing investigation in this field includes balancing delta arms or pantographs [2,5].

In this research paper, we utilize kinematic equations for force-balancing [8] and synthesis [9] of closed-chain four-bar linkage (4BL) mechanisms, to design a 2 degree of freedom (DOF) planar manipulator. We apply the force balance equations of a four-bar linkage to an optimization methodology which calculates the counter-masses that balance the manipulator while minimizing the total mass. The dynamic behavior of unbalanced and balanced designs is verified in simulation and compared.

2 Background

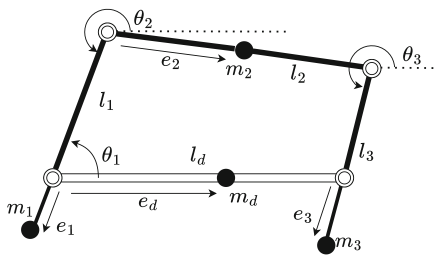


Fig. 1. A force-balanced four bar linkage diagram.

We use the generalized coordinates and kinematic parameters as shown in Fig. 1, assuming that the CoM is inline for each link. For link i with revolute joint causing motion about generalized coordinate θ_i, l_i is the length, e_i is the inline vector to the center of mass m_i, with d being the fixed bar. The conditions required for force-balance of a 4BL mechanism are derived in [8]:

$$m_1 e_1 + m_2(1 - \frac{e_1}{l_2})l_1 = 0 \qquad m_3(l_3 - e_3) + m_2\frac{e_2}{l_2}l_3 = 0 \tag{1}$$

For any set of values of $\theta_1, \theta_2, \theta_3$ that fulfill the closed loop constraints, the center of mass of the balanced 4BL is fixed and can be found in terms of coordinates e_i with respect to the axis of fixed bar l_d. It is important to note that the kinematic parameters defined here are aggregate for the entire link, and include counter-masses.

3 Method

We utilize the force balanced equations of a 4BL to design a 2-DOF manipulator, made by synthesizing two force-balanced 4BLs as outlined in [9] for a dynamically balanced four bar linkage. An illustrative figure of the mechanism is shown in

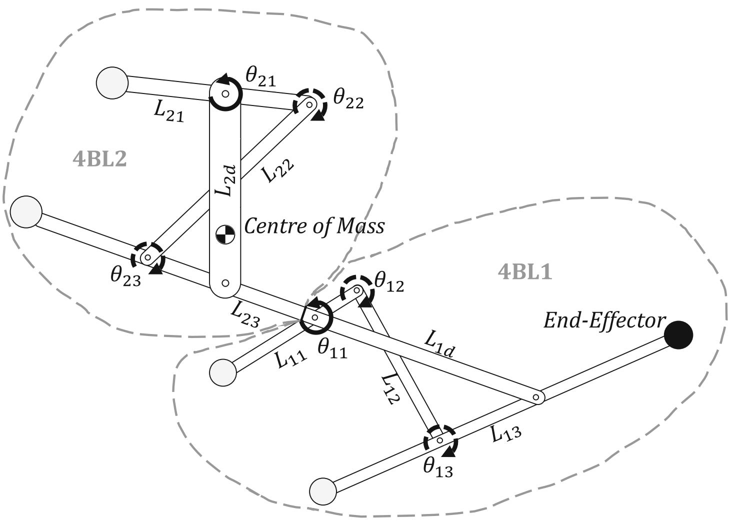


Fig. 2. The 2DOF manipulator built from mounting 4BL1 on 4BL2. Joints are indicated by their coordinate θ_{ji} and links are L_{ji}. Actuated Joints (solid arrow), passive joints (dashed arrow), and counter-masses (shaded circle) are indicated.

Fig. 2. Kinematic parameters and coordinates use dual indexes ji, where j is the 4BL and i is the link or joint index. 4BL1 with $j = 1$ is attached to 4BL2 with $j = 2$, which is mounted on the base.

To conserve space, simplify the mechanism design, and reduce interference, we opt for a crossed over kinematic configuration for link 2 with the counter-masses attached to links with $i = \{1, 3\}$. Constant parameters are the link lengths, as well as positions and loads of additional mechanical components (e.g. joints, motors, and payload at the end effector). These are aggregated into a single link mass m_i with center of mass at e_i. Each 4BL is then optimized through extending the link geometry and attaching counter-masses through a constrained optimization function, where the objective function is minimization of the total mass min $\sum_{i=1}^{3} m_i$, variables are the counter-masses, subject to (1).

The optimization is done recursively, e.g. 4BL1 is optimized first, and then modeled as a single mass load applied on L_{23}, which is then also optimized. The combined manipulator is 2-DOF as θ_{11} moves the end effector through constrained motion of θ_{13}, and 4BL1 moves in relation to θ_{23} which moves constrained as actuated by θ_{21}.

4 Results

We applied the method outlined to design a 2-DOF force balanced manipulator that is comparable to a serial link manipulator with maximum extension of 0.5 m. 4BL1 is mounted with extension $e_{23} = 0.35$ m on 4BL2 and the end-effector is mounted at an extension of $e = 0.35$ m on 4BL1, to achieve this maximum extension. We set the length proportions $l_{j1} = l_{j3}$ and $l_{j2} = l_{jd}$, so that singularities occur when L_{j1}, L_{jd}, and L_{j3} align simultaneously. The ratio of link lengths selected for $L_{j1} : L_{j2}$ results in a roughly elliptical and convex workspace with an area of $0.386\,\text{m}^2$.

We model the mechanism with revolute joints (consisting of bearings), link mass, motors (attached to the fixed bars) and an end-effector payload of 300 g, similar to an equivalent 2-DOF serial link manipulator used to move a load. The original, unbalanced mechanism with these loads is then balanced by extending the links to include counter-masses.

Table 1 shows the geometric and kinematic parameters for each link with additional loads (joints and payload) aggregated with the link profile, for both the unbalanced and balanced manipulator mechanisms. The mass of the unbalanced mechanism is 1.464 kg and the balanced mechanism is 2.330 kg (44% heavier).

Table 1. Parameters for the design of the synthesized 4BL, for the links shown in Fig. 2. Parameters are as denoted in Figs. 1 and 2, and $_{cm}$ indicates the counter-mass separately. I is the moment of inertia of each link about joint θ_{ij}.

		Unbalanced			Balanced				
Link	l (m)	m (kg)	e (m)	$I(10^{-3}$kg .m$^2)$	m_{cm} (kg)	e_{cm} (m)	m (kg)	e (m)	$I(10^{-3}$kg .m$^2)$
L_{21}	0.16	0.0728	0.0386	3.076	0.06124	−0.1579	0.1687	0.0569	2.490
L_{22}	0.2	0.0816	0.0538	0.5933	–	–	0.0822	0.0541	0.5906
L_{23}	0.16	0.1057	0.0997	0.9266	0.7576	−0.1579	0.8984	−0.126	21.32
L_{2d}	0.2	0.1636	0.0729	3.891	–	–	0.1636	0.0729	3.891
L_{11}	0.12	0.0640	0.0247	0.1342	0.05417	−0.12	0.1445	−0.0450	1.199
L_{12}	0.15	0.0706	0.0350	0.2548	–	–	0.0706	0.0350	0.2548
L_{13}	0.12	0.1079	0.1041	0.7137	0.2149	−0.12	0.3492	−0.0462	4.408
L_{1d}	0.15	0.1526	0.0532	2.117	–	–	0.1526	0.0532	2.117

The design was implemented in MATLAB, through modeling of the links and optimization of counter-masses using the `fmincon` function, and finding the corresponding joint velocities for an end-effector trajectory using inverse kinematics as a constrained least squares minimization problem. The end-effector was moved smoothly to the waypoint $[x, z]^T = [0.15, 0.1]^T$ relative to its original position in 1 s, holds for 0.5 s, and moves back to the starting position in 1 s. The unbalanced and balanced models were implemented in ADAMS to obtain accurate estimates of reaction torques/forces on the base (as measured at the mount point), and the joint torques. The results are shown in Fig. 3.

We observe force balancing in the optimized mechanism through a constant linear reaction force for the optimized manipulator, whereas the unbalanced fluctuates to a maximum of 0.775 N. Furthermore we can observe a much lower maximum absolute reaction torque of 0.595 N.m for the balanced mechanism as

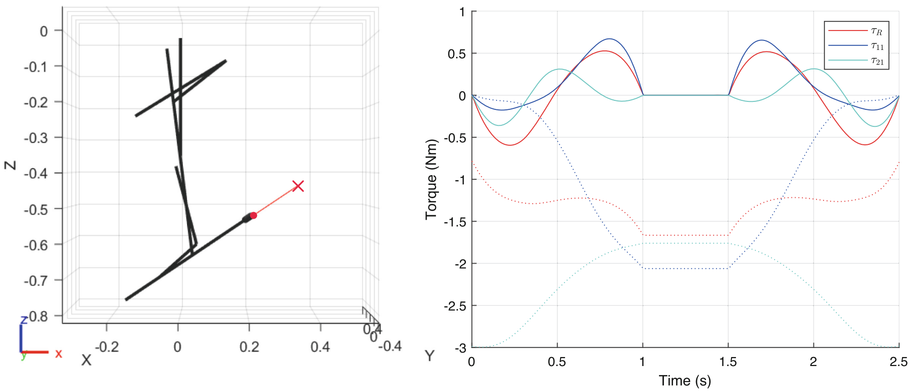


Fig. 3. Trajectory tracked at left: indicating start/end position (dot) and tracked way-point (cross). Torques at right: τ_R is the reaction torque, actuated joint torques are τ_{11} and τ_{21} for balanced (solid) and unbalanced (dotted).

compared to 1.66 N.m for the unbalanced mechanism. The joint torques are also reduced by 67–92% due to reduction of gravity torques.

5 Conclusions

A 2-DOF force-balanced manipulator composed by synthesizing two 4BLs offers several benefits, the most important being a reduction in reaction torques and fluctuations in the reaction force. By aligning the CoM of the manipulator to the base mounting point, we also eliminate any reaction torques caused by gravity forces. Although the inertia of the mechanism increases due to the addition of counter-masses and extension of links, we find that the increase in motor and reaction torques from this is less with respect to gravity-induced torques in the unbalanced case.

Force balancing also helps to reduce the disturbance forces and vibrations between the base and manipulator. This makes this manipulator suitable for mobile platforms as it facilitates more precise execution of manipulation tasks. As the reaction forces are constant and the reaction torques are reduced, it also enables a much simpler control mechanism for the mobile base. In future research to further explore this concept, we will investigate different tasks such as load picking/interaction with the environment, and the dynamic interaction between the base and manipulator.

References

1. Abuzayed, I., Itani, A.R., Ahmed, A., Alkharaz, M., Jaradat, M.A., Romdhane, L.: Design of lightweight aerial manipulator with a CoG compensation mechanism. In: 2020 Advances in Science and Engineering Technology International Conferences (ASET). IEEE (2020)

2. Clark, A.B., Baron, N., Orr, L., Kovac, M., Rojas, N.: On a balanced delta robot for precise aerial manipulation: Implementation, testing, and lessons for future designs. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2022)
3. Ollero, A., Tognon, M., Suarez, A., Lee, D., Franchi, A.: Past, present, and future of aerial robotic manipulators. IEEE Trans. Rob. **38**(1), 626–645 (2022)
4. Pasetto, A., Vyas, Y., Cocuzza, S.: Zero reaction torque trajectory tracking of an aerial manipulator through Extended Generalized Jacobian. Appl. Sci. **12**(23) (2022)
5. Suryavanshi, K., Hamaza, S., van der Wijk, V., Herder, J.: ADAPT: a 3 degrees of freedom reconfigurable force balanced parallel manipulator for aerial applications. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 11,936–11,942 (2023)
6. Vyas, Y., Pasetto, A., Ayala-Alfaro, V., Massella, N., Cocuzza, S.: Null-space minimization of center of gravity displacementof a redundant aerial manipulator. Robotics **12**(2) (2023)
7. Wei, B., Zhang, D.: A review of dynamic balancing for robotic mechanisms. Robotica **39**(1), 55–71 (2020)
8. van der Wijk, V.: Methodology for analysis and synthesis of inherently force and moment-balanced mechanisms. Ph.D. thesis, University of Twente (2014)
9. Wu, Y., Gosselin, C.M.: Synthesis of reactionless spatial 3-DoF and 6-DoF mechanisms without separate counter-rotations. Int. J. Robot. Res. **23**(6), 625–642 (2004)

A Cognitive Architecture for Socially Assistive Robots

Devis Dal Moro(✉), Magí Dalmau-Moreno(✉), Josep Bravo(✉), Federica Gabriella Cornacchia Loizzo, and Daniel Serrano

Eurecat, Centre Tecnològic Catalunya, Robotics and Automation Unit, Barcelona, Spain
{devis.dalmoro,magi.dalmau,josep.bravo}@eurecat.org

Abstract. Recent studies have tried to approach through Socially Assistive Robotics (SAR) the challenge of supporting elderly individuals living alone. In this paper, we introduce a cognitive architecture for SAR, incorporating planned and reactive behaviors. We emphasize the role of generative AI to provide a high-level semantic understanding and reasoning over dynamic environments, allowing more natural and flexible interactions with the user. The architecture has been deployed using the TIAGo mobile robotic manipulator in a simulated kitchen environment.

Keywords: Socially assistive robots · HRI · cognitive architecture

1 Introduction

Recently conducted demographic analyses are reporting a rise in the number of elderly individuals living alone, with a harmful impact on elderly people's well-being and mental health [5,6]. The impact of Socially Assistive Robot (SAR) does not replace human relationships but serves as a valuable technology aid, surpassing standalone software capabilities. Indeed, a physical presence offers a platform for more expressive and empathetic interactions. Motions of the robot are not exclusively means to complete a task, but can be synchronized to complement verbal communication, i.e. to transmit expressions [8] and aid in the user's ability to understand the robot [4]. Therefore, developing natural relational capabilities becomes crucial. A SAR agent requires high-level understanding and reasoning in dynamic, open environments, similar to human capabilities. Recent advancements in AI could provide tools for a significant step forward.

In this paper, we propose a cognitive robotics architecture for socially assistive robots, empowered by generative AI and fitted in the renowned Sense-Deliberation-Act design pattern, through the usage of behavior trees, a de facto standard in robotics. We recently deployed such architecture in the development of the Never Home Alone (NHoA) project[1]. The latter aims to create an intelligent and social robotic system that assists elderly people in maintaining their independence in their homes and preventing loneliness and isolation by promoting an active, connected and healthy lifestyle in the day to day.

[1] NHoA website: https://nhoa-project.eu/.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 384–388, 2024.
https://doi.org/10.1007/978-3-031-76428-8_71

2 Related Work

Recent studies [1] highlight the growing acceptance among older individuals for socially assistive robots (SARs), indicating a willingness to embrace these robots for future use, especially in tasks related to physical assistance. SARs show promise in enhancing social connections, mitigating loneliness, improving medication compliance, and fostering independence. Several social robots have been used in the literature to conduct studies with users for providing companionship, assistance, and various services, especially targeted towards the elderly. Research examples include robots like ARI Robot [2]. While pet robots offer emotive experiences, their limitations often lie in the lack of social interactions. Tabletop robots excel in static monitoring and stimulation but necessitate a smart environment and wearables for comprehensive monitoring. Mobile social robots, although making strides, may lack certain physical assistance capabilities. Mobile manipulators, in assistive robotics, provide potential comprehensive solutions for domestic tasks, but persistent challenges include size and adaptation to home settings, necessitating predictable and understandable robot motions [3] for enhanced trust and natural interaction.

3 Methodologies

Figure 1 depicts the high level architecture of our SAR agent. It implements a Sense-Deliberation-Act pattern through three different main modules running independently and exchanging information through a shared space of memory. The sense module groups together a mixture of submodules, retrieving low-level sensor data and converting into high-level semantic information about the surrounding environment and the person we are interacting and engaging with. Deliberation phase processes this constantly updated understanding of the world and, combining with a medically provided plan and an adaptive patient profile, decides and refines on the actions to be put in a queue for execution. Finally, the Act module takes the role of the controller, coordinating the start and stop of mutual exclusive actions and concurrent ones, i.e. a head following motion can run concurrently with a conversation routine.

The deliberation phase is the core component of our cognitive architecture. Its task is to manage two execution queues, selecting the high level skill to be executed by the SAR agent. It can be split in two relevant main phases: 1. skill selection, 2. skill refinement. In the first step, we select a skill either from any reactive trigger or from a daily plan (e.g. a medical proposed agenda). Reaction behaviors are triggered in response to a critical event detected in the scene (e.g. patient has fallen to the ground), or some voice activated commands (e.g. user request for help in case of need). When activated, actions to be executed are pushed to a priority queue. The latter is always prioritized: as long as it is not empty, the daily plan queue remains in a waiting state until it can be resumed. The second step processes potentially refine the actions arguments and decide whether some parallel actions shall be activated, before passing the

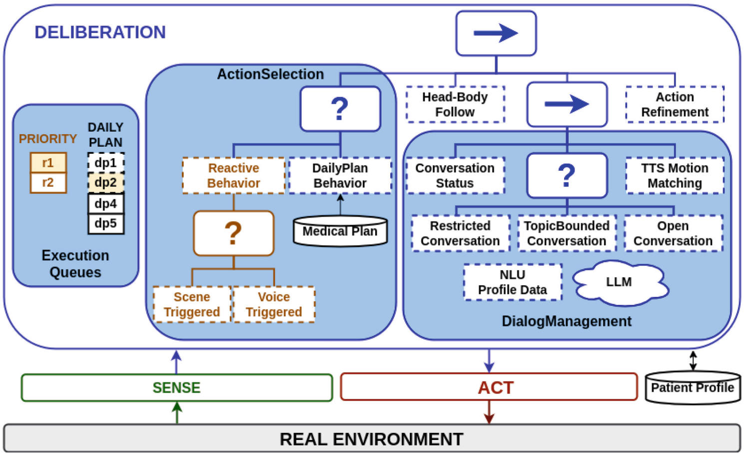


Fig. 1. Diagram of proposed SAR architecture consisting of three main modules running independently: sense, deliberation and act.

command to the act module. Action refinement involves a decision process based on specifics of the skill (e.g. selecting the next thing to say and how to say it during a conversation) or of the user profile (e.g. customizing the next physical exercise during a training session). Given its relevance, exclusively conversation-related refinement is exploded in the figure, leaving a more general-purpose box at the end for all the other actions. Throughout all the deliberation process, generative AI, in conjunction with 3D scene understanding modules, enables our system to dynamically interpret and respond more naturally to external triggers, provided by the user or the environment, making our SAR agent able to reason more effectively about high-level concepts. For instance, implying state-of-the-art Visual Language Models (VLM) and Large Language Models (LLM), each object in our scene could be mapped to distinct embedding spaces, enabling both visual and textual encoding. This would allow empowering the NLU capabilities of the SAR robot to match something said by the elderly patient to some object in the scene and its position.

One of the most relevant deliberative refinement tasks occurs during a conversation when, after receiving an answer from the user, we need to generate both verbal and motion responses. In this situation, we distinguish three main cases: 1. restricted conversation 2. topic bounded conversation 3. completely open conversation. The first option presents a clearly defined and constrained dialog scenario with specific answer possibilities for the purpose of adhering to an established protocol and gathering accurate medical information (e.g. medical questionnaire). To introduce a degree of flexibility and adaptation in the dialog, string distance metrics or small NLU models shall be employed to attempt to match the user's answer to the set of possible ones. More intriguing are the other two conversation options. Both leverage recent state-of-the-art LLMs, such as GPT-4 [7], for a natural response generation. Nuanced fine-tuning mechanisms take place through prompt engineering techniques that take into account the role of the SAR agent, the conversation history, the user profile and the relevant high-level concepts available and reconstructed from the scene. The topic-bounded option considers an additional element to constrain the scope of the conversation. The final step in dialog deliberation is a query to an LLM to evaluate the possibility of matching the text-to-speech (tts) for the robot to a predefined expressive motion set of the robot, performed only if space and safety measures allow doing so.

4 Use Case

The presented methodology has been deployed in a realistic scenario. As a SAR agent we used TIAGo[2]. The architecture has been implemented through Behaviour Trees. The daily routine has been created with healthcare professionals and elderly participants. TIAGo engages in environment observation, guiding the elderly to a seated position and initiating interactive dialogues. TIAGo transitions into health monitoring by guiding the user through a structured medical questionnaire in a closed conversation, ensuring accuracy and efficiency. The robot then acts as a virtual trainer, leading the patient through personalized exercise routines, monitoring performance and providing positive feedback. Finally, TIAGo provides the user with an overview of the day, fostering a sense of connection. The demonstration is flexible, allowing the user to interrupt the plan and initiate open conversations, seek objects in the room, request assistance, or alter the predetermined routine. The robot's gaze and head following the user throughout the entire demo contribute to a seamless and responsive user experience (Fig. 2).

Fig. 2. Images taken from dry runs of the demo. On the left TIAGo approaching and chatting with a (supposedly) elderly patient, on the right guiding him through a physical exercise session.

[2] https://pal-robotics.com/es/robot/tiago/.

5 Conclusions

We propose a cognitive architecture empowering a SAR with high-level reasoning capabilities empowered by generative AI to naturally engage and support an elderly patient living alone through daily tasks. Both planned and reactive behaviors are displayed by the robot. The architecture and specific use cases have been deployed in a realistic home environment with a TIAGo Robot.

Further experimentation will be conducted in real, in-home case studies with elderly patients, retrieving further data to evaluate effectiveness of SAR architecture in this field and compare it with similar frameworks. The possibility of learning new skills through human interactions and feedback, along with the ability to express more effectively emotions through semantic mapping are also left for future work.

Acknowledgements. This work was financially supported by the Catalan Government through the funding grant ACCIO-Eurecat (Project Shark) and within the NHoA project with code PLEC2021-007868, funded under MCIN/AEI/10.13039/50110001 1033 and by the European Union "NextGenerationEU"/PRTR.

References

1. Bardaro, G., Antonini, A., Motta, E.: Robots for elderly care in the home: a landscape analysis and co-design toolkit. Int. J. Soc. Robot. **14**, 657–681 (2022). https://oro.open.ac.uk/78674/
2. Cooper, S., et al.: ARI: the social assistive robot and companion. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 745–751 (2020). https://doi.org/10.1109/RO-MAN47096.2020.9223470
3. García, N., Suárez, R., Rosell, J.: Planning hand-arm grasping motions with human-like appearance. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3517–3522 (2018). https://doi.org/10.1109/IROS.2018.8594432
4. Kruse, T., et al.: Legible robot navigation in the proximity of moving humans. In: 2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 83–88 (2012). https://doi.org/10.1109/ARSO.2012.6213404
5. Leigh-Hunt, N., et al.: An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health **152**, 157–171 (2017). https://doi.org/10.1016/j.puhe.2017.07.035. ISSN 1476-5616
6. Ong, A.D., Uchino, B.N., Wethington, E.: Loneliness and health in older adults: a mini-review and synthesis. Gerontology **62**(4), 443–449 (2016). https://doi.org/10.1159/000441651. ISSN 1423-0003
7. OpenAI et al.: GPT-4 Technical Report (2023). arXiv: 2303.08774 [cs.CL]
8. Zhou, A., Dragan, A.D.: Cost Functions for Robot Motion Style (2018). https://arxiv.org/abs/1809.00092v1

Towards Autonomous Robotic Procedure for Ultrasound-Guided Percutaneous Cardiac Interventions for Mitral Valve Repair

Angela Peloso(✉), Riccardo Munafò, Veronica Ruozzi, Anna Bicchi, Xiu Zhang, Elena De Momi, and Emiliano Votta

Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy
angela.peloso@polimi.it
https://www.deib.polimi.it/

Abstract. The paper outlines the early stages of a robotic platform designed to improve safety and repeatability of Transcatheter Edge-to-Edge Repair (TEER) procedures. The platform integrates artificial intelligence (AI) software for image interpretation, sensor-equipped catheters for virtual monitoring, robotic actuators for manipulation, and a mixed-reality interface for real-time monitoring. The AI software can automatically identify Mitral Valve (MV) anatomical features from ultrasound images. The platform also includes a path planning module and an inverse kinematic controller for safe navigation. The use of novel sensors and automatic actuation allows for precise control. Real-time simulation of the catheter's interactions provides accurate analysis of anatomical deformations. These developments represent significant progress in percutaneous intracardiac procedures, with the potential to make TEER procedures safer and more accessible.

Keywords: percutaneous cardiac interventions · mitral valve · 3D transesophageal echocardiography · path planning · inverse kinematic control · mixed reality

1 Introduction

Here, we report our recent advances toward the development of a shared autonomy robotic platform designed to make percutaneous intracardiac procedures safer, easier to learn, more repeatable, and less physically and mentally demanding. The work is carried out within the European Union's EU Research and Innovation Programme Horizon 2020 framework under the project ARTERY, grant agreement No. 101017140. We illustrate these advances with reference to TEER using the MitraClip^{TM} System (Abbott, IL, USA).

A. Peloso, R. Munafò, E. De Momi and E. Votta—These authors contributed in the same way.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 389–393, 2024.
https://doi.org/10.1007/978-3-031-76428-8_72

2 Methods

This shared autonomy robotic platform is characterized by six key technologies:

1. AI-based software to simplify the interpretation of intraprocedural imaging and the identification of the target structures in the intracardiac space;
2. AI-based control software to automatically identify the set of maneuvers to be performed at the proximal end of the catheter to safely reach the target pose of the catheter tip, as well as to control and adjust on-the-fly the associated catheter route to the target;
3. novel sensors embedded on the relevant catheters, which can hence be virtually monitored;
4. robotic actuators that can implement the required maneuvers at the proximal end of the guide sheath and of the Mitraclip catheter;
5. robotic actuators to manage the Real-Time 3D Transesophageal Echocardiography (RT3DTEE) probe;
6. a mixed reality (XR) user interface allowing operators to interact with the robotic platform with improved ergonomy and from the distance

Here we provide an overview specifically on the key-points from 1 to 5.

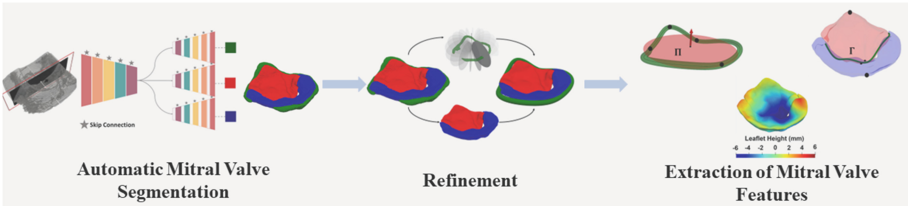


Fig. 1. Representation of the embedded pipeline: a 3D surface is extracted from MV substructures segmentation automatically generated by the MdResUNet [2]. This surface is refined before anatomical features (valve plane (Π), coaptation line (Γ), and leaflet height) are automatically computed.

2.1 Real-Time Image Processing

The AI-based software developed for intraprocedural imaging processing serves to automatically identify anatomical features of MV from RT3DTEE images, assisting in the precise placement of the Mitraclip device. The software utilizes a deep learning-based pipeline, analyzing systolic frames from 3DTEE scans and reconstructing a 3D surface of the MV in its closed configuration (Fig. 1). The architecture outperformed a state-of-the-art baseline ResUNet, and provided an unprecedented reconstruction and multi-class segmentation of mitral annulus, anterior leaflet, and posterior leaflet [1].

2.2 Catheter Navigation

To reach the autonomous movement of the catheter, we adopted imitation learning combining two approaches: Generative Adversarial Imitation Learning (GAIL) and Behavioural Cloning (BC). GAIL [3] uses a generator to propose actions based on proximal policy optimization with external rewards (R_{ext}, indicating which actions are good to accomplish the task) and a discriminator to differentiate these actions from expert demonstrations, providing an intrinsic reward (R_{int}, degree of similarity). During the training process, the generator gets better at mimicking the demonstrations, while the discriminator gets stricter in learning to differentiate the two types of paths. BC enhances the imitation of expert behavior. The model was trained in a simulation environment and evaluated among 100 simulated paths.

2.3 Catheter Actuation and Control

The tendon-driven robotic catheter was automated using two types of stepper motors (Nema 23, JoyNano - Nema 17 Sainsmart). This allowed for control of the catheter's flexion in the mediolateral and anteroposterior planes and tip insertion movement. The system uses a model-free feedback controller with a multilayer perceptron to map the tip position and the actuator state. The robot explores its workspace, recording the tip pose corresponding to specific motor commands, with data collected through three 6-degrees-of-freedom electromagnetic sensors (EM). The accuracy of the Inverse Kinematic Controller (IKC) was evaluated against a model-based approach using a Proportional-Integral-Derivative controller [4], showing notable improvement in the yaw angle associated with the mediolateral bending plane and minimal error in the pitch angle corresponding to the anteroposterior plane bending (Fig. 2).

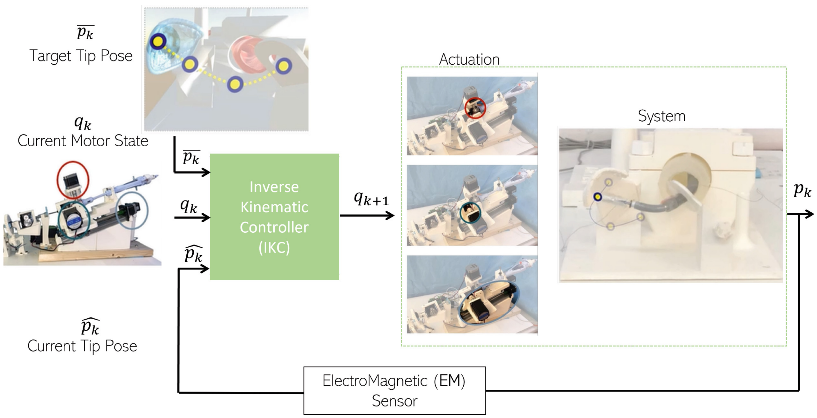


Fig. 2. Scheme of the Inverse Kinematic Controller. The inputs are the desired pose at time $(k+1)$, $\bar{\boldsymbol{p}}_{k+1}$, the current servomotors position, $\boldsymbol{q}_k$, and the current tip pose $\hat{\boldsymbol{p}}_k$, measured by EM sensor. The output is the position of the servomotor at the next time instant, $\boldsymbol{q}_{k+1}$

2.4 3DTEE Actuation

The RT3DTEE probe can be moved in four ways for precise imaging of the heart: advancing and withdrawing the main body of the probe along the esophagus, rotating it around the long axis, and bending the probe tip in both directions using co-axial handwheels and an internal cable-driven continuum mechanism within the endoscope. Our proposed 3DTEE robot motorizes these movements using custom mechanisms (Fig. 3), adapting to different probes like Philips' x7-2t (Philips, Netherlands) and GE's 6 VT-D ((6 VT-D, GE HealthCare, United States). The robot uses servo motors (XM430-W350-R, Dynamixel, Keara) and spur gears for efficient power transfer and consistent rotation. This allows for precise control over the ultrasound image plane.

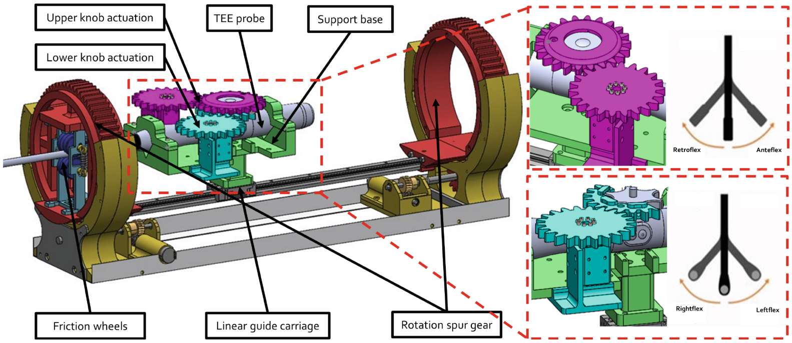


Fig. 3. Mechanical assembly of the RT3DTEE

2.5 Real-Time Vessel Deformation

To minimize damage in catheter guidance, a real-time simulation of catheter interaction was developed using sensor-equipped catheters [5]. The 3D anatomy of a male's right femoral vein and inferior vena cava was reconstructed from CT scans. A numerical simulation was developed in the Sofa framework. The catheter's 3D configuration is imported into the simulation, enabling interaction with the vessel. Real-time deformation visualization consequent to the catheter advancement was possible thanks to the embedding of Fibre Bragg Grating (FBG) and EM sensors in the distal part. An XR interface was implemented in Unity and visualized through a Hololens^{TM} 2 headset (Fig. 4). The Robotics Operation System (ROS) protocol is used for data flow, simulating vessel deformation due to catheter interaction, and updating the virtual model in the headset.

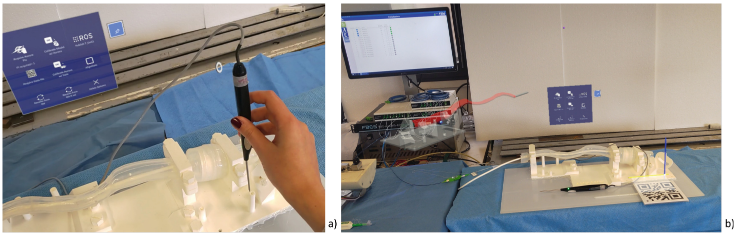


Fig. 4. Data recording received from sensors in the same frame as the anatomical model (a); interaction between the sensorized catheter and the vessel reproduced in a digital twin, visualized in XR through the Microsoft Hololens$^{\text{TM}}$ 2 headset (b).

3 Discussion and Conclusion

The paper presented the early stages of a shared autonomy robotic platform designed to improve percutaneous intracardiac procedures, specifically TEER for MV repair. The platform combines AI software, sensor-equipped catheters, robotic actuators, and an XR user interface. The AI software uses a deep learning pipeline for real-time image processing and path planning for safer guidance inside critical anatomical regions. The catheter and RT3DTEE probe are automated and equipped with sensors for precise control and real-time simulation of the catheter's interaction with the vessel. This allows for an enhanced ergonomic experience and real-time visualization without invasive imaging techniques like fluoroscopy. Future work will focus on refining these prototypes into a fully integrated platform to enhance TEER procedures. This work is a significant step in the field of percutaneous interventions, and such innovations will play a crucial role in shaping its future.

References

1. Munafò, R., et al.: A deep learning-based fully automated pipeline for regurgitant mitral valve anatomy analysis from 3D echocardiography. IEEE Access (2024)
2. Kerfoot, E., Clough, J., Oksuz, I., Lee, J., King, A., Schnabel, J.: Left-ventricle quantification using residual U-Net. Statistical Atlases and Computational Models of the Heart (2019)
3. Ho, J., Ermon, S.: Generative adversarial imitation learning. In: Advances in Neural Information Processing Systems (2016)
4. Zhang, X., et al.: Robotic actuation and control of a catheter for structural intervention cardiology. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2022)
5. Ha, X.T., et al.: Robust catheter tracking by fusing electromagnetic tracking, fiber Bragg grating and sparse fluoroscopic images. IEEE Sens. J. (2021). https://doi.org/10.1109/JSEN.2021.3107036

Comparison Between Force and Ultrasound Image-Based Controllers for Autonomous Robotic Ultrasound Acquisition in Different Tissue Types

F. J. Domingo Gil, A. G. de Groot(✉), and F. J. Siepel

Robotics and Mechatronics Group, University of Twente, Enschede, The Netherlands
a.g.degroot@utwente.nl

Abstract. The main challenges in automatic robotic ultrasound (US) scanning include ensuring acoustic coupling between the tissue and the US probe, while minimizing tissue deformation. In this comparative study, a force controller, confidence-driven controller (CDC), and hybrid force-image controller are tested to perform automatic robotic ultrasound (US) scan. The controllers have been tested using a 7DoF robot equipped with a force sensor and a linear US transducer. Obtaining a robotic 3D US acquisition would increase the reproducibility and accuracy of the diagnostic, and at the same time, would simplify the registration of 3D US volumes with MRI images. The robot follows a reference trajectory which is adjusted by any of the controllers to ensure acoustic coupling. The CDC makes the corrections based on confidence maps. The force and hybrid controllers correct the probe contact based on the contact force of the end-effector with the tissue. The CDC and hybrid controller ensure normal orientation to the tissue by controlling in-plane rotations based on the confidence map. The three controllers are tested in three phantoms of different stiffness.

Keywords: autonomous US acquisition · force controller · confidence maps

1 Introduction

Breast cancer remains a global health concern, as 1 out of 8 women are diagnosed with breast cancer in their life. Early detection through reliable imaging modalities such as mammography, MRI, and ultrasound (US) is essential for positive prospects [5]. While US-guided biopsies are preferred for their real-time feedback, they face challenges like depth assessment in 2D scans, leading to potential misdiagnosis. The possibility of obtaining direct 3D volumes has been studied, by utilizing mechanical 3D probes or freehand scanners [2]. However, all these solutions depend on highly-trained human operators, what reduces the repeatability of the process. To address this challenge, the possibility of performing autonomous robotic US acquisition is studied.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 394–398, 2024.
https://doi.org/10.1007/978-3-031-76428-8_73

Robotic systems offer advantages in trajectory precision, reduced variability, and higher repeatability. Despite these benefits, challenges like patient movement or tissue deformation may affect US image quality during autonomous robotic US acquisition. To solve these issues, control strategies, including image feedback, force feedback, and their combination, have been developed. Image feedback, uses confidence maps, to ensure acoustic coupling and optimize image quality [1,4]. Force feedback maintains a constant force to minimize tissue deformation, while a hybrid approach combines force and image feedback for optimal control.

In this study, a force feedback controller and a hybrid force-image feedback controller for automated robotic US scanning are developed. Its behaviour is tested in three different phantoms with different stiffness, and compared to the behavior of a confidence map controller. Testing the controllers in different tissue types will provide insight into the optimal situation to use each of them.

2 Materials and Methods

2.1 Robotic Setup

For this research, a KUKA LBR MED 7 robotic arm was used. Attached to the flange of the KUKA arm is an ATI Mini-40 Force/Torque (F/T) sensor (Manufacturer page). Additionally, at the top of the sensor, an end-effector (EE) which includes a linear US transducer is fixed. Three phantoms are used to mimic different tissue types and complement the robotic setup. The phantoms were fabricated with a solution mixture of Plastisol and Softener in the three following concentration ratios: 100/0, 80/20, and 60/40 (Plastisol/Softener).

2.2 Control Strategy

The robot follows a predefined trajectory, which ensures that the whole tissue is scanned. However, trajectory errors and environmental uncertainties, like patient movements or breath, can affect acoustic coupling. Additionally, for optimal registration with modalities such as MRI, tissue deformation must be minimized. Three different controllers are implemented to address these challenges.

2.3 Confidence Driven Controller (CDC)

As Welleweerd et al. reported in their study, confidence maps can be used to control two degrees of freedom of the probe: the in-plane rotation and the translation in the z-direction of the EE frame, ψ_{ee} [4]. Please refer to [4] for a detailed explanation of this controller.

2.4 Force Controller

To use the force sensor data, a calibration procedure was developed. It ensures bias calibration, removal of the EE static mass [3], and reference frame alignment.

In the force control algorithm, F_z can be used to satisfy a constant contact force setpoint that ensures acoustic coupling, by controlling the z-translation in the EE reference frame. The error is determined by the difference between the force setpoint, F_{set}, and the current force reading in the z-direction, F_z: $e = F_{set} - F_z$. To control F_z a simple PD controller is used. The trajectory pose $H(i)^0_{ref}$ of the robot is adjusted by $H(i)^{ref}_{adj}$ at time i as follows:

$$H(i)^0_F = H(i)^0_{ref} H(i)^{ref}_{adj} = H(i)^0_{ref} H(i-1)^{ref}_{adj} H_{F\Delta z} \tag{1}$$

with

$$H_{F\Delta z} = \begin{bmatrix} I^{3\times 3} & d\hat{z} \\ 0^{1\times 3} & 1 \end{bmatrix}$$

$H(i)^{ref}_{adj}$ includes the current and previous outputs of the force control algorithm. The force control algorithm outputs a translation, Δz, in z-direction, $H_{F\Delta z}$.

2.5 Hybrid Force-Image Controller

An important factor in maximizing the quality of the US images is maintaining the US transducer normal to the scanning region. The force controller in the current setup lacks this capacity, therefore a hybrid force-image controller is used for normal orientation control. Similar to the previously explained force controller, the z-translation in the EE reference frame is controlled by F_z and a force setpoint. However, the in-plane rotation is controlled by the visual servoing algorithm, based on the confidence-weighted barycentre. Equation 1 is extended according to:

$$\begin{aligned} H(i)^0_{Fvs} = H(i)^0_{ref} H(i)^{ref}_{adj} = H(i)^0_{ref} H(i-1)^{ref}_{adj} H(i)^{adj}_{Fvs} \\ H(i)^{adj}_{Fvs} = H_{F\Delta z} H_{\Delta\theta} \end{aligned} \tag{2}$$

with

$$H_{F\Delta z} = \begin{bmatrix} I^{3\times 3} & d\hat{z} \\ 0^{1\times 3} & 1 \end{bmatrix} H_{\Delta\theta} = \begin{bmatrix} Rot_y(\theta) & 0^{3\times 1} \\ 0^{1\times 3} & 1 \end{bmatrix}$$

In addition to the z-translation obtained from the force control algorithm, the control algorithm outputs in this case a rotation $\Delta\theta$ around the y-axis of the transducer, $H_{\Delta\theta}$ as in the case of the CDC.

3 Results

Figure 1 shows a comparison of the mean confidence and its corresponding contact force, F_z, for each of the 10 experiments for each confidence setpoint in each phantom using the CDC. It can be seen that the confidence stays constant in the 3 phantoms, while the force varies. Furthermore, it is observed how F_z for 80% contact in the hard phantom differs a lot between tests.

To assess image quality and the effect of in-plane corrections, the mean barycenter error was used. Figure 2 shows the mean barycenter error of each

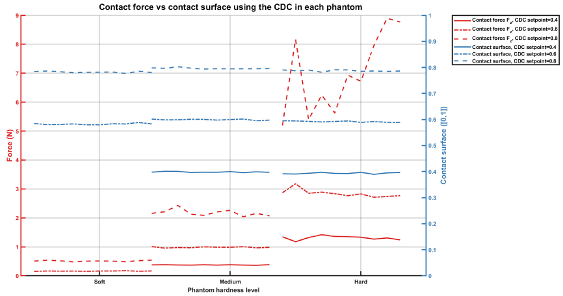


Fig. 1. Comparison of the contact force and contact surface using the CDC for the 3 setpoints in each phantom.

of the ten experiments performed with each controller for each setpoint. The CDC and hybrid controller both have in-plane normal control, while the force controller does not. The absence of in-plane correction significantly increases the barycenter error in all cases except the 80% hard, situation, which has similar values to the CDC and Hybrid controller.

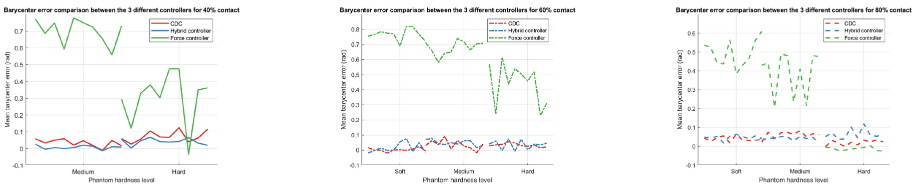


Fig. 2. Mean barycenter error for each of the 10 experiments for 40%, 60% and 80% contact in each phantom.

4 Discussion

In autonomous robotic US acquisition it is essential to ensure and optimize acoustic coupling and image quality while minimizing breast deformation. This comparative study shows three different controllers designed to satisfy the needs: a force controller, a CDC, and a hybrid force-image controller.

As observed in Fig. 1 the CDC is a more universal controller, since the setpoint to be used is independent of the phantom stiffness. On the contrary, the F_z obtained for each CDC setpoint varies depending on the phantom stiffness. These results show that the F_z needed to obtain a certain contact surface is highly dependent on the hardness of the phantom being scanned. Therefore, the force and hybrid controllers need individual setpoints for each phantom.

In Fig. 1 it is observed as well that, there is a lack of data for 40% confidence in the soft phantom. During the experiments, the CDC controller was tuned trying to simulate the results from Welleweerd et al. [4], therefore, the same controller was used for all the setpoints and phantoms. However, for 40% contact in the soft phantom, the controller was unstable and was losing contact with the tissue. An optimized tuning for each situation could make the experiments for 40% confidence in the soft phantom feasible, and could also improve all the results.

The results regarding in-plane rotation based on the barycenter of the confidence map also show great performance. While this is not a new feature and previous studies have demonstrated similar outcomes [4], this study has shown the potential of combining force feedback for z-translation control and confidence maps for in-plane rotation control. The barycenter error shows no significant difference between the CDC and Hybrid controller, while the barycenter error for the force controller, which does not control in-plane rotation is very high in most cases.

5 Conclusion

A force controller, a CDC, and a hybrid force-image controller for automatic robotic US acquisition were compared. The three controllers focus on optimizing acoustic coupling and image quality while minimizing breast deformation. The controllers were tested by scanning three different phantoms of three different stiffness. Results show that the CDC performs similarly in the three phantoms, applying a similar amount of deformation in any case using the same confidence setpoint. Therefore, the CDC has been proven to be independent of the stiffness, while the force and hybrid controllers are not. Considering that breast tissue can vary among individuals and does not have a uniform stiffness due to its complex anatomy, the CDC appears to be the most optimal controller for automatic robotic US scan of breast tissue. In addition, the results of the barycenter error show the importance of having in-plane rotation control to optimize the image quality.

References

1. Chatelain, P., Krupa, A., Navab, N.: Confidence-driven control of an ultrasound probe. IEEE Trans. Robot. **33**(6), 1410–1424 (2017). https://doi.org/10.1109/TRO.2017.2723618. ISSN 15523098
2. Huang, Q., Zeng, Z.: A review on real-time 3D ultrasound imaging technology. BioMed Res. Int. **2017** (2017). https://doi.org/10.1155/2017/6027029. ISSN 23146141
3. Vougioukas, S.: Bias estimation and gravity compensation for forcetorque sensors. In: Proceedings of International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering, pp. 82–85. Citeseer (2001)
4. Welleweerd, M.K., et al.: Automated robotic breast ultrasound acquisition using ultrasound feedback. In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 9946–9952 (2020). https://doi.org/10.1109/ICRA40945.2020.9196736. ISSN 10504729
5. Wilkinson, L., Gathani, T.: Understanding breast cancer as a global health concern 1. Br. J. Radiol. **95**, 20211033 (2021). https://doi.org/10.1259/bjr.20211033

Integration of Visual SLAM in Robot-Assisted Minimally Invasive Surgery: Advances, Challenges, and Solutions

Muzammil Khan[1,2](✉), Françoise Siepel[1], and Theo Ruers[2]

[1] University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
{f.j.siepel,m.khan}@utwente.nl

[2] The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
t.ruers@nki.nl

Abstract. Robot-assisted surgery (RAS) has demonstrated notable advancements in visualization, instrument dexterity, ergonomic improvements, and decreased infection risks when compared to conventional surgical methods. However, within minimally invasive surgery (MIS) contexts, RAS encounters notable challenges in navigating surgical tools effectively. Recent advancements in robot navigation techniques have transitioned from rudimentary wheel odometry and dead reckoning to sophisticated Visual SLAM (Simultaneous Localization and Mapping) methods, capable of addressing complex indoor and outdoor environments. Nevertheless, the integration of Visual SLAM within Robot-Assisted Minimally Invasive Surgery (RAMIS) applications remains substantially restricted due to various factors, including limited field of view, challenges in stereopsis, soft tissue deformations, insufflation effects, and instrument occlusions. This study provides an extensive overview of ongoing efforts towards the development of Visual SLAM algorithms tailored for establishing precise RAMIS systems. It delves into the encountered challenges, delineates the essential features required for establishing a precise Visual SLAM-driven RAMIS system, and explores a diverse range of approaches, which can potentially enhance Visual SLAM functionality within RAMIS contexts.

Keywords: Minimally invasive surgery · Robot-assisted surgery · Visual SLAM

1 Introduction

Robot-assisted surgery (RAS) has been introduced as a solution to address the inherent limitations encountered in laparoscopic surgery. Improved surgical dexterity is achieved through the deployment of endo-wristed instruments equipped with seven degrees of freedom [1]. However, akin to Minimally Invasive Surgery

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 399–404, 2024.
https://doi.org/10.1007/978-3-031-76428-8_74

(MIS), RAS encounters various challenges, including restricted field of view, diminished depth perception, absence of tactile feedback, instances of instrument occlusions, absence of fixed bony landmarks for orientation, and the need for external tracking of endoscope movement. These factors collectively present significant hurdles in the navigation of surgical tools during a RAMIS procedure, which is a critical requirement for establishing spatial correlation between the endoscope, patient anatomy, and surgical instruments.

Advancements in robot navigation techniques have transcended rudimentary approaches like wheel odometry and dead reckoning, evolving towards more sophisticated Visual SLAM approach. Visual SLAM allows a robot to autonomously determine its position within an unfamiliar environment without relying on external tracking systems, and simultaneously, it constructs a detailed map of its surroundings by leveraging the visual data accessible through video cameras. To enhance robot navigation performance further, Visual SLAM has been augmented with a variety of sensors beyond video cameras, which encompasses sonar sensors, infrared sensors, and LASER scanners [2]. However, the integration of Visual SLAM into RAMIS systems remains significantly constrained due to various factors, including the low-texture of tissue surfaces, challenges associated with stereopsis, endoscope size, insufflation, and occlusions.

To address the aforementioned challenges and facilitate the successful integration of Visual SLAM algorithms into RAMIS systems, researchers have primarily concentrated on incorporating the following key characteristics into developed Visual SLAM algorithm [3]:

1. Estimation of dense depth maps and comprehensive $3D$ surface reconstruction using monocular endoscope videos.
2. Endoscope pose estimation to facilitate accurate tracking.
3. Integration of dynamic surface reconstruction techniques to accommodate organ deformations occurring during the surgical procedure.
4. Implementation of surgical instrument masking within the reconstructed environment to afford the practitioner an unobstructed view, free from occlusions.

This study provides an extensive overview of ongoing efforts towards the development of Visual SLAM algorithms tailored for establishing precise RAMIS systems. It delves into the encountered challenges, delineates the essential features required for establishing a precise Visual SLAM-driven RAMIS system, and explores a diverse range of approaches, which can potentially enhance Visual SLAM functionality within RAMIS contexts.

The subsequent section discusses the advancements and obstacles pertinent to Visual SLAM algorithms in surgical environments. In Sect. 3, different promising approaches to enhance Visual SLAM functionality within RAMIS environments are delineated. Section 4 concludes the paper with future remarks.

2 Visual SLAM: Advances and Challenges

In 2017, Mahmoud et al. [4] exploited ORB-SLAM, renowned as one of the top-performing monocular SLAM algorithms, within a surgical context by extending its functionality to facilitate the reconstruction of a semi-dense map. However, the resultant $3D$ map inadequately depicts the surface characteristics of textureless soft organs like the liver. To address the challenge of sparse and semi-dense $3D$ reconstruction, Mahmoud et al. [5] introduced a real-time dense SLAM methodology designed specifically for monocular RGB laparoscope imagery. This approach operates concurrently across parallel threads and does not rely on fiducial or external tracking devices, facilitating seamless integration into surgical workflows. However, the system exhibited limited performance in handling soft tissues, attributed to its inability to effectively manage pronounced deformations during surgical manipulation. Recasens et al. [1] introduced a self-supervised monocular SLAM methodology that generates pseudo-RGBD frames and subsequently tracks camera pose using photometric residuals. Notably, this relaxes the requirement of collecting ground truth data for training their proposed model, but does not integrate any deformation model to accommodate scene deformations. Zhou et al. [6] introduced an innovative stereo laparoscopy video-based non-rigid SLAM technique termed EMDQ-SLAM, emphasizing camera motion tracking and tissue deformation estimation between consecutive video frames. This methodology leverages the expectation maximization and dual quaternion (EMDQ) algorithm in conjunction with SURF features. Furthermore, it mitigates the accumulation of errors inherent in EMDQ tracking by integrating a g2o-based graph optimization approach. Quantitative assessment reveals an average error spanning 0.8–2.2 mm across various scenarios. However, stereo endoscopes are manufactured with two lens channels installed in one standard endoscope, which causes lenses to have low diameters. It results in dramatically increased image distortion and reduced image sharpness [7]. Ozyoruk et al. [8] assert that deep learning methodologies present the potential for the development of dense topography reconstruction and pose estimation techniques applicable to endoscopic videos. They introduced Endo-SfMLearner, an unsupervised monocular depth and pose estimation methodology integrating residual networks with a spatial attention module. This architectural design aims to direct the network's focus towards discernible and highly textured tissue regions. The methodology incorporates a brightness-aware photometric loss mechanism to enhance robustness against rapid frame-to-frame illumination variations commonly encountered in endoscopic videos. Nonetheless, this approach does not address the elimination of unwanted pixels. Zha et al. [9] introduced EndoSurf, a supervised neural-field-based technique adept at learning the representation of a deforming surface from RGBD sequences. This methodology utilizes a signed distance function (SDF) field and radiance field to predict SDFs and colors for surface points, respectively. These predictions facilitate the synthesis of RGBD images through differentiable volume rendering. Furthermore, EndoSurf effectively eliminates undesired pixels, including surgical tools, blood, and smoke, from the reconstructed view. However, as a supervised learning framework, it requires substantial amounts of annotated datasets for training.

3 Enhancing Visual SLAM Functionality: Solutions

3.1 Specialized Object Detectors

Specialized object detectors refer to the popular object detectors such as YOLO-Tiny [11] that are modified to detect only a certain class of objects. Furthermore, recent works such as Crowd-SLAM [10] have demonstrated the effective integration of Visual SLAM techniques within dynamic outdoor environments by employing a specialized YOLO-Tiny module. This method is designed to facilitate the successful localization of a camera within crowded settings while simultaneously mapping the surrounding environment by excluding the crowds. This achievement is realized through the operation of four parallel threads: Object Detection, Tracking, Local Mapping, and Loop Closing. Object detection relies on Crowdhuman YOLO Tiny (CYTi), a YOLO-Tiny [11] specialization tailored for crowded environments, while the remaining three threads operate within the ORB-SLAM2 framework [12]. Therefore, in the RAMIS context, specialized object detectors can be beneficial for Visual SLAM algorithms in terms of excluding the unwanted pixels corresponding to surgical tools without the need of ground truth masks.

3.2 Scene Flow Estimation

Scene flow estimation aims to compute the $3D$ motion vectors of all points in a scene, thereby providing a detailed understanding of how objects are moving. It enables the segmentation of a scene into static (rigid) and dynamic (non-rigid) regions. Recent frameworks such as EffiScene [13] have advanced the capability to learn unsupervised scene flow estimation for intricate outdoor environments, featuring moving vehicles and clouds, thus obviating the necessity for ground truth data. EffiScene achieves this by simultaneously learning four fundamental low-level vision tasks: optical flow estimation, scene depth estimation, camera pose estimation, and motion segmentation. In a surgical environment, scene flow can aid Visual SLAM in understanding the scene dynamics including organ deformations.

3.3 Physics-Based Deformation Modelling

Achieving realism in reconstructed surfaces makes utmost physical accuracy very crucial. Various physics-based models have been proposed in literature to reconstruct $3D$ surfaces from $2D$ images, including techniques such as snakes, balloon forces, and deformable superquadrics [14]. More recently, in the realm of dynamic surface reconstruction, Thapa et al. [15] devised a deep neural network capable of real-time estimation of depth and normal maps from a monocular image sequence depicting undulating fluid surfaces. This was achieved by computing the refractive distortion of a reference background pattern beneath the fluid surface. Hence, physics-based models, such as those grounded in fluid dynamics and refraction principles, hold promise in addressing non-rigid deformation issues encountered in organs like the liver under monocular camera settings.

3.4 Structure from Motion (SfM)

SfM algorithms traditionally aim to reconstruct the $3D$ structure of a scene from an unordered collection of $2D$ images, often without real-time camera localization considerations. Ummenhofer et al. [16] innovatively formulated SfM as a learning problem, employing end-to-end trained stacked encoder-decoder networks to compute depth and camera motion from successive unconstrained image pairs. The pivotal component of the pipeline is an iterative network capable of refining its own predictions. Beyond depth and motion estimation, the network also provides estimates of surface normals, optical flow between images, and confidence levels of the matching. Additionally, a notable technical contribution of this study involves the introduction of a specialized gradient loss mechanism to address scale ambiguity in structure from motion. Consequently, SfM approaches can be promising in mitigating challenges associated with large endoscope motions and localization issues.

4 Conclusion

In this study, various state-of-the-art methodologies aimed at enhancing Visual SLAM for the establishment of precise RAMIS systems have been reviewed. Although these approaches have demonstrated notable improvements in performance, they encounter distinct challenges inherent to the surgical environment. Furthermore, different approaches have been identified, which were previously employed in outdoor and indoor settings to enhance Visual SLAM capabilities, noting their applicability for Visual SLAM in the RAMIS context. It is anticipated that these methodologies will be leveraged as promising solutions to address the aforementioned challenges in future endeavors.

References

1. Recasens, D., Lamarca, J., Fácil, J.M., Montiel, J.M.M., Civera, J.: Endo-depth-and-motion: reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. IEEE Robot. Autom. Lett. **6**(4), 7225–7232 (2021). https://doi.org/10.1109/LRA.2021.3095528
2. Merzlyakov, A., Macenski, S.: A comparison of modern general-purpose visual SLAM approaches. In: International Conference on Intelligent Robots and Systems, pp. 9190–9197 (2021). https://doi.org/10.1109/IROS51168.2021.9636615
3. Schneider, C., Allam, M., Stoyanov, D., Hawkes, D.J., Gurusamy, K., Davidson, B.R.: Performance of image guided navigation in laparoscopic liver surgery–a systematic review. Surg. Oncol. **38**, 101637 (2021). https://doi.org/10.1016/j.suronc.2021.101637
4. Mahmoud, N., et al.: ORBSLAM-based endoscope tracking and $3D$ reconstruction. In: International Workshop on Computer-Assisted and Robotic Endoscopy, pp. 72–83 (2017). https://doi.org/10.1007/978-3-319-54057-3_7
5. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE Trans. Med. Imaging **38**(1), 79–89 (2018). https://doi.org/10.1109/TMI.2018.2856109

6. Zhou, H., Jayender, J.: EMDQ-SLAM: real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy video. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 331–340 (2021). https://doi.org/10.1007/978-3-030-87202-1_32
7. Nicolau, S., Soler, L., Mutter, D., Marescaux, J.: Augmented reality in laparoscopic surgical oncology. Surg. Oncol. **20**(3), 189–201 (2011). https://doi.org/10.1016/j.suronc.2011.07.002
8. Ozyoruk, K.B., et al.: EndoSLAM dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos. Med. Image Anal. **71**, 102058 (2021). https://doi.org/10.1016/j.media.2021.102058
9. Zha, R., Cheng, X., Li, H., Harandi, M., Ge, Z.: Endosurf: neural surface reconstruction of deformable tissues with stereo endoscope videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 13–23 (2023). https://doi.org/10.1007/978-3-031-43996-4_2
10. Soares, J.C.V., Gattass, M., Meggiolaro, M.A.: Crowd-SLAM: visual SLAM towards crowded environments using object detection. J. Intell. Robot. Syst. **102**(2), 1–16 (2021). https://doi.org/10.1007/s10846-021-01414-1
11. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2021). https://doi.org/10.48550/arXiv.1804.02767
12. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an open-source slam system for monocular, stereo, and RGB-D cameras. IEEE Trans. Rob. **33**(5), 1255–1262 (2017). https://doi.org/10.1109/TRO.2017.2705103
13. Jiao, Y., Tran, T.D., Shi, G.: Effiscene: efficient per-pixel rigidity inference for unsupervised joint learning of optical flow, depth, camera pose and motion segmentation. In: International Conference on Computer Vision and Pattern Recognition, pp. 5538–5547 (2021). https://doi.org/10.48550/arXiv.2011.08332
14. Salzmann, M., Fua, P.: Deformable Surface 3*D* Reconstruction from Monocular Images. Springer (2022)
15. Thapa, S., Li, N., Ye, J.: Dynamic fluid surface reconstruction using deep neural network. In: International Conference on Computer Vision and Pattern Recognition, pp. 21–30 (2020)
16. Ummenhofer, B., et al.: Demon: depth and motion network for learning monocular stereo. In: International Conference on Computer Vision and Pattern Recognition, pp. 5038–5047 (2017)

DemoDatenPro: Methods for Semi-automatic Disassembly and Data Acquisition in Circular Economies

Michael Hofmann(✉), Matthias Propst, Markus Ikeda, and Andreas Pichler

Profactor Gmbh, Im Stadtgut D1, 4407 Steyr-Gleink, Austria
{michael.hofmann,matthias.propst,markus.ikeda, andreas.pichler}@profactor.at
http://www.profactor.at

Abstract. The central focus of the circular economy is on the technological advancement of recycling materials, devices, and machines. Notable efforts are evident in the automotive industry, where the European government mandates an 85% recycling rate for cars. In the circular economy framework, manual labor predominantly drives processes of reuse and repair. The challenges arise from the vast array of similar yet distinct products, the lack of machine-readable information, and the difficulty in automating processes related to reuse and repair. Addressing these challenges, the ongoing project DemoDatenPro aims to enhance the acquisition of disassembly information by utilizing a machine-readable "digital product passport" and showcasing the outcomes through a robotic assistant. Subsequent developments strive to refine human-robot interaction and explore the application of artificial intelligence to generalize disassembly information. This paper introduces key research questions, initial concepts, and planned research demonstrator associated with the DemoDatenPro project.

Keywords: circular economy · robotics · AI

1 Introduction and Problem Statement

The recycling of materials, devices and machines is coming into the focus of technological developments, by the circular economy. Leading attempts of the European government [1] to recycle 85% of a car are already influencing the construction of cars. Similar rules are expected for other goods. Special attention is given to the topics of:

- Usage of value-able raw materials e.g. batteries in electric powered cars [2].
- Re-usage of large machines and vehicles like airplanes, trains, wind-turbines, with a life span of 20 to 30 years [3].
- Re-circulation of fabrics [4].

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 405–409, 2024.
https://doi.org/10.1007/978-3-031-76428-8_75

– Repair and re-usage of consumer electronics [5]

The Ellen MacArthur Foundation defines the circular economy as an economy that is: *Restorative and regenerative by design, and which aims to keep products, components and materials at their highest utility and value at all times, distinguishing between technical and biological cycles.* [7]

Beside the reduction of waste dumps the preservation of the environment, the dependency of imported raw materials is considered as an issue. Many materials like lithium [8] or carbon fibers [6] are produced outside of Europe and high quality recycling could reduce the dependency form outer European manufacturers.

The circular economy is an alternative to the traditional linear economy [9]. Instead of dumping products at the end of life, it tries to keep the resources as long as possible in the life-cycle, as seen in Fig. 1.

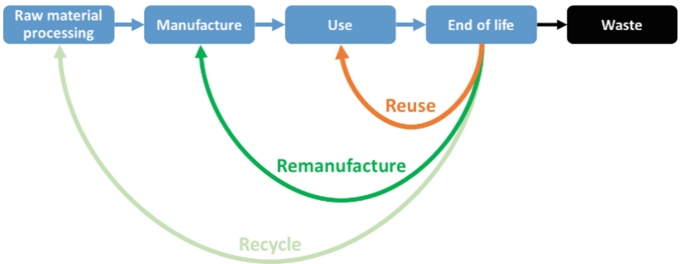


Fig. 1. Circular Economy [9]

This change also implies a change in the product design and production facilities. On the one side, the product design has to enable reuse or re-manufacturing. On the other side the production facility has to document the actual used materials and production processes. Not every product can be transformed to a circular economy.

The identification of relevant product groups and material life cycles along the process chain is therefore one of the initial problems. Further more, the manual labor is the state of the art method in the circular economy process chain. The transformation of manual labor to at least semi-automatic disassembly in a human robot collaboration is a challenging issue. In this sense, also the development of digital tools in soft- and hardware can help to increase the recycling, repair and reuse. Therefore the following three problems are targeted by the DemoDatenPro project:

Problem 1. Extraction and storage of disassembly information for different states of the product life cycle. This is a major requirement for future semi-automatic process, because accurate information about the product has to be available in a later state of the product life cycle. Disassembly information can be used by a human or as a knowledge basis for robot guided processes.

Problem 2. The amount of information available to the system. If no, or insufficient information are available about the product, the system has to generalize to specific tasks to behave autonomously. This is essential for the given situation by re-manufacturing, as a high mixture and low volume is given.

Problem 3. The interaction with humans is inevitable and therefore a smooth interaction between a semi autonomous robot and a human has to be possible. A common understanding of the work to be done, and a smooth transition between human and robotic tasks has to be accomplish.

2 Approach and Demonstrator

This section presents how the DemoDatenPro project will work on the proposed problems of the previous chapter. The demonstrator depicted in Fig. 2 is designed to work on three levels: overview, workbench and focus level. The overview-level intents to recognise human actions and ergonomics of the human, the level of detail is low and positions are not very accurate. The workbench-level is used to classify and detect objects, positions are accurate enough to guide a human or to be used is input for the robot. The focus level is mainly used for robotic processes, the high accuracy is used to improve the process quality.

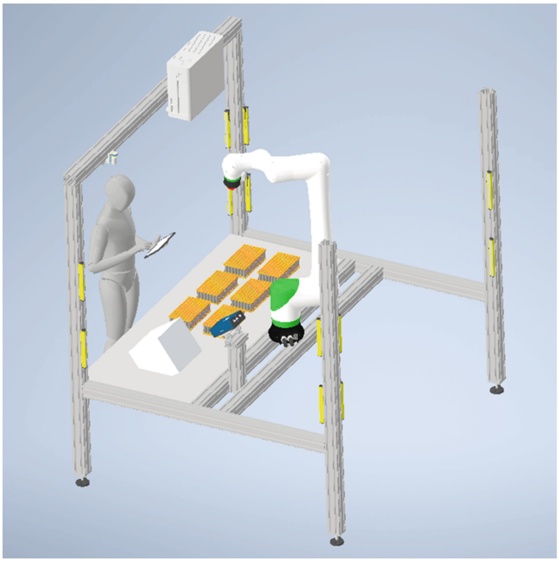


Fig. 2. DemoDatenPro demonstrator: A collaborative robot assists the disassembly of a product. Additional information can be augmented with an projector. Several sensors are collecting disassembly information.

Extraction and storage of disassembly information will be targeted by observing the human while disassembling a product. Relevant information like position and type of screws are stored in a *digital product passport*. The extraction of information is part of an continuous interaction between human and the system. Projections are used to validate anticipated actions. If the input of the human is not correct, or has to be adjusted a position tracked tool will be used to clarify the input. The disassembly information, which are extracted during this process has to be stored in appropriate data formats. If possible, data structures are used, which allow a geometric link between process data and observed actions. For logical information e.g. ontological data structures can be used. These information will forma digital-twin.

Development of methods for generalisation of disassembly information is used to process the data extracted in the previous section. These generalised information are used to complete anticipated processes while observing human actions. The generalisation will be done with deep neural networks. The problem of lack of training information will be targeted with different few-shot-learning approaches like data randomisation, data augmentation or simulation. Current research targets a few-shot-pipeline with YOLO [10] and simulation of screwing-heads to identify the correctness of screwing process in the focus-level.

Interactive usage of disassembly information in an human robot collaboration is used to demonstrate the methodological developments of the previous sections and to demonstrate a future automatisation of low volume high mixture process as seen in the circular economy. The basic automatisation principle is a skill- and workflow oriented programming of the robot. Interaction and robot processes are mapped to specific skills, which are proposed by the system to the human. The executed process are using the generalized knowledge to increase the autonomy while executing skills at the focus level.

3 Conclusion

This paper presented the DemoDatenPro project which targets three problems in the circular economy domain. These three problems are the extraction of disassembly information, the generalisation of disassembly information and the interactive usage of disassembly information. All three problems emphasize the human robot collaboration. The project will target these problems with innovative human-machine interaction concepts, driven by a digital product passport to emphasize the low volume high mixture characteristics of the circular economy. Further more a robot assistance is introduced to increase the automation during the process.

Acknowledgment. This work was funded by the project DemoDatenPro an FTI Initiative Kreislaufwirtschaft from the Upper Austrian Government.

The research has been funded by the Government of Upper Austria through the program EFRE, the DemoDatenPro.

References

1. EU Directive 2000/53/EC
2. Natkunarajah, N., Scharf, M., Scharf, P.: Scenarios for the return of lithium-ion batteries out of electric cars for recycling. Procedia CIRP **29**, 740–745 (2015). ISSN 2212-8271
3. Sealy, C.: "New lease of life for carbon fiber reinforced plastics" reinforced-plastics. https://www.reinforcedplastics.com/content/features/new-lease-of-life-for-carbon-fiber-reinforced-plastics. Accessed 22 Jan 2024
4. M-Brain GmbH. The European market potential for recycled fashion. https://www.cbi.eu/market-information/apparel/recycled-fashion/market-potential. Accessed 22 Jan 2024
5. Bracquené, E., et al.: Quantifying the environmental impact of clustering strategies in waste management: a case study for plastic recycling from large household appliances. Waste Manag. **126**, 497–507 (2021). ISSN 0956-053X
6. Bhutada, G.: "Charted: Lithium Production by Country (1995–2020)" visualcapitalist. https://www.visualcapitalist.com/charted-lithium-production-by-country-1995-2020/. Accessed 22 Jan 2024
7. Ellen MacArthur Foundation. Circular economy introduction. http://www.ellenmacarthurfoundation.org/circular-economy. Accessed 22 Jan 2024
8. Carbon fibre 2020 (Knoxville, Tenn, US) Preconference seminar by Tony Roberts, AJR Consultancy
9. Remanufacturing as best practice of the Circular Economy A Position Paper of the Automotive Parts Remanufacturer Association - Europe
10. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475 (2023)

A Flexible Robotic-Based Architecture for Cyber-Physical Sorting Systems in Waste Management Industry

Konstantinos Kokkalis(✉), Fotios K. Konstantinidis, Georgios Tsimiklis, and Angelos Amditis

Institute of Communication and Computer Systems (ICCS), National Technical University of Athens, 9 Iroon. Polytechniou Str., Zografou, GR 15773 Athens, Greece
konstantinos.kokkalis@iccs.gr

Abstract. The pressing necessity for highly efficient waste segregation effectively leads to utilization of automated sorting systems capable of high throughput and great adaptability to highly mixed streams. This works proposes an architecture for a Cyber-Physical Sorting System (CPSS) relying on robotic actuation. Its modular design aims to be adjustable to different waste stream requirements with minimum effort by supporting multi-modal sensing, manipulation of diverse materials and geometries, as well as integration of different industrial robots.

Keywords: Waste sorting · Industrial robotics · Sustainable robotics

1 Introduction

As the waste generation problem has intensified in the last decades, new initiatives have constantly been adopted by the EU aiming to embrace circularity and circular economy. EU's new Circular Economy Action Plan [1] has introduced a new agenda focusing among others on the design and production of high quality products able to be reused, repaired and recycled, paying special attention to key value chains such as electronics, batteries, textiles and plastics. However, recovering wasted materials remains an acute challenge. due to streams' heterogeneity, which impedes the application of many such techniques and necessitates high-performance segregation [2]. Traditional automatic sorting systems rely on a narrow set of physical properties such as ferromagnetism and cannot efficiently classify blended materials of the aforementioned value chains, while human labor, currently employed in many stages of sorting processes, remains expensive, hazardous and difficult to secure. In consequence, robotic material sorting systems have gained significant ground with multiple commercial products coming into market [3,4]. Due to their simplicity most of these systems utilize only visual sensory equipment e.g. RGB cameras and machine learning techniques to segment and classify the incoming streams [5]. This restriction to single-modal sensing limits the system's capability to separate highly mixed systems. Furthermore,

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 410–415, 2024.
https://doi.org/10.1007/978-3-031-76428-8_76

research on task and motion planning for moving objects sorting has focused on a limited range of commercial robots with relatively few degrees of freedom [6], on simple manipulation strategies [7] and simple geometries. This lack of universal solutions to handle diverse streams has motivated this work, which aims to introduce an architecture of a generic robotic sorting system that can integrate a variety of sensory inputs and planning schemes based on the needs of each sorting application.

2 Overall Architecture

The need for intricate sensing, planning, control and actuation posed by the sorting process naturally leads to a cyber-physical system with multiple components constantly exchanging information. In Fig. 1 the architecture of the system is presented along with the interactions of its main components.

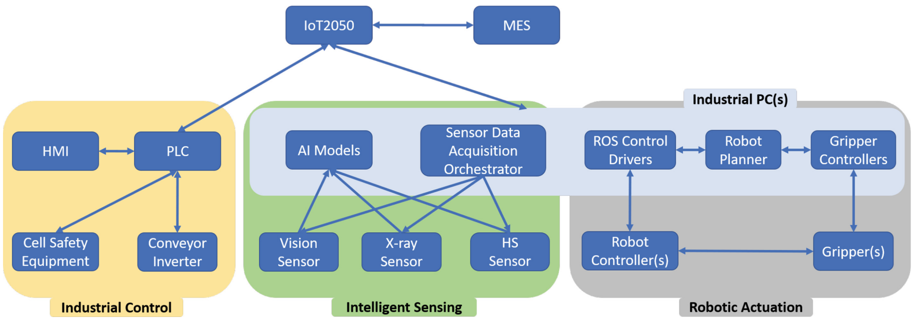


Fig. 1. The proposed architecture of the Cyber-Physical Sorting System

Following a bottom-up analysis of the figure, hardware devices related to sensing and actuation can be noted on the lowest level. One level above, components responsible for planning and processing of input data are located i.e. a single (or a series of) industrial PC, a Programmable Logic Controller (PLC) and a Human Machine Interface (HMI). They incorporate all the intelligence of the system, and the control of the aforementioned low-level devices relies on them. On the third level a IoT2050 gateway acts as a "mediator" supporting the interchange of data among the PLC and the industrial PCs [8].

As part of larger industrial processes, Cyber-Physical Sorting Systems are required to be constantly monitored and tracked in order to assist decision making in a higher level [9]. Furthermore, they should be able to coexist with other CPSs already installed in the facilities. Thus, a link with an existing Manufacturing Execution System (MES) should be possible utilizing the IoT2050 gateway and sharing essential data of the components with the lower levels. The architecture of the system is designed to achieve primarily modularity and flexibility. The devices on the bottom level should be easily changeable with others when a different waste stream is considered. For example, a delta robot with its simple 4-DoF kinematics may be sufficient for a process, but if more complex end-effector poses are required, a serial robot with 6 DoF would be necessary. Thus, the second level of the architecture should be able to cope with hardware replacements with minimum changes.

3 Subsystems

A vertical division of the architecture, on the other hand, would result in 3 different subsystems: a) Industrial Control, b) Intelligent Sensing and c) Robotic Actuation. In this Section each of these subsystems will be examined in detail.

3.1 Industrial Control Subsystem

As any typical industrial automated system, a sorting system requires the utilization of multiple hardware components designed for optimal efficiency and precision, yet limited intelligence. These components are tasked to perform critical functions of the overall system and comprise the Industrial Control Subsystem. Examples of low-level components of this subsystem are safety-critical equipment such as emergency push-buttons, safety interlock switches and area scanners. These simple devices are capable of controlling the basic operation of the entire system in case of abnormalities with high reliability and fast response.

The control of the conveying mechanism largely affects the efficiency of the overall system, since inaccuracies will lead to poor synchronization with the robotic system resulting in inferior performance. Therefore, precise control of the corresponding motor inverter is essential to the system. The PLC processes the data from the low-level components in real-time and ensures safe operation according to industrial standards. Furthermore, commands from the HMI are passed to the PLC, while real-time data from the low-level components are exchanged with the rest of the subsystems and the MES through IoT2050.

3.2 Intelligent Sensing Subsystem

Starting from the low-level devices of the subsystem, there is a variety of sensors to be chosen with different working principles in different regions of the electromagnetic spectrum. Their selection should be based on the materials composing the stream to be segregated. The sensory output is fed to AI models aiming to achieve three objectives: image segmentation, material classification and object localization of the moving objects on the conveyor. For segmentation, deep learning models trained only on vision features have been proved adequate. However, material classification using only RGB cameras, especially in dirty and dusty conditions, remains challenging. For these cases, Near Infrared (NIR), Short-Waved Infrared (SWIR), X-Ray sensors and others can provide useful features to be exploited by multi-model material classifiers. Many present-day sensors are capable of acquiring data at a high rate, while others require longer acquisition times. Thus, synchronization among sensors necessitates special care. Furthermore, running complex ML techniques can be computationally expensive and continuous execution may result in longer inference times. To balance and control the collection of data from the sensors a Sensor Data Acquisition Orchestrator should be employed, as described in [10,11].

3.3 Robotic Actuation Subsystem

As the material flows through the sorting system the last components involved are the actual robots, responsible for picking and placing the objects. The motors of a single robot are commanded by an industrial controller running a real-time OS. Through different communication protocols (usually TCP or UDP), it is possible to receive setpoints in the joint space in order to execute a relatively smooth trajectory. The control of the grippers, on the other hand, may be performed from the robot controller on which they are mounted or from a separate driver running on the industrial PC [8]. Ensuring the demanded modularity of the architecture requires the development of robot-agnostic planning and control pipelines. The ROS2 framework and libraries encourage these flexible schemes, where the computational load is distributed among different pieces of software. Thus, a task and motion planning pipeline independent of the robot model is created based on ROS2 and MoveIt2! stack [12]. Although most components of the subsystem are aiming to be robot agnostic, the demand to create hardware-specific drivers for robots and grippers cannot be eliminated. Therefore, the ROS Control Driver and Gripper Controller modules are robot-specific, and their purpose is to receive commands from the Robot Planner using a standard API based on ROS2 and translate it to setpoints for the corresponding hardware. Finally, these modules can provide feedback on the state of the low-level devices to close the control loop.

3.4 Information Flow Between Subsystems

Task and motion planning of the robots (performed by the Robot Planner module) requires multiple sources of data coming from the Industrial Control and Intelligent Sensing subsystems as shown in Fig. 2. More precisely, the generation of the grasping poses for each object moving along the conveyor relies on the geometric characteristics as captured by the sensor set and the inference of the AI models.

The grasping poses and the ID gripper to be used are then passed to the Trajectory Planner submodule along with conveyor speed and the object material as already classified. This submodule aims to come up with feasible, collision-free and optimal trajectories that would pick the object and then place it to the appropriate bin. Finally, these trajectories for multiple conveyed objects are given as input to the Task Planner, which determines at every cycle which object should be picked based on some optimization criteria or simple scheduling rules. In a multi-robot system, the module also assigns the object to a specific robot.

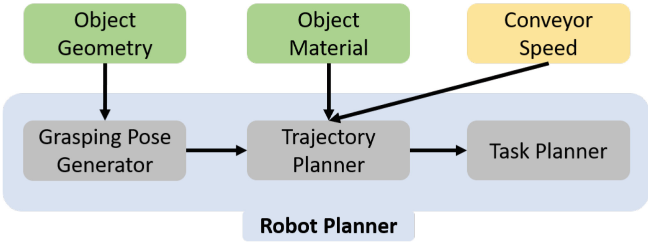


Fig. 2. Data flow for task and motion planning

4 Summary and Future Work

This work introduced a modular and flexible architecture for Cyber-Physical Sorting Systems. Its essential components were analyzed based on a layered approach, while each of them can be attributed to one of the three existing subsystems: Industrial Control, Intelligent Sensing and Robotic Actuation. Finally, the information flow to perform the separation was also presented. Future work will compare multiple implementations of this architecture for different value chains and streams.

Acknowledgements. This research was financially supported by the European Union's Horizon Europe research and innovation program under grant agreement No. 101138789 (project W2W).

References

1. European Commission. A new Circular Economy Action Plan (2020)
2. Konstantinidis, F.K., Myrillas, N., Mouroutsos, S.G., Koulouriotis, D., Gasteratos, A.: Assessment of industry 4.0 for modern manufacturing ecosystem: a systematic survey of surveys. Machines **10**(9), 746 (2022)
3. Lukka, T.J., Tossavainen, T., Kujala, J.V., Raiko, T.: Zenrobotics recycler - robotic sorting using machine learning (2014)
4. TOMRA: FLYING BEAM™. https://www.tomra.com/en/waste-metal-recycling/products/technologies/flying-beam. Accessed 18 Dec 2023
5. Kiyokawa, T., Takamatsu, J., Koyanaka, S.: Challenges for future robotic sorters of mixed industrial waste: a survey. IEEE Trans. Autom. Sci. Eng. **21**, 1023–1040 (2022)
6. Ku, Y.D., Yang, J.H., Fang, H.Y., Xiao, W., Zhuang, J.T.: Optimization of grasping efficiency of a robot used for sorting con. and demolition waste. Int. J. Autom. Comput. **17**, 691–700 (2020)
7. Mattone, R., Divona, M., Wolf, A.: Sorting of items on a moving conveyor belt. Part 2: performance evaluation and optimization of pick-and-place operations. Robot. Comput. Integr. Manuf. **16**(2), 81–90 (2000)
8. Balakera, N., Konstantinidis, F.K., Tsimiklis, G., Latsa, E., Amditis, A.: IIoT network system from data collection to cyber-physical system transmission under the industry 5.0 era. In: International Congress on Information and Communication Technology, pp. 929–941. Springer (2023)

9. Konstantinidis, F.K., Sifnaios, S., Tsimiklis, G., Mouroutsos, S.G., Amditis, A., Gasteratos, A.: Multi-sensor cyber-physical sorting system (CPSS) based on industry 4.0 principles: a multi-functional approach. Procedia Comput. Sci. **217**, 227–237 (2023)
10. Konstantinidis, F.K., et al.: Multi-modal sorting in plastic and wood waste streams. Resour. Conserv. Recycl. **199**, 107244 (2023)
11. Sifnaios, S., Zorzos, I., Arvanitakis, G., Konstantinidis, F.K., Tsimiklis, G., Amditis, A.: Exploration and mitigation of the impact of lighting conditions on multi-spectral image classification. In: International Conference on Imaging Systems and Techniques (IST), pp. 1–6. IEEE (2023)
12. Coleman, D., Sucan, I., Chitta, S., Correll, N.: Reducing the barrier to entry of complex robotic software: a MoveIt! case. CoRR (2014)

Check for updates

Towards Measuring the Ease of Robotic Disassembly

Christoffer Sloth(✉) and Iñigo Iturrate

University of Southern Denmark, Odense, Denmark
{chsl,inju}@mmmi.sdu.dk

Abstract. This paper addresses the use of robotics and AI for recycling and remanufacturing low-value components. To make the robotic disassembly of low-value components viable, such components should be easy to disassemble. This initial work provides simple measures associated to ease of disassembly by robots that may be used for redesigning components. The disassembly consists of three phases: Identification, extraction, and disassembly. The paper focuses on the extraction phase of the disassembly.

Keywords: disassembly · robotics · e-waste

1 Introduction

The focus on recycling and remanufacturing is increasing, as current practice is unsustainable [7,11]; as an example only 17.4% of the documented e-waste was recycled in 2019 [4]. To address this issue, regulations such as the WEEE Directive and the RoHS Directive are being implemented to increase recycling of e-waste. It is, however, difficult to implement technological solutions for completing the recycling although design for assembly and disassembly has been considered for a long time [2]. A few successful automatic disassembly systems have been implemented; however, these are all manufacturer-centric. According to [1], consumer electronic products tend to be complex and difficult to disassemble; consequently they are not recycled or remanufactured. Subcomponents will often have a much longer operational life than the product life [3]. Therefore, the potential for recycling is large, but to exploit the potential, automated disassembly strategies should be developed for distributed disassembly of general products rather than only manufacturer-centric disassembly. Robotics and AI play a crucial role in exploiting the potential, where the challenges of distributed disassembly are similar to the transition to low-volume customised production.

2 Problem Formulation

To emphasize the scope of the research, we take an outset in the reuse of HAN connectors. HAN connectors can be embedded into various products; hence, the

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 416–420, 2024.
https://doi.org/10.1007/978-3-031-76428-8_77

connectors should not only be easy to disassemble, they should also be easy to identify and extract from any product. Connectors are low-value components; hence, the introduction of specialized tools is not viable for performing the disassembly. Therefore, the connector should be designed to be easily identified, extracted, and disassembled by a standard robot with generic tooling.

To redesign products to fulfill the above criteria, one needs to quantify the ease of robotic disassembly, i.e., how easy it is for a robotic system to identify, extract, and disassemble the components of a given product.

3 Ease of Disassembly

This section presents measures related to ease of disassembly of a component within an arbitrary product. As illustrated in Fig. 1, the disassembly of a component of a product is comprised of the following three phases: 1) Identifying a component in the product, 2) Extracting the component from the product, 3) Disassembling the component.

Fig. 1. Overview of phases in the disassembly of a component in a product.

The following subsections address the three phases, with an emphasis on the extraction.

3.1 Identification of Component in Product

To identify an object in a product consisting of a multitude of components is challenging. It requires that the component has distinctive features that can be seen from any admissible viewpoint of the component. It may be only necessary to determine the class of object to enable extraction; otherwise, the metric presented in [5] may be used for the identification. Current methodology typically considers datasets such as COCO [10] when developing detection and segmentation algorithms.

3.2 Extraction of Component from Product

The quantification of ease of extracting a component from a product involves several measures. First, the pose of the component should be determined, then the component should be unmated and grasped, and finally the component should be extracted.

Pose Estimation. It is nontrivial to characterize objects that are easy to pose estimate, as it highly depends on the applied pose estimation algorithm. General computer vision guidelines list rules of thumb e.g. objects should have texture, objects should not be shiny, and the object geometry should be simple. In disassembly, consideration should also be taken in relation to degradation of texture over time, and change of geometry due to wear. If a particular pose estimation framework is considered then it may be possible to compute a distribution of pose estimates based on a given viewpoint [6]. For the considered pose estimation, the viewpoint is not known; thus, the distribution of pose estimated should be computed from a subset of possible viewpoints. Finally, equivalence classes of poses given by symmetries of the component should be established. Current methodology typically considers datasets such as BOP [8] when developing pose estimation algorithms.

Unmating. The method used for mating components together can significantly affect the ease of disassembly. As an example, unscrewing may be easier for a standard robot than opening a snap-fit. The difficulty of unmating relates to the required dexterity and space, number of contact points and forces. Finally, disassembly should be possible with standard tools to allow distributed disassembly.

Grasping. The quality of a grasp can be determined from the contact points on an object and the model of the contact [12], and a grasp can be represented by a grasp matrix G. Since the component is integrated into a product, the area of the component that is free for grasping is not known a-priori. Consequently, a set of admissible grasps is defined as a set of grasp matrices $\mathcal{G}$. One grasp quality metric is given by the smallest singular value of a grasp matrix [9]. To design a component that is easy to grasp inside a product, it should be easy to grasp from any admissible grasp; hence, we define the following grasp quality measure:

$$Q_{\text{grasp}} = \min_{G \in \mathcal{G}} \sigma_{\min}(G), \tag{1}$$

where $\sigma_{\min}(G)$ denotes the smallest singular value of G. The grasp quality is good when Q_{grasp} is large.

Extraction. The ease of extraction is related to the required dexterity. In [13] an index of difficulty is introduced based on the required resolution of position and force along a given trajectory. The combination of the two measures can be used for quantifying the ease of extraction:

$$Q_{\text{extract}} = \sum_i \log_2(w_i) + \sum_i \log_2(f_i), \tag{2}$$

where w_i is the normalized resolution of position and f_i is the normalized minimum applied force along the ith axis. The measure from [13] is rewritten such that the extraction is easy when Q_{extract} is large.

3.3 Disassemble Component

The ease of the component disassembly can be accomplished with measures given in [2]. This process highly depends on whether a distributed disassembly system should be able to perform the component disassembly or it should be accomplished by a centralized disassembly system.

4 Some Open Questions

Although we have attempted to present some of the challenges in the different phases of the disassembly and characterize them with a metric that measures their difficulty, many open questions remain. Here, we make some considerations as to future questions that will impact robotic disassembly:

1. *Can the design rationale behind existing benchmarks, e.g., for object detection, segmentation, and pose estimation, be formalized in a way that measure ease of disassembly?*
   The computer vision research community currently evaluates many of their findings on open datasets or benchmarks [8,10]. However, the criteria for the choice and difficulty of test cases in these is not formalized. Characterizing this may give valuable insights for design for disassembly.
2. *Which information from the product design pipeline can be passed on to the robotics system designers to ease disassembly automation?*
   As an example, knowledge of the location of distinguishing object features in the product design could help in the design of robot systems by, e.g., optimizing camera viewpoint and reducing ambiguity.
3. *To what extent should robot technologies be developed to better handle current product designs vs. products be redesigned to better fit the current state of the art of robotic automation?*

5 Conclusion

In this paper, we presented some metrics that can be used for assessing the ease of disassembly of components. The redesign of components to enable easy disassembly is pivotal for increasing the recycling of e.g. e-waste. More work is needed to guide the component redesign such that robotics and AI can contribute even more to generate a sustainable society.

References

1. Bogue, R.: Robots in recycling and disassembly. Ind. Robot Int. J. Robot. Res. Appl. **46**(4), 461–466 (2019)
2. Boothroyd, G., Alting, L.: Design for assembly and disassembly. CIRP Ann. **41**(2), 625–636 (1992)

3. Foo, G., Kara, S., Pagnucco, M.: Challenges of robotic disassembly in practice. Procedia CIRP **105**, 513–518 (2022)
4. Forti, V., Balde, C.P., Kuehr, R., Bel, G.: The global e-waste monitor 2020: quantities, flows and the circular economy potential (2020)
5. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R.: Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recogn. **51**, 481–491 (2016). https://doi.org/10.1016/j.patcog.2015.09.023
6. Haugaard, R.L., Hagelskjær, F., Iversen, T.M.: SpyroPose: SE(3) pyramids for object pose distribution estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2082–2091 (2023)
7. Hjorth, S., Chrysostomou, D.: Human-robot collaboration in industrial environments: a literature review on non-destructive disassembly. Robot. Comput. Integr. Manuf. **73**, 102208 (2022)
8. Hodan, T., et al.: BOP: benchmark for 6D object pose estimation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 19–34 (2018)
9. Kim, B.H., Oh, S.R., Yi, B.J., Suh, I.H.: Optimal grasping based on non-dimensionalized performance indices. In: Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180), vol. 2, pp. 949–956. IEEE (2001)
10. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
11. Poschmann, H., Brueggemann, H., Goldmann, D.: Disassembly 4.0: a review on using robotics in disassembly tasks as a way of automation. Chem. Ing. Tech. **92**(4), 341–359 (2020)
12. Roa, M.A., Suárez, R.: Grasp quality measures: review and performance. Auton. Robot. **38**, 65–88 (2015)
13. Sturges, R., Wright, P.: A quantification of dexterity. Robot. Comput. Integr. Manuf. **6**(1), 3–14 (1989). https://doi.org/10.1016/0736-5845(89)90080-X

Robotic Ease of Disassembly Metric (Re-DiM) for Flexible Cooperative Remanufacturing of Bike Batteries

Terrin Pulikottil[1,2(✉)], Wouter Sterkens[1,2], Mathijs Piessens[1,2], and Jef R. Peeters[1,2]

[1] Department of Mechanical Engineering, KU Leuven, Leuven, Belgium
terrin.pulikottil@kuleuven.be
[2] Flanders Make@KU Leuven, Leuven, Belgium

Abstract. The success and economic viability of remanufacturing is intrinsically linked to the efficiency of the product's disassembly. For complex products such as bike batteries, the synergy of human and robot cooperative disassembly presents a flexible solution that integrates the benefits of both manual and robotic approaches. To assess the barriers that hinder the adoption of such flexible robotic systems for remanufacturing, a Robotic ease of Disassembly Metric (Re-DiM) is developed in the presented research. This method is applied to evaluate the technical feasibility and time efficiency for a flexible robotic disassembly system for diverse bike batteries. Based on the specifications and limitations of the cooperative disassembly system and necessary toolset under development, various criteria are defined that connectors must meet for effective robotic disassembly. Based hereon and the disassembly time estimations, the challenges of robotic disassembly are underscored for bike battery disassembly.

Keywords: Remanufacturing · Robotic disassembly · Disassembly metric · Human-robot cooperation

1 Introduction

The European Commission's circular economy action plan emphasizes the pivotal role of remanufacturing practices for End-of-First-Life products like mobility batteries [1]. Unlike traditional manufacturing, remanufacturing aligns seamlessly with the objectives of a circular economy by focusing on not just economic benefits but also ecological preservation by maximizing resource use and extending the product lifetime [2]. In remanufacturing, disassembly is pivotal but also highly challenging when compared to assembly. This is due to varying product conditions, orientations, missing parts and/or product information and smaller lot sizes as different generations of products often return over a longer period of time. Despite manual disassembly offering high flexibility to handle complex tasks and to deal with uncertainties, several drawbacks are to be considered: high operational cost, limited scalability, regional worker scarcity, training and knowledge transfer, safety and health risks. While robotic disassembly could alleviate

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 421–426, 2024.
https://doi.org/10.1007/978-3-031-76428-8_78

these issues, it faces various challenges, such as high capital costs and limited adaptability for intricate tasks, as dedicated tooling designs are considered unfeasible when jointly remanufacturing different product generations. Therefore, human-robot systems offer a balance between efficiency, flexibility, expenses (operational and capital) and reduces health and safety risks. Similar to a Flexible Manufacturing System (FMS), a human-robot disassembly system has the potential to collaborate with a variety of similar products to offer a solution for a Flexible Re-Manufacturing System (Re-FMS). These systems are also known to cope better with unpredictability in task frequency, return volume, and product model variations. Despite various attempts in recent years by researchers in disassembling critical products like mobility batteries [3–6], there is still high uncertainty of how such a flexible disassembly system should look like. In prior research, solutions often solely focus on only a part of the disassembly process e.g., only un-fastening of screws [3, 4] or gripping of components for removal [5, 6]. Nonetheless, in the shift towards Industry 5.0 with a focus on sustainable and human-centric production systems, where disassembly tasks involve orchestration between humans and robots, the development of fully integrated remanufacturing systems becomes of increasing importance.

For such a flexible disassembly system, a metric to assess the barriers and to evaluate the ease of both manual and robotic disassembly plays a critical role for remanufacturers in process development, task allocation and scheduling activities. While the authors have in prior research recommended disassembly metrics for manual disassembly, such as e-DiM metric [7], there exists currently no metric that encompasses both manual and robotic disassembly. Therefore, this research extends the scope of the existing e-DiM by introducing a Robotic ease of Disassembly Metric (Re-DiM) that considers both human and robotic disassembly operations for flexible bike battery disassembly.

2 Materials and Methods

Material: Nine distinct bike batteries (Fig. 1-b) from various manufacturers were evaluated, featuring diverse designs, including tube casings, snap fits, and adhesives: Bosch PowerPack 400 & 500, Bosch PowerTube 625, Flyer 36 V STB Derby Cycle, Gazelle Impulse, Gazelle Bronze, Schimano BT-E6000, TranzX BL-03, and JD-PST. For these batteries, the disassembly sequence during remanufacturing includes opening the casing, extracting the Battery Management System (BMS), detaching nickel strips from Li-ion cells, removing cells from the holder, and optionally cleaning the cell surface, to then replace the cells and reverse the process. This paper solely concentrates on the disassembly operation, omitting any discussion or analysis of the reassembly process.

Flexible Disassembly Cell: The research platform under development represents an efficient and affordable Re-FMS for a diverse range of products (Fig. 1-a). This cell is equipped with a 6-dof robot with cooperative features (like laser scanner), as it is expected to be industrialized in the coming decennia for the EoFL treatment. The preference for a cooperative cell over a collaborative one arises from its advantages in terms of speed, precision, payload capacity, and human safety while using unsafe tool. The system also includes conveyor with human working station, positioning and clamping fixtures and a tool changer with a set of tools including a 3D vision camera. Depending on the product family, the cell can be further extended with additional and dedicated stations for specialized operations like milling, stamping, press fit removal etc.

For a bike battery disassembly, the envisaged toolset includes: (i) A screwdriver with interchangeable commonly available bits: Phillips, torx, hex and slotted (for un-fastening task) (ii) A two-finger electric gripper with a min. Stroke length of 140 mm and a max. Finger width of 10 mm (for manipulation tasks) (iii) A pneumatic nipper with max. Cutting pressure of 500 N and min. Blade height of 10 mm (for metal wire cutting). As busbar removal is considered most feasible by milling operations, a specialized milling station is presumed. The toolset list and its dimensions are defined on what is most realistic, but they are to be further defined, for example based on consultation by re-manufacturers or standardization organization.

Robotic Ease of Disassembly Metric (Re-DiM): Re-DiM is a time metric which uses as input the connector specifications and calculates based on disassembly parameters the estimated operational time (Fig. 2). Similar to eDIM, the Re-DiM metric adopts a user-friendly spreadsheet format to ensure ease and seamless compatibility, without requiring any additional software installation. In this spreadsheet each row corresponds to a specific connector, and their sequence dictates the evaluated disassembly order.

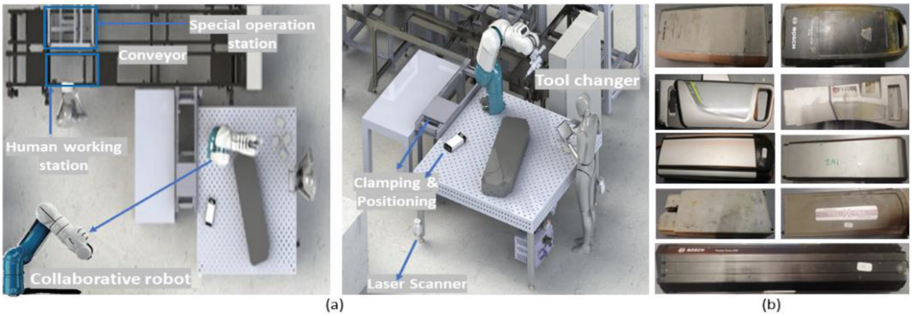


Fig. 1. (a) Flexible disassembly cell (b) 9 Bike batteries (from top left-clockwise: Bosch 400 & 500, Gazelle Impulse, Schimano BT, TranzX JD, Bosch 625, TranzX BL, Gazelle Bronze and Flyer 36 V)

Disassembly Information: It consist of four key sections and should be provided by the product designers. First, the Product Description outlines component and connector details. Manual disassembly description detailing connector visibility, tools, positioning, manipulation, and component removal. Robot disassembly criteria assesses connector's robotic disassembly feasibility. Finally, the Robot disassembly description elaborates on robotic disassembly specifics, including task types, tools used, positioning, fixture manipulation, grasping type, and removal methods.

Disassembly Parameters: It includes disassembly time parameters and robotic disassembly criteria parameters. Manual disassembly time parameters are calculated using the MOST technique [7]. For robotic disassembly parameters, calculation incorporates a 50% maximum robot speed assumption based on a robot's maximum cartesian speed

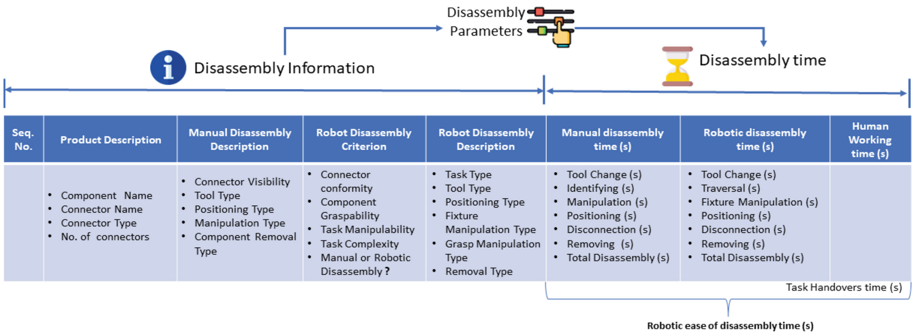


Fig. 2. Robotic ease of disassembly metric

of 10 m/s, accounting for trapezoidal motion profiles and short distances during disassembly. Fixed action times of 5 s are assumed for tool changing, clamping, de-clamping, grasping, wire cutting and precise positioning and 10 s for task handover. For screw disconnection, the thread length, pitch, and a screwdriver speed of 600 rpm are considered. Busbar removal time by milling station is considered as 3 s/joint. Robotic disassembly criteria parameters is determined considering the limitations of the flexible disassembly cell in terms of the robot and tool design like connector standard design, graspable component, gripper's ability to access the object, task feasibility in achieving using a single robotic arm.

Disassembly Time: Manual (M_t) and robotic (R_t) disassembly time is calculated by summing up tool change, identification/traversal, manipulation, positioning, disconnection, and removal time. When a task can be performed by either a robot or a human during disassembly, the metric attributes the task to the robot and calculates the remaining tasks as human working time (H_t). By counting the task handovers between the human and robot the task handover time (C_t) is calculated. Finally, the robotic ease of disassembly (Re-DiM) time is calculated as,

$$\text{robotic ease of disassembly time} = min\left[M_t, \left(\frac{R_t + C_t}{n} + H_t\right)\right]$$

Here, n is the cost factor that depends on parameters like cost for robot usage, robot availability, labor cost etc. (e.g., n = 2, assuming robot has half the labor cost).

3 Results and Discussion

Criteria assessment highlights the connector type that are challenging for robotic disassembly with the proposed flexible disassembly cell. Cable plugs present difficulties during disassembly due to their variable positions, snap fits and PCB connectors pose challenges with limited clearance for gripper manipulation, and along with friction fits these connectors (designed for human hand) require more than two-finger operations. In analyzing the re-diM results (Table 1) of the 9 bike batteries, it is evident that the designs with casings connected by screws (Bosch and Shimano) aided in a shorter disassembly time in manual and higher automation in robotic disassembly compared to other designs, such as the tube design (Bosch PowerTube and Gazelle Innergy Bronze) and those connected by snapfits (Flyer Derby Cycle) or adhesives (Gazelle Impulse, Gazelle Innergy Bronze, TranzX BL-03). Tube design, snapfits and adhesives increases the time for manual disassembly and is mostly impossible to be disassembled robotically with the proposed flexible disassembly cell (without a dedicated tooling). Busbar removal is considered the most time consuming task for manual disassembly and due to the need for complex removal process, high precision (nickel busbar thickness range between 0.15–0.2 mm) and safety concerns a dedicated station is presumed to be required. Overall, the significant differences in disassembly times stresses also the impact and need for a better alignment of product and process design to assure the viability of robotic re- and demanufacturing.

Table 1. Re-DiM disassembly time results for 9 bike batteries

Bike Battery Model	No. of Li-ion Cells	Manual disassembly time M_t (s)	*Robotic disassembly* time R_t (s)	*Human Working time* H_t (s)	*Task Changeover time* C_t (s)	*Re-DiM* time (s)
Bosch powerpack 400	40	1193	445	148	40	391
Bosch powerpack 500	40	1202	329	347	60	541
Bosch powerpack 625	50	1295	589	165	90	504
Flyer 36 V STB	50	1584	392	546	60	772
Gazelle Impulse	50	1460	312	359	60	664
Gazelle Bronze	52	1532	437	671	70	924
Shimano BT E6000	40	1097	368	228	30	427
Tranzx BL-03 (pouch)	6	503	166	428	50	503
Tranzx JD-PST	52	1248	480	395	60	665

Conclusion: The paper introduces a flexible disassembly system and the Re-DiM disassembly time metric designed for cooperative human-robot disassembly. Illustrated through a case study on bike battery disassembly, the paper emphasizes the challenges associated with robotic disassembly for the remanufacturing of these batteries for cell replacement. Future research directions includes reassessment of disassembly limitations and parameters based on novel tool designs, the integration of enhanced and adaptive robot control linked with a digital product passport based on empiric experiments.

Acknowledgement. This work is carried out under SCANDERE project granted from the ERA-MIN3 program under Grand No. 101003575 and CaDaNs project funded by Vlaanderen Circulair.

References

1. European Commission: A new Circular Economy Action Plan (2020)
2. Huang, J., et al.: A case study in human–robot collaboration in the disassembly of press-fitted components. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture **234**(3), 654–664 (2020). https://doi.org/10.1177/0954405419883060
3. Gerbers, R., Wegener, K., Dietrich, F., Dröder, K.: Safe, Flexible and Productive Human-Robot-Collaboration for Disassembly of Lithium-Ion Batteries. In: Recycling of Lithium-Ion Batteries: The LithoRec Way, pp. 99–126. Springer International Publishing, Cham (2018). https://doi.org/10.1007/9783319705796
4. Zhang, H., Zhang, Y., Wang, Z., Zhang, S., Li, H., Chen, M.: A novel knowledge-driven flexible human–robot hybrid disassembly line and its key technologies for electric vehicle batteries. J. Manuf. Syst. **68**, 338–353 (2023). https://doi.org/10.1016/j.jmsy.2023.04.005
5. Qu, W., Li, J., Zhang, R., Liu, S., Bao, J.: Adaptive planning of human–robot collaborative disassembly for end-of-life lithium-ion batteries based on digital twin. J. Intell. Manuf. (2023). https://doi.org/10.1007/s10845-023-02081-9
6. Kay, I., Farhad, S., Mahajan, A., Esmaeeli, R., Hashemi, S.R.: Robotic disassembly of electric vehicles' battery modules for recycling. Energies **15**(13), Art. no. 13 (2022). https://doi.org/10.3390/en15134856
7. Vanegas, P., et al.: Ease of disassembly of products to support circular economy strategies. Resour. Conserv. Recycl. **135**, 323–334 (2018). https://doi.org/10.1016/j.resconrec.2017.06.02

Cognitive and Robotic Assistance to Increase Efficiency in Li-Ion Battery Re-manufacturing

Matthias Propst(✉), Michael Hofmann, Markus Ikeda, and Andreas Pichler

Profactor Gmbh, Im Stadtgut D1, 4407 Steyr-Gleink, Austria
{matthias.propst,michael.hofmann,markus.ikeda, andreas.pichler}@profactor.at
http://www.profactor.at

Abstract. The high level of resources consumed in battery-system production is greatly disproportionate to the time-frame of use in electric vehicles. In order to alleviate this problem, second-life usage, e.g. traction batteries for home power storages, is gaining increased interest. The goal of reuse and re-purpose involves re-manufacturing as important aspect of circular economy. In this context, the productivity of assembly, maintenance and disassembly methods is key for the economic re-usability of battery systems. As of current, industry is lacking established, semi-automatic methods to enable disassembly and maintenance of Li-ion battery systems. The ongoing project BatteryLife aims to develop robotic and cognitive assistance systems to foster economically efficient re-manufacturing of battery packs and modules. This paper presents main research questions, initial requirements, targeted methods, and planned research demonstrators.

Keywords: battery system · re-manufacturing · second-life · production assistance · human-robot collaboration

1 Problem Statement

Producing battery systems demands a high amount of resources and causes high environmental effects. Li-Ion batteries for electric vehicles (EVs) are used until approx. 20 to 30 percent drop of their initial capacity. Second-life usage of retired EV batteries in less demanding applications (e.g. battery energy storage systems - BESS) has emerged to a fast growing market [1] within a circular economy. This life cycle (see Fig. 1) of battery systems includes a) assessing the state of health (SoH) and deriving re-purposing applications, b) re-manufacturing, and c) recycling of reusable raw-materials. While the ecological advantages of second-life battery usage were confirmed by a wide range of publications [2], there are open technical challenges to increase the economic efficiency of second-life batteries. Firstly, there is a large variety in the design of battery modules, packs

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 427–432, 2024.
https://doi.org/10.1007/978-3-031-76428-8_79

and individual cells. This includes the chemistry of cells, as well as the mechanical/assembly design of modules and packs. The way how battery systems are designed is crucial to enable efficient analysis of the SoH and re-manufacturing for second-life usage.

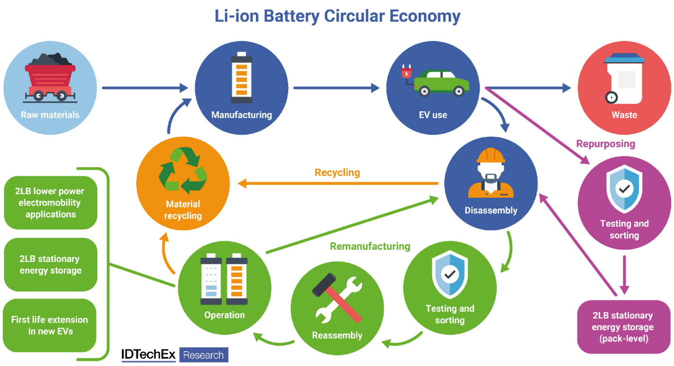


Fig. 1. Circular economy of Li-Ion battery systems [1].

As of current, manual labour is the dominant nature in re-manufacturing processes. The mentioned variety of mechanical designs can be considered as main reason. A transition towards (semi-)automated disassembly and re-assembly is required to improve cost efficiency of battery re-manufacturing. Moreover, there are open questions concerning the safety of semi-automated process, especially when having the human worker included in the process.

Considering the aforementioned problems, the contribution of this paper is to discuss requirements and possible methods towards realization of cognitive and physical assistance systems. Those shall be designed to support the human worker in a human-robot-collaborative (HRC) setup to perform SoH assessment and re-manufacturing of EV battery systems. Moreover, specific robotic skills, relevant for semi-automated re-manufacturing are highlighted. An example of a state-of-the-art battery pack is considered to underpin the difficulties in implementing the skills. Finally, targeted demonstrators to show the applicability of assistance systems in second-life re-manufacturing are introduced, briefly.

2 Requirements and Methods

This section describes requirements and targeted methods towards answering the research questions and addressing the problems discussed in Sec. 1. Firstly, an overview of the required cognitive system modules is presented. The second part focuses on robotic processes of interest.

Situation Awareness and Cognitive Architecture. The requirements for the cognitive system can be divided into four main areas: a) Architecture for mapping the process/task semantics, b) Situation detection and c) Situation-driven process execution, and d) projection-based user-guidance.

A digital description of **application-specific process knowledge and their semantics** is necessary to enable situational planning and execution of process steps. Semantic software frameworks enable the causal, logical and spatial description of relationships between successive actions. This representation of knowledge is based on existing frameworks like Knowrob [3] and Behavior-Trees [4] and will be designed to provide flexibility to cope with process variants.

Situation detection includes the recognition (translation and rotation) of features such as screws (heads), weld seams and individual parts. Moreover, the detection of the actions performed by the human, as well as the state of the (robotic) assistant are relevant. Human action recognition requires the consideration of multimodal data sources, such as hand tracking and the status of smart tools.

Building upon the semantic description of process knowledge and situation detection, a planning and reasoning system should enable a **situation-driven process** execution. The purpose is to interpret the current state of the process and to derive the next suitable action or assistance including parametrization. Finally this module initiates execution by human worker, robot system, or based on HRC. The fluidity of the interaction, significantly determines the acceptance of the assistance system.

In order to enable fluent interaction between human and assistive systems an intuitive interface is inevitable to collect feedback from the human and to guide the user through the process. This can be achieved based on a **projection-based spatial augmented reality (SAR) interface** using laser-projector devices. Important aspects include a lean integration with the cognitive architecture to enable context specific adaptation of the projected information and interfaces. The system will apply Deep-learning neural network technology to detect relevant features and parts, and to dynamically adapt the projected content. An important aspect is the projection of hints for the user, especially to indicated dangerous spots (e.g. high-voltage contacts).

Robotic Screwing. Assemblies of state-of-the-art battery systems commonly include a high number of screw fittings. Those are needed to combine cells, modules and packs together (see Fig. 2). High precision, high repeatability, and continuous monitoring of the process is required to ensure high quality.

There is need to implement robotic screwing and unscrewing of different screw types, based on state-of-the-art screwing tools. Special focus will be put on the challenging task of improving stability of local screw detection in case of rusty, dirty or worn screw heads. The targeted approach is to apply deep neural network based detection algorithms, in combination with generated data-sets based on 3d rendering engines (e.g. Blender Bolt Add-On including textures.

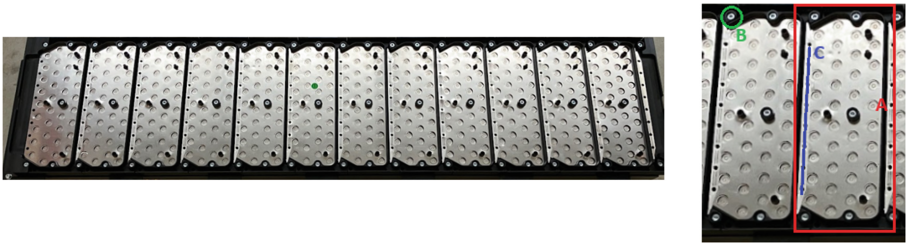


Fig. 2. Example of a battery pack consisting of twelve modules (left) [5]. Individual battery modules (A) are combined with screw fittings (B) to battery packs, and conductively connected by welded pole-sheets (C) (right).

Robotic Separation of Welded Connections. Laser welding is an established technique to conductively connect battery modules, minimizing both mechanical strains and electrical loss for improved efficiency. However, welded joints cause added difficulties if disassembling of battery packs is required. Figure 2 (right) shows an example of welded pole-sheets (see annotation C), which potentially need to be separated for e.g. replacing a module.

Therefore, the target is to implement a robotic separation process that complies with the high requirements of a) accuracy, b) minimizing impact on surrounding material and parts (e.g. dust, metal chips, heat dissipation) and c) operation in a HRC setting. Initially, the performance and suitability of different separation tools including e.g. angle grinder, milling spindle or oscillating cutter, will be analysed. Such a setup, naturally raises critical questions of safety - the goal is to consider safety aspects to fulfill minimum requirements of the standard ISO/TS 15066 [6].

3 Targeted Demonstrators

This section provides gives a brief outline on the targeted demonstrations.

Collaborative Disassembly and Maintenance includes the full setup of cognitive and physical assistance to support the human operator in maintaining (exchange of battery module in pack) a battery system correctly as depicted in Fig. 3. The goal is to guide the user through manual processes, and semi-automatically execute screwing and separation process steps through a robot system.

Qualification for Second-Life shows off a guided battery system SoH measurement, including result documentation, based on cognitive (projection) assistance. Both demonstrations in common provide situative hints to the user to avoid dangerous situation through e.g. high voltage.

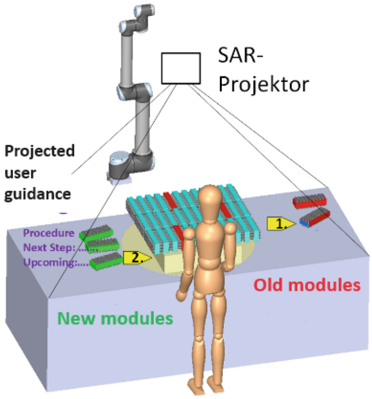


Fig. 3. Demonstrator outline for collaborative disassembly and maintenance.

4 Conclusion and Future Work

This publication introduced the project BatteryLife, which investigates and develops cognitive and physical assistance to improve efficiency in re-manufacturing of EV battery systems for second-life. The main research problems include a) conception and implementation of assistance systems in a HRC setup, b) understanding relevant assistive features for battery re-manufacturing, and c) to gain understanding how battery system design can be improved to foster efficient re-manufacturing. A major goal is to demonstrate the applicability of fluent human robot collaboration in the use-cases *disassembly and maintenance* and *assessing state-of-healh* of state-of-the-art battery systems. The next steps include the implementation of proof-of-concept demonstrators for robotic dis-assembly processes, and development of a comprehensive concept for the cognitive assistance-system architecture.

Acknowledgements. This research was funded by the project BatteryLife, an FTI Initiative Kreislaufwirtschaft of the Upper Austrian Government.

References

1. IDTechEx research Ltd: second-life electric vehicle batteries 2023–2033. https://www.idtechex.com/tw/research-report/second-life-electric-vehicle-batteries-2023-2033/924. Accessed 18 Dec 2023
2. Dong, Q., Liang, S., Li, J., Kim, H.C., Shen, W., Wallington, T.J.: Cost, energy, and carbon footprint benefits of second-life electric vehicle battery use. In: iScience. Elsevier (2023)
3. Tenorth, M., Beetz, M.: Representations for robot knowledge in the KnowRob framework. In: Artificial Intelligence, vol. 247, pp. 151-169. Elsevier (2017)
4. Iovino, M., Scukins, E., Styrud, J., Ögren, P., Smith, C.: A survey of behavior trees in robotics and AI. In: Robotics and Autonomous Systems, vol. 154. Elsevier (2022)

5. MIBA BATTERY SYSTEMS GmbH: TERRA platform battery system. https://www.mibabattery.com/en/products/terra-platform. Accessed 03 Jan 2024
6. international organization for standardization: ISO-TS 15066: robots and robotic devices: collaborative robots. ISO (2016)

Navigating Sustainability: A Real-World Examination of Life Cycle Assessment in Early-Stage Robotics

Paula Preuß[1,2](✉) and Michel Joop van der Schoor[3]

[1] TU Berlin, Berlin, Germany
[2] Reiner Lemoine Institut, Berlin, Germany
paulavpreuss@outlook.com
[3] Vrije Universiteit Brussel, 1050 Brussels, Belgium
Michel.Joop.van.der.Schoor@vub.be

Abstract. This study introduces a practical application of life cycle assessment (LCA) for an early-stage service robot. The research specifically focuses on and applies prospective LCA methodology, incorporating scale-up methods to establish future use phase scenarios. Through these scenarios, the emerging technology can be compared with the reference process, giving initial insights about its potential environmental impacts, while still considering and highlighting uncertainties inherent in early-stage LCA. These results can then serve as a feedback loop to improve upon further iterations of the assessment.

Keywords: prospective LCA · early stage LCA · robotics · scale-up

1 Motivation

Life cycle assessment can assist in identifying opportunities to improve the environmental performance of a product and as a comparison tool between processes. However, although robotics constitute a rapidly growing sector, with sales of professional service robots growing by 37% in 2021 [1], LCA is not an established method in this field. Saidani et al. observed a lack of research in their conducted literature review for autonomous systems in 2021 [2]. In 2022, Pradel et al. conducted, to their knowledge, the first LCA on commercial weeding robots, although about five hundred of them were commercially available in 2021 [3]. Therefore, the aim of this study was to apply LCA methodology to robotics at an early development stage. The specific use case was applied to a mobile service robot for public litter bin emptying developed at TU Berlin within the project Marble [4].

Conducting LCA at an early design stage offers the most potential to implement changes that will steer the product towards environmental sustainability. However, this stage also coincides with the greatest uncertainty and least available data [5]. Hetherington et al. summarized the challenges when conducting

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 433–437, 2024.
https://doi.org/10.1007/978-3-031-76428-8_80

early-stage LCA into four main aspects: comparability, scaling issues, data collection and quality, and uncertainty [6]. Considering these challenges, the outputs of a prospective LCA should not be taken as precise results. Their value rather lies in their contribution to the solution space by raising questions and suggesting possible constraints [7], giving an indication of future developments [8], and identifying the parameters that may have the greatest influence on environmental impacts [5].

2 Methodology

A comparative LCA aims to benchmark newly developed technologies against existing processes. To ensure comparability, the temporal and technological development must be aligned to the same development stage for all technologies [9]. Gavankar et al. propose using existing concepts of technology readiness level (TRL) and manufacturing readiness level (MRL) to describe the stage of technology development in LCA [10,11]. Scale-up methods are then applied to model the novel technology at a higher TRL. In general, scale-up found in the literature is often process-based and focused on efficiency gains and improvements in production yields [8,11–13].

To ensure comparability between Marble and the reference process, a scale-up of the technology was conducted. The robot is largely composed of third-party components, e.g. electric motors and sensors that present high MRLs, and the assembly of the modules doesn't include any novel processes. Therefore, the focus in this study was instead set on modeling potential future use cases as opposed to scaling up manufacturing processes. Specifically, the use phase describes the process of driving to and emptying public litter bins over the robots' lifetime, as well as any necessary maintenance.

The scale-up was conducted based on the framework proposed by Tsoy et al. [12], which consists of three steps: projected scenario definition, preparation of a projected LCA flowchart and projected data estimation. To avoid a temporal mismatch between scaled-up foreground and background systems, the energy mix for the use phase was modeled at a future point in time.

3 Life Cycle Assessment

The functional unit chosen was the emptying of all bins in Berlin during the span of one year. The chosen functional unit omits real market penetration and is chosen for best comparability against the reference process. The current process is covered by a fleet of partly electrified litter compactor trucks. To account for temporal development, the fleet was assumed to be completely electrified.

To conduct the study, robot-aided scenarios were developed with the help of scenario flowcharts. Figure 1 provides an exemplary flowchart of a robot-aided scenario. Here, the robots are deployed throughout the city and escorted by vehicles, which aid in litter collection and battery swapping. The flowcharts were then used to conduct energy simulations taking into account robot speed, energy consumption and travel distances. From these simulations, a total number of robots and vehicles needed to cover the functional unit was derived.

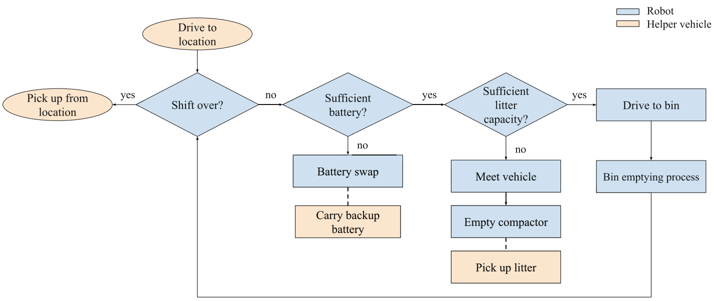


Fig. 1. Exemplary scenario flowchart

Based on the scenario flowcharts, an LCA flowchart was developed, which includes raw material extraction, manufacturing, use (including maintenance) and End-of-Life (EoL) of the robots and vehicles included in the scenarios. The life cycle inventory (LCI) was constructed using proxy processes from the *ecoinvent* 3.8 database. Raw material extraction and EoL data was based on material composition of robot components, which was derived from manufacturer data, CAD data and validations on the existing lab prototype. Assumptions made for manufacturing steps were based on expert judgement, while assumptions for the use phase were largely based on simulation data, taking into account a future energy mix. Maintenance data, as well as data relating to the vehicles was based on information about the reference process. To assess the results of the inventory analysis, the ReCiPe method was used.

4 Results

The specific results of the LCA are not discussed within the scope of this paper. However, through conducting the LCA, impact categories and robot components associated with the highest environmental impacts were able to be identified, and scenarios were compared along the entire life cycle. Here, assumptions made regarding the impacts of different life cycle phases were compared against the results. Through conducting a sensitivity analysis, scenario parameters with the highest uncertainty were identified. Since the conducted LCA is prospective, the uncertainty present in the results should be communicated carefully.

5 Uncertainty

To communicate uncertainty in an accessible way to nonexpert audiences, Gavankar et al. propose a graphical uncertainty diamond [14]. The uncertainty diamond is comprised of four axes and displayed in Fig. 2. This assessment is based on the researchers' subjective perspective. In this case, both *uncertainty due to scenarios* and *uncertainty due to current issues external to the process* were rated as high. Being a prospective LCA, external issues present an intrinsically higher uncertainty. Electricity mix and legal frameworks are only some of the assumptions made that may be subject to change. Similarly, developed scenarios rely on basic mathematical relations and functional assumptions that present a high level of uncertainty. This includes parameters like number of robots assigned to each vehicle or robot lifetime. In turn, these parameters also present the highest sensitivity. *Statistical uncertainty* was given a medium score: since the inventory was built exclusively with the help of proxy data, variability in the data is inevitable. Additionally, newly developed modules are subject to change regarding materials and mass as the development progresses. However, none of the materials or processes used to construct the LCI are novel, and proxy data was accessible. *Uncertainty due to lack of information* was assumed to be low, since functional relationships and information used to build scenarios, e.g. robot characteristics like speed or energy consumption were known, and a prototype was available for data collection and model validation.

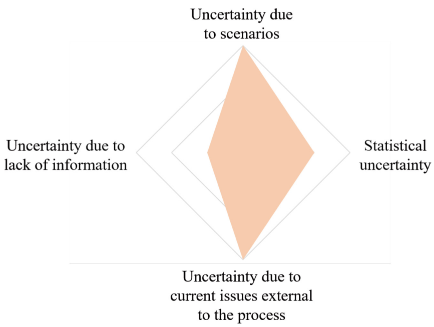


Fig. 2. Uncertainty diamond

6 Conclusion and Outlook

In this study, methodology for prospective LCA was applied to a service robot case study focusing on scale-up of the use phase. The uncertainties present in the results highlight the importance of reviewing these scenarios and using prospective LCA as an iterative method, instead of a quantitative assessment. As of this stage, the LCA can aid in identifying components that could be re-designed to reduce environmental impacts and help to identify under which frameworks a use case could outperform the reference process, and eliminate scenarios which could not. By integrating LCA as a feedback loop, future iterations can be used to correct and improve assumptions while benefiting from existing insights.

Additionally, highlighting existing uncertainties can help to address these in a targeted way.

References

1. Müller, C., Graf, B., Pfeiffer, K. (eds.): IFR Statistical Department. World Robotics 2022 - Service Robots. Frankfurt am Main (2022)
2. Saidani, M., et al.: Comparative life cycle assessment and costing of an autonomous lawn mowing system with human-operated alternatives: implication for sustainable design improvements. Int. J. Sustain. Eng. **14**, 704–724 (2021). ISSN 1939-7038
3. Pradel, M., de Fays, M., Seguineau, C.: Comparative Life Cycle Assessment of intra-row and inter-row weeding practices using autonomous robot systems in French vineyards. Sci. Total Env. **838**, 156441 (2022)
4. Van der Schoor, M.J., Göhlich, D.: Integrating sustainability in the design process of urban service robots. Front. Robot. AI 10 (2023). ISSN 2296-9144
5. Bergerson, J.A., et al.: Life cycle assessment of emerging technologies: evaluation techniques at different stages of market and technical maturity. J. Ind. Ecol. **24**, 11–25 (2020). ISSN 1530-9290
6. Hetherington, A.C., Borrion, A.L., Griffiths, O.G., McManus, M.C.: Use of LCA as a development tool within early research: challenges and issues across different sectors. Int. J. Life Cycle Assess. **19**, 130–143 (2014). ISSN: 1614-7502
7. Villares, M., Işıldar, A., van der Giesen, C., Guinée, J.: Does ex ante application enhance the usefulness of LCA? A case study on an emerging technology for metal recovery from e-waste. Int. J. Life Cycle Assess. **22**, 1618–1633 (2017). issn 1614-7502
8. Buyle, M., Audenaert, A., Billen, P., Boonen, K., van Passel, S.: The future of ex-ante LCA? Lessons learned and practical recommendations. Sustainability **11**, 5456 (2019)
9. Thonemann, N., Schulte, A., Maga, D.: How to conduct prospective life cycle assessment for emerging technologies? A systematic review and methodological guidance. Sustainability **12**, 1192 (2020)
10. Gavankar, S., Suh, S., Keller, A.A.: The role of scale and technology maturity in life cycle assessment of emerging technologies: a case study on carbon nanotubes. J. Ind. Ecol. **19**, 51–60 (2015). ISSN 1530-9290
11. Moni, S.M., Mahmud, R., High, K., Carbajales-Dale, M.: Life cycle assessment of emerging technologies: a review. J. Ind. Ecol. **24**, 52–63 (2020). ISSN 1530-9290
12. Tsoy, N., Steubing, B., van der Giesen, C., Guinée, J.: Upscaling methods used in ex ante life cycle assessment of emerging technologies: a review. Int. J. Life Cycle Assess. **25**, 1680–1692 (2020). issn 1614-7502
13. Cucurachi, S., van der Giesen, C., Guinée, J.: Ex-ante LCA of emerging technologies. Procedia CIRP **69**, 463–468 (2018). ISSN 2212-8271
14. Gavankar, S., Anderson, S., Keller, A.A.: Critical components of uncertainty communication in life cycle assessments of emerging technologies. J. Ind. Ecol. **19**, 468–479 (2015). ISSN 1530-9290

Multi-modal Electronics State Evaluation for Robotic Demanufacturing

Yifan Wu[1,2,3](✉), Chuangchuang Zhou[1,2,3], Wouter Sterkens[1,2,3], and Jef Peeters[1,2]

[1] Department of Mechanical Engineering KU Leuven, Celestijnenlaan 300, BOX2422, 3001 Leuven, Belgium
yifan.wu1@kuleuven.be
[2] Flanders Make@KU Leuven, 3000 Leuven, Belgium
[3] PSI-EAVISE-KU Leuven, Jan Pieter de Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium

Abstract. State detection is of great importance for evaluating the residual value of end-of(-first)-life electronics, which is often determining for the optimal depth of disassembly. To increase the cost-efficiency and reduce the subjectivity that is inherent to human involvement in the state evaluation process, the presented research developed an automated detection system based on Faster R-CNN-FPN using the fusion of color (RGB) and depth (D) images to identify two most encountered defects in end-of-first-life laptops: missing battery and missing cover. A Cross-Attention Fusion (CAF) module is introduced to enhance the detection accuracy. A dataset containing 513 high-quality RGB and 513 corresponding depth images of laptops has been created and annotated. Experimental results show neck fusion with a CAF module achieves the highest detection mAP of 91.1%, highlighting the potential for automated electronics state detection for vision-guided robotic demanufacturing using data fusion techniques.

Keywords: EEE · Computer vision · State detection · Battery removal · RGB-D fusion · Vision-guided robotic demanufacturing

1 Introduction

The rapid pace of innovation and consequent upgrades contribute to a significant decrease in the average lifetime of electric and electronic equipment (EEE) [1]. To promote a circular economy, reuse, repair, refurbishing, and remanufacturing are to be upscaled. However, throughout the use phase and reverse logistics various defects can be induced to the product, including damage or loss of components. Since defects or missing components can impede normal functionality, an intact surface is not only an aesthetic consideration but also a key factor in determining the feasibility of reuse. Additionally, for recycling processes, the removal of batteries from discarded devices is necessary due to the presence of hazardous substances, posing environmental and health risks. In compliance with EU legislation, the proper removal of battery from end-of(-first)-life electronics is mandated [2]. In this context, the state detection of products to be refurbished, reused, and recycled is imperative.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 438–442, 2024.
https://doi.org/10.1007/978-3-031-76428-8_81

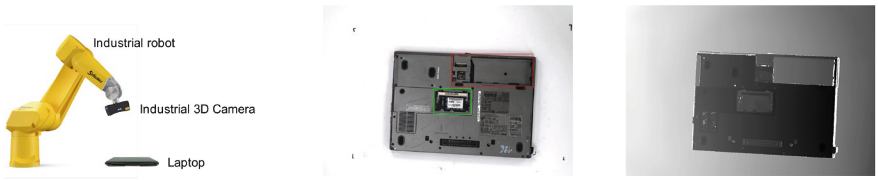
Nowadays, employees rely heavily on their knowledge and experience to inspect the EEE state with the human eye [3]. However, the manual visual inspection reduces the cost-efficiency and scalability of the whole reuse process [4]. In order to facilitate the battery collection from EEE, batteries from tablets were manually detected for a robotic scraping process, leading to an low-efficient battery removal process [5]. Therefore, automatic detection based on vison-guided robotics is necessary to replace manual inspection [6]. Recent research conducted in different fields has demonstrated the benefits of using computer vision for defect and battery detection, respectively. Shafia et al. proposed a MobileNet-based SSD detector for defect detection on hollow cylindrical surfaces [7]. Sterkens et al. [8] verified the feasibility of battery detection in X-ray images. Whereas prior research achieved promising results, the accuracy of state detection is still limited. One of the challenges is the difficulty of distinguishing the defects from the surface of the device and the missing battery area, which is sometimes highly similar in color and texture.

To address these issues, a depth camera is introduced in the presented study as an additional data source, providing range information unaffected by color and lighting variations, allowing for a more comprehensive understanding of the spatial layout. For the evaluation of the applicability of the adopted RGB-D fusion approach, laptops are chosen as case study products due to their significant market share for refurbishment and component- and product reuse, and the required battery dismantling prior to recycling. The main contributions to this paper are the following: (1) A multimodal fusion detection method based on Faster-RCNN-FPN [9] is implemented to construct a state detection system. (2) The integration of a Cross-Attention Fusion (CAF) module is adopted to improve state detection accuracy. (3) A laptop dataset of two important classes: missing battery and missing cover (piece) was created to verify the performance of the developed multimodal state detection system.

2 Materials and Methods

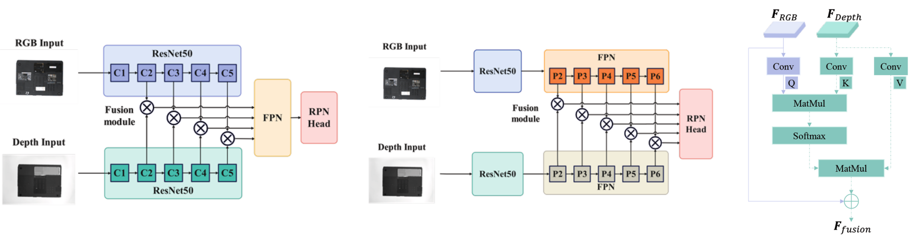
To assess the proposed method, a laptop dataset comprising 513 RGB and corresponding depth images is collected. Color and depth images were acquired using an integrated RGB-D (3D) camera Mech-Mind ProS. The camera is mounted on a Stäubli RX-160 Industrial robot, enabling image capture from different viewpoints as a kind of data augmentation, as shown in Fig. 1(a). Ground truth bounding boxes were annotated on RGB images and converted to relative coordinates for annotating depth images, as shown in Fig. 1(b)(c). Subsequently, this dataset is randomly split into an 80% training set and a 20% test set for all experiments.

The classic Faster R-CNN-FPN network is used as the model's infrastructure. The Faster R-CNN-FPN has been modified to input RGB and depth images. Faster R-CNN-FPN consists of the backbone structure, the FPN (Feature Pyramid Network), and the RPN (Region Proposals Network) head. In this case study, ResNet 50 [10] is utilized for feature extraction, which consists of five convolutional blocks. The feature layers {C1–5} are the output of these five convolutional blocks. The FPN with five feature layers {P2–6} is used for the multi-scale feature fusion. The head of Faster R-CNN-FPN is the region proposal network (RPN), which can output the object's class probability and prediction box. The backbone fusion model fuses the feature layers {C2–5} extracted from two ResNet50 backbone networks by fusion module. In contrast, the neck fusion model fuses the feature layer {P2–6} extracted from two FPN networks by fusion module, as illustrated in figure Fig. 2(a)(b).



(a) (b) (c)

Fig. 1. Setup and image examples in the dataset of laptops. (a) Camera setup (b) an example of a captured RGB image and (c) a depth image. Annotation is shown on the RGB image, where the missing cover is a green bounding box and the missing battery with a red bounding box.



(a) (b) (c)

Fig. 2. RGB-D fusion methods based on Faster R-CNN-FPN [13]: (a) backbone fusion network and (b) neck fusion network. (c) Cross-Attention Fusion (CAF) module.

Common fusion modules only perform simple pixel operations. Consequently, it is difficult to establish connections between long-range pixels, resulting in limited captured cross-modal information. Therefore, inspired by the CCNet [11], as shown in Fig. 2(c), a Cross-Attention Fusion (CAF) module is developed to establish connections between RGB and depth long-distance pixels using dot product (multiply) operation, allowing the network to focus more on the important information with small additional computational consumption. The output of CAF can be computed as:

$$\mathrm{CAF}(Q, K, V) = \mathrm{softmax}\left(QK^T\right)V \tag{1}$$

where Q, K, V are feature maps extracted from RGB and depth subnetwork. Q and V are operated to a dot product operation and then activated through a softmax layer, after which they are operated to a dot product operation with V. Then, the output is generated by the addition of extracted RGB feature maps and the processed depth feature maps.

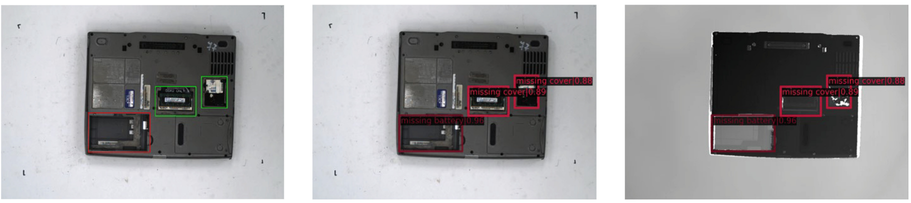
3 Experiments and Results

Two different mid-term fusion methods, backbone network fusion, and neck fusion, and three fusion modules, add, concatenation, and CAF module are adopted. For all experiments, a computing platform running Windows 10 Pro is used, which includes an NVIDIA RTX3090 24 GB GPU. A fine-tuning method [6] is applied for training using the pre-trained ResNet50_FPN on ImageNet-1K. RandomFlip is applied as the data augmentation method. The input images are all resized to 1333×800. The learning rate is set to 0.01 with the number of iterations for warming up equal to 500. The Stochastic Gradient Descent (SGD) with a momentum of 0.9 and a weight decay of 0.0001 is used. The total number of training epochs is 100, with a batch size of 2. Mean average precision (mAP) @0.5 is used for the evaluation metrics.

Evaluation results are shown in Table 1. It can be found that in the case of Faster R-CNN-FPN, the best result is neck fusion with CAF, which has an increase of 9.8% in mAP compared with the RGB-based network. Both the AP of missing battery and the AP of missing cover in the RGB-D methods are higher than the two single-modal methods. Therefore, it can be concluded that RGB-D fusion methods can significantly improve the accuracy based on Faster R-CNN-FPN on this dataset. In addition, it can be concluded that our proposed CAF module outperforms add fusion and cat fusion. Figure 3 shows the visualization results derived from the best model. The missing battery and missing cover can be classified and located accurately on the RGB and depth image. The experiments show that multi-modal fusion methods enable the detector to learn faster cross-modalities feature information at different depth layers to further improve the accuracy of detectors.

Table 1. Comparison of different methods.

Method	Single Modality		Backbone fusion			Neck fusion		
	RGB	Depth	Add	Cat	CAF	Add	Cat	CAF
AP(%) of missing battery	86.5	91.8	93.1	87.6	86.7	**94.4**	89.9	91.2
AP(%) of missing cover	76.1	65.7	76.3	76.6	86.5	77.8	80.8	**90.9**
mAP (%)	81.3	78.7	84.7	82.1	86.6	86.1	85.4	**91.1**



(a) (b) (c)

Fig. 3. State detection result examples using neck fusion with CAF: (a) ground truth bounding boxes (b) predicted bounding boxes on RGB image (c) predicted bounding boxes on depth image.

4 Conclusion

In the present research, a multimodal state detection system using RGB and depth images is developed for robotic demanufacturing. The best-performing method is the neck fusion network using the CAF module, achieving a mAP of 91.1%. This highlights the effectiveness of the CAF module. The success of the RGB-D fusion state detection system shows its potential as a guiding mechanism for robots in subsequent disassembly steps for which RGBD cameras are commonly adopted for robot path planning and manipulation. In future research, detection methods based on multi-view images will be explored to detect missing components from commonly used non-flatted electric products. Besides, additional EEE images will be collected to continuously optimize the performance of the system, enabling the network to identify the state of different devices and different types of defects.

Acknowledgment. The authors would like to express their gratitude for the dataset collection support from Ing. Dieter De Marelle, and master students Lander Govaerts and Evert Verbesselt in the LCE Research Group at KU Leuven. This work is partially supported by the China Scholarship Council (Grant NO. CSC202206020066) and Flanders Make, the strategic research center for the manufacturing industry, through the DEMAN project.

References

1. Rudolf, S., et al.: Extending the life cycle of EEE—findings from a repair study in germany: repair challenges and recommendations for action. Sustainability. **14**, 2993 (2022). https://doi.org/10.3390/su14052993
2. EUR-Lex - 32006L0066 - EN - EUR-Lex, https://eur-lex.europa.eu/eli/dir/2006/66/oj
3. Sterkens, W., et al.: Product label identification with OCR for model-specific reuse, repair, and recycling: a case study for washing machines. Journal of Industrial Ecology. n/a
4. See, J.E.: Visual Inspection: A Review of the Literature. https://doi.org/10.2172/1055636
5. Zhou, C., Engelen, B., Zaplana, I., Peeters, J.: Design of a robotic system for battery dismantling from tablets. Procedia CIRP. **105**, 273–277 (2022)
6. Zhou, C., Sterkens, W., Díaz-Romero, D.J., Zaplana, I., Peeters, J.: You Only Demanufacture Once (YODO): WEEE retrieval using unsupervised learning. Resources, Conservation and Recycling **190**, 106826 (2023). https://doi.org/10.1016/j.resconrec.2022.106826
7. Shafi, I., Mazahir, A., Fatima, A., Ashraf, I.: Internal defects detection and classification in hollow cylindrical surfaces using single shot detection and MobileNet. Measurement **202**, 111836 (2022). https://doi.org/10.1016/j.measurement.2022.111836
8. Sterkens, W., Diaz-Romero, D., Goedemé, T., Dewulf, W., Peeters, J.R.: Detection and recognition of batteries on X-Ray images of waste electrical and electronic equipment using deep learning. Resources, Conservation and Recycling **168**, 105246 (2021)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**, 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
11. Huang, Z., et al.: CCNet: criss-cross attention for semantic segmentation. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision (2019)

Portable, Robotic Material Recovery in a Box

Michail Maniadakis[1](✉), Antonios Liapis[2], Jef Peeters[3], Vasilis Makridis[4], Laurent Paszkiewicz[5], Fredy Raptopoulos[6], Javier Grau Forner[7], Myrto Pelopida[8], Friederike Kleijn[9], and Nikos Vythoulkas[10]

[1] Foundation for Research and Technology Hellas, Herakleion, Crete, Greece
mmaniada@ics.forth.gr
[2] Institute of Digital Games, University of Malta, Msida, Malta
[3] Department of Mechanical Engineering, KU Leuven, Leuven, Belgium
[4] Hellenic Recovery Recycling Corporation, Athens, Greece
[5] IRIS, Barcelona, Spain
[6] ROBENSO, Herakleion, Crete, Greece
[7] AIMPLAS, Valencia Parc Tecnologic, Valencia, Spain
[8] AXIA Innovation, Munchen, Germany
[9] International Solid Waste Association, Rotterdam, Netherlands
[10] Perifereiakos Foreas Diaxeirisis Stereon Apovliton Ionion Nison, Argostoli, Greece

Abstract. In recent times, there has been a growing integration of robotic technology in waste separation, offering a robust solution for managing the continually increasing volume of post-consumer waste. While existing solutions are typically implemented in large-scale Material Recovery Facilities (MRFs) designed for handling substantial waste quantities, the current market offerings for robotic waste sorting systems prove to be less cost-effective for smaller or less accessible areas. However, in contrast to bulk processing, sorting waste close to its source has demonstrated enhanced material recovery in terms of both quality and quantity. To address this disparity, the Horizon Europe RECLAIM project is dedicated to developing a portable, robotic MRF (prMRF) tailored for small-scale waste sorting and material recovery. RECLAIM adopts a modular multi-robot/multi-gripper approach for material recovery and employs an AI-powered computer vision module for material categorization. Additionally, the project introduces an innovative Recycling Data-Game that encourages citizens to actively participate in RTD activities, contributing annotations crucial for training the AI models utilized in material categorization.

Keywords: Portable Robotic MRF · Robotic Waste Sorting · Material Recovery

1 Introduction

Recent advancements in technology have facilitated the modernization of industrial Material Recovery Facilities (MRFs), now supported by intelligent and autonomous robotic waste treatment equipment [1]. This approach is currently employed by large-scale MRFs typically situated near major urban areas for municipal waste treatment

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 443–448, 2024.
https://doi.org/10.1007/978-3-031-76428-8_82

[2]. Despite the high efficiency of these industrial solutions, the transportation of waste to these large plants incurs high costs and complicates material recovery due to the compression applied to reduce volume before transport.

However, there are instances where this operational model does not efficiently address waste treatment needs. Such scenarios include natural disasters, social, cultural, and athletic events with surges in waste volumes during specific periods, rural or remote areas, and transport hubs. To cater to the recycling needs of these cases, a recent market trend focuses on implementing lightweight, flexible, and portable waste management units that can be swiftly deployed in areas requiring waste treatment, especially the recovery of valuable recyclable waste. Despite the substantial demand for these systems, they currently lack the integration of smart, high-tech solutions that could significantly enhance their productivity. Consequently, waste sorting in these specialized sectors is still carried out manually today.

The RECLAIM project (https://www.reclaim-box.eu/) utilizes established and well-tested AI-driven robotic waste management technologies, currently undergoing refinement and integration into an advanced "portable, robotic Material Recovery Facility" (prMRF). This endeavor aims to substantially elevate local-scale material recovery activities, imparting them with efficiency comparable to industrial standards. Additionally, the project embraces a citizen science approach. Beyond heightening social awareness of the European Green Deal, it facilitates and motivates citizens to engage in project RTD activities by contributing annotations crucial for the deep learning process used in training the AI-based waste categorization module.

2 The Essence of the RECLAIM Approach

In pursuit of its objective, the RECLAIM project implements concrete steps to create innovative solutions within four essential technology pillars aimed at facilitating the recovery of valuable materials from post-consumer waste streams. The pertinent pillars are outlined below.

PIL-1: Advanced AI for Material Localization and Categorization (AI-ILC) - Leverage and improve existing AI technologies to create efficient solutions that integrate visual [3] and hyperspectral [4] information, achieving high-performance localization and categorization of recyclable waste. These solutions should be applicable in real and harsh conditions, supporting material recovery efforts in challenging environments.

PIL-2: Modular, Low-cost, High Performance, Multi-robot/Multi-gripper Recyclable Recovery - Advocate for a new modular and easily scalable structure for robotic waste sorting, relying on cost-effective, high-productivity Robotic Recycling Workers RoReWos (see Fig. 1) equipped with diverse grippers tailored for various material types [5]. The RoReWos' simplified architecture reduces implementation costs and doubles the rate of "picks per invested euro". The development of RoReWos is realized by their grouping in teams, in a way that enhances the productivity of the composite system.

PIL-3: Portable Robotic Material Recovery Facility (prMRF) - The RECLAIM project strives to integrate waste loading and ferrous material sorting equipment, along with AI-powered robotic waste sorters, within the prMRF container box. The prMRF accomplishes fully automated waste treatment and the efficient recovery of recyclable materials with a minimal number of human workers undertaking simple supporting tasks. The prMRF container can be easily transported to designated locations and become fully operational within a few hours, ensuring continuous and effective waste sorting for extended periods.

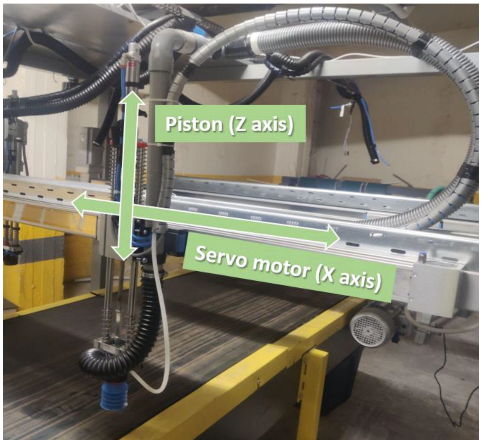


Fig. 1. One of the RECLAIM's low-cost Robotic Recycling Workers

PIL-4: Environmental Gaming for Social Awareness, Data Collection and Annotation - Enhance public awareness of recycling through an innovative Recycling Data-Game (RDG). This game not only sheds light on challenges related to waste treatment but also motivates citizens to engage in project activities using a citizen science approach, contributing data for AI-ILC training [6]. Simultaneously, RDG serves as a platform to communicate the fundamental principles of AI and Data Science to the wider public.

The above described implementations advocate for the decentralized processing of recyclable waste, anticipating elevated recovery rates and enhanced material purity. The improved sorting efficiency is further facilitated by the higher quality, freshness, and uncompressed state of the materials presented to the RoReWos.

3 Focused Deployments

The RECLAIM technological advancements will be put into practice and subjected to thorough evaluations through rigorous real material recovery tasks. RECLAIM aims at deploying a prMRF at diverse pilot sites to assess its performance under real and particularly demanding conditions, recovering materials from post-consumer packaging waste.

Specifically, the deployment of prMRF will take place in the Ionian Islands group, Greece, offering an opportunity to evaluate its performance in distinct use-case scenarios.

The Ionian group comprises seven main islands, widely recognized as popular tourist destinations. All these islands encounter significant challenges in waste management, particularly during the summer period. However, the relatively small size of these islands and the high seasonality of waste production do not favor investments in a full-size Material Recovery Facility (MRF).

To date, packaging waste generated by residents and tourists is typically compacted and packaged for transportation to the mainland for material recovery. The residue from this waste treatment must either return to the islands for landfilling or remain on the mainland, incurring notably high costs for landfilling. In addition to the increased logistics expenses, particularly in roundtrip transfers (island-mainland-island), the quality of recyclable materials transferred to the mainland for processing often diminishes due to inadequate temporary storage and compression. This leads to heightened material degradation and entanglement, resulting in increased difficulty in recovery and lower recycling rates. Consequently, valuable recyclables are more likely to end up in landfills instead of being recycled.

Much like the situation in Greece, the utilization of prMRF for the localized processing of fresh, uncompressed, high-quality waste and the recovery/sorting of recyclables in a decentralized fashion is crucial for all rural areas. By enhancing the economic feasibility of decentralized sorting, RECLAIM offers a viable strategy to significantly boost recovery rates for valuable recyclable materials by sorting in close proximity to the source. Th RECLAIM approach also helps in reducing costs and minimizing environmental impact by eliminating the need for back-and-forth transportation of waste.

The geographical setting of the Ionian Islands group offers a distinctive chance to evaluate both the foundational technologies and the business case for the prMRF. RECLAIM is in direct collaboration with partners, holding the responsibility for waste management across all the Ionian islands. This will play a crucial role in investigating how prMRF can effectively address the specific requirements of different islands. Specifically, RECLAIM has outlined three scenarios of particular interest for the Ionian Islands group, as well as for various other regions in Europe encountering similar challenges. The three scenarios are outlined below.

Scenario-1: Material recovery from mixed recyclables streams. The primary recycling method employed in Greece involves citizens collecting mixed post-consumer municipal packaging waste in bins, which is later sorted at material recovery facilities. However, as noted earlier, particularly for the Ionian Islands, this process results in high logistical costs and increased material loss, especially when compression and improper temporary storage occur at transport stations. RECLAIM will evaluate the operation of prMRF in treating mixed recyclable packaging waste streams, with a specific focus on the PMD stream (Plastics, Metals, Drinking cartons). This stream involves the positive separation of multiple material types and represents the most challenging scenario due to the high material complexity examined in the RECLAIM project. Addressing this challenge will be a focal point throughout the entire duration of the project.

Scenario-2: Cleaning of citizen separated, material streams. In a section of Corfu, which is part of the Ionian Islands group, the promotion of recyclable separation by citizens involves the use of dedicated bins for collecting recyclable packaging materials in distinct streams (a recycling approach also embraced by many North European countries). The gathered material streams require further processing to enhance their purity and, consequently, their value in the secondary market. RECLAIM will evaluate the operation of prMRF in treating material-specific streams, employing negative separation to eliminate undesired objects. This scenario will be explored concurrently with Scenario-1, offering the opportunity to adapt prMRF for both positive and negative material selection. By comparing Scenarios 1 and 2, the project aims to assess the advantages and disadvantages of citizen-led separation. While citizen separation is anticipated to yield higher sorting efficiencies, it also incurs higher collection costs.

Scenario-3: Close to the source material recovery. The inherent portability of the prMRF allows for its installation in proximity to areas with high production of packaging waste, such as transportation hubs (e.g., ports, airports), festivals, sporting events, exhibitions, etc. Unfortunately, until now, the fresh, high-quality waste generated in such areas has been collected with garbage trucks and is commonly mixed with other municipal waste. Consequently, RECLAIM will leverage the portability of the prMRF to investigate the feasibility of local, close-to-source robotic material recovery and sorting, comparing its efficiency with the current practice. This scenario will be explored in the final six months of the project by installing the prMRF at a central location, serving as an open and live demonstrator for EU citizens (residents and tourists), local authorities, governments, circular economy stakeholders, and potential investors.

4 Conclusions

RECLAIM lays the foundation for the adoption of innovative and decentralized waste sorting and management strategies. This is achieved through the creation of low-cost, portable, easily installable, and highly productive prMRFs, designed to facilitate local waste treatment as well as direct and efficient recovery of recyclable materials. Consequently, the prMRF is anticipated to play a crucial role in establishing a global circular economy model that is leakage-free, ensuring complete material recovery from post-consumer waste in any location, even in the most remote areas. This is expected to pave the way for the widespread adoption of prMRF in the market, contributing significantly to Europe's goal of becoming zero-polluting, climate-neutral, sustainable, and globally competitive.

References

1. Wilts, H., Garcia, B.R., Garlito, R.G., Gómez, L.S., Prieto, E.G.: Artificial intelligence in the sorting of municipal waste as an enabler of the circular economy. Resources **10**(4), 28 (2021)
2. Raptopoulos, F., Koskinopoulou, M., Maniadakis, M.: Robotic pick-and-toss facilitates urban waste sorting. 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), pp.1149–54 (2020)
3. Koskinopoulou, M., Raptopoulos, F., Papadopoulos, G., Mavrakis, N., Maniadakis, M.: Robotic waste sorting technology: toward a vision-based categorization system for the industrial robotic separation of recyclable waste. IEEE Robot. Autom. Mag. **28**(2), 50–60 (2021)
4. Alexakis, G., Maniadakis, M.: Hyperspectral classification of recyclable plastics in industrial setups. In: Artificial Intelligence Applications and Innovations. AIAI 2023. IFIP Advances in Information and Communication Technology, vol 675. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-34111-3_35

5. Zhang, H., Peeters, J., Demeester, E., Kellens, K.: Deep learning reactive robotic grasping with a versatile vacuum gripper. IEEE Trans. Rob. **39**(2), 1244–1259 (2023). https://doi.org/10.1109/TRO.2022.3226148
6. Sfikas, K., Liapis, A., Yannakakis, G.: Controllable exploration of a design space via interactive quality diversity. GECCO Companion, pp. 167–170 (2023)

Robot Design with Sustainability-Impact-Based Requirements

Michel Joop van der Schoor(✉)

Vrije Universiteit Brussel, 1050 Brussels, Belgium
Michel.Joop.van.der.Schoor@vub.be

Abstract. In the design of complex mechatronic products such as robots, requirements are necessary to specify functions and track surrounding conditions, interfaces, stakeholder needs and overall objectives. These are predominantly based on the performance in terms of function, value and cost. It is well established that each product has impacts on the environment, society and economy throughout its life cycle and thus on sustainability. However, these impacts are in their entirety not accounted for by performance-based requirements. Therefore, this paper explores the use of sustainability-impact-based requirements along the early development phase of an urban service robot. On the basis of the case study the paper attempts to characterize sustainability-impact-based requirements and explain their role in robot design. It will also describe problems arising due to this process and method.

Keywords: Sustainability · Robot Design · Requirements

1 Introduction

Defining and understanding requirements is a cooperative, iterative and incremental process supported by methods such as checklists, benchmarking, user participation or the scenario technique [1]. The background to these methods is the development of an understanding of the context of the product and the necessary processes over the life cycle phases to ensure the requirements list to be as comprehensive and complete as possible [2]. Requirements ought to be formulated in a clearly specified and measurable manner as to easily validate the performance of the first concepts and prototypes. This is primarily applied to the functionality and the value/costs of the developed product [3].

Oftentimes impacts on the ecological, social and economic dimension of sustainability that occur along the life cycle due to raw material extraction, production, use phase and the end of life are neglected. To support a more sustainable product development a comprehensive list of requirements needs to reflect upon these impacts and set objectives accordingly. The adoption of impact-based requirements aims to change the very inception of a product and contribute to sustainability and the understanding of possible impacts on society.

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 449–453, 2024.
https://doi.org/10.1007/978-3-031-76428-8_83

2 Methodology

During the product development process (PDP), decisions are made that affect the entire product life cycle. The resulting processes have an impact on all dimensions of sustainability: the environment, society and economy [4–6]. Developers must therefore anticipate which processes will take place due to the creation of the product in order to avoid or influence them according to the set of sustainability objectives (e.g. SDGs). Particularly in the early phases of product development, when requirements are gathered and defined, the possible influence is estimated highest. Studies on the ecological impact of products observed that approx. 60–80% of the environmental impact is determined in this phase [7].

Life cycle engineering (LCE) poses a key element for the assessment of a product from raw material extraction to its disposal. Tools such as the life cycle assessment (LCA), social life cycle assessment (SLCA) or life cycle costing (LCC) have been developed over the last three decades [5, 6, 8, 9]. They contain certain impact categories, that can vary depending on the chosen method. For the LCA the ReCiPe method is commonly used and consists of 18 impact categories ranging from global warming potential and resource scarcity to acidification and eutrophication [10]. The SLCA poses different stakeholder categories with impact subcategories such as fair labor, local employment, wealth distribution or transparency [5]. A method suggested by Kohl et al. supplements impact categories in the social dimension especially for automation initiatives during the use phase with a focus on workers and the intrusion of public spaces [11]. Furthermore, van der Schoor et al. proposed checklists to support the requirements engineering in the social and ecological dimension of sustainability [12].

2.1 Sustainability Impact-Based Requirements

As requirements are used for the evaluation and assessment, one can also take the assessment methods and reverse engineer requirements that are based on these well-established impact categories in the different dimensions of sustainability. This was done during the case study in the project MARBLE, where an urban service robot was developed in support for the municipal waste management [12]. In order to follow a holistic sustainability approach for gathering requirements, three key points were identified:

Intended functionality. This refers to the envisioned functionality of the robot and its application area. To comprehend possible impacts during the use phase a thorough assessment of the robot and its process integration is necessary. Operating in a certain environment a robot will affect different stakeholder groups that are directly or indirectly involved. The implementation of functionalities, outward design choices and the programmed behavior can impact the interaction with and perception by stakeholders. Similarly, energy consumption, maintenance and repair will be affected by these choices made during the development phase. Therefore, requirements focusing on these impacts can alter the outcome significantly, when identified correctly.

In the scope of this assessment and requirements deduction the robot's purpose and possible implications should be included as well. This proves to be difficult as resulting rebound effects, misusage or malpractice are unpredictable. Nevertheless, the knowledge of such effects and the use of value-driven objectives [12] will create sensitivity and might prevent them to a certain degree.

Production and disposal. These comprise impacts due to the life cycles: Raw material extraction, production and development, distribution, disposal and recycling/recovery. Requirements in this category can be deducted directly from the impact categories mentioned in the LCA or SLCA. Although the LCA is linked to the ecological and the SLCA to the social dimension, they are not that selective. The LCA also covers impacts on human health and the SLCA contains categories concerning socio-economic aspects. The LCC focuses on the entirety of the product's costs and is thus linked to requirements of financial constraints impacting solely the economic sustainability.

Holistic approach. The third key point underlines the importance of the holistic approach. As a robot can contribute to one aspect of sustainability it may simultaneously inhibit others. Since the goal must be to create an overall sustainable robot, it is paramount to include requirements that cover all dimensions of sustainability.

3 Application

As the initial motivation of the development of a robot resides in a certain function to be achieved, requirements to ensure the functional quality and also manage financial constraints are still mandatory. However, by adding new requirements to complete a comprehensive understanding of the robot's impact, design choices linked to function or cost have to be regarded from their sustainability aspects as well.

During the project MARBLE a first prototype was developed and requirements gathered on the functional and impact level. Assessing the relevant areas concerning the intended functionality (see Fig. 1) and taking impact categories from life cycle methods into account aids in the formulation of a first set of impact-based requirements (examples see Table 1).

The choice of materials poses a good example for the various implications. On the functional side used materials have to withstand weather conditions and working loads. On the impact side the weight will contribute to the energy consumption, hence CO_2 emissions and operational costs. Furthermore, the extraction of materials can differ greatly in their detrimental effects on the environment as well as in their suitability for recycling and reuse. Also, the origin of the material and the vendor may result in different policies and work ethics which determine social and socioeconomic impacts, e. g. forced and child labor or the equal distribution of earned profits among value chain actors [5].

During the use phase social requirements mainly derive from the stakeholder assessment. The functionality of the robot can increase accessibility by taking over physical straining tasks when implemented in an inclusive way as for instance visually impaired require audio cues. The same is relevant for operation in public spaces, where the robot could create a hostile environment by neglecting certain needs of vulnerable groups, too quick and unforeseeable movements, great exerted forces or scary design.

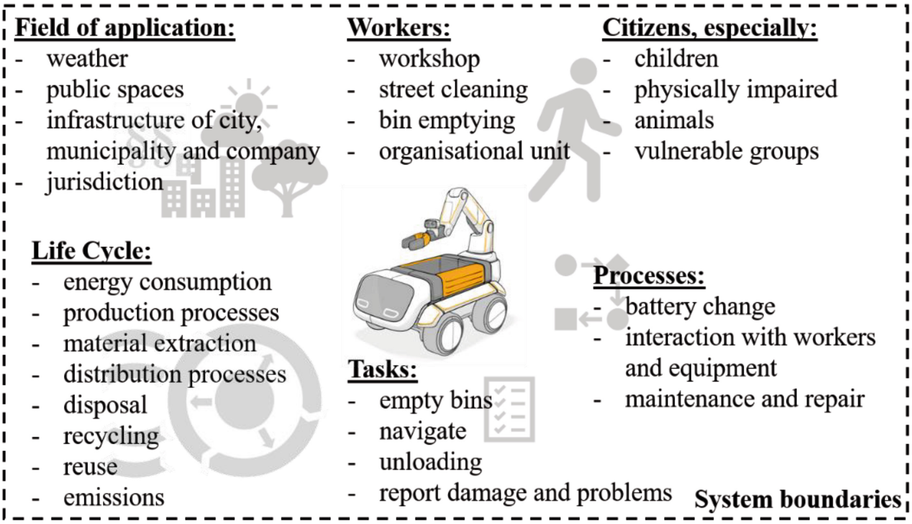


Fig. 1. Relevant areas for sustainability-impact-based requirements in the project MARBLE

Table 1. Examples of sustainability-impact-based requirements within MARBLE

Dimension	Requirement
Ecological	Elec. Energy used in the process shall not come from fossil resources
Ecological	Emit less CO_2 than the original process without robots
Ecological	Minimal material diversity for recycling
Social	No capturing of sensitive personal data
Social	No creation of hostile environments for workers and citizens
Social	Increase equality and accessibility in the work environment
Economic	Cost of production and operation shall be affordable to municipality
Economic	Earned profits shall be equally distributed along value creation chain

4 Implications on Robot Design

Considering all life cycle phases with a focus on the development of the use phase facilitates a comprehensive understanding of the robot system and its implications. Identifying conflicts at an early development stage is advantageous, as it allows for more manageable modifications to the concept with relatively lower associated costs.

Nevertheless, the early development stage presents its own set of challenges. The limited information about the robot and the absence of real-world experiments, particularly in terms of stakeholder interaction, pose difficulties in validating impact-based requirements. Critical information regarding the social and ecological implications of materials and mechatronic hardware is often undisclosed. Uncovering this information necessitates extensive research and assessment efforts, leading to increased costs and the integration of various disciplines into the development process.

This integration, while crucial for obtaining essential information, introduces complexities in project management. Communication within the development team becomes intricate, involving multiple disciplines and impacting mutual understanding. Compounding these challenges, the predetermined objectives may yield contradictory and competing requirements, thereby increasing the potential for conflict.

Despite the increased workload, the insights gained from a comprehensive list of impact-based requirements in the early stages of the PDP underscore the importance of placing holistic sustainability considerations at the core of the development process. This approach not only acknowledges potential conflicts but also supports a more nuanced and well-informed evaluation of sustainability throughout the development lifecycle.

References

1. Bender, B., Gericke, K.: Entwickeln der Anforderungsbasis: Requirements Engineering. In: Bender, B., Gericke, K. (eds.) Pahl/Beitz Konstruktionslehre, pp. 169–209. Springer, Berlin Heidelberg, Berlin, Heidelberg (2021)
2. McAloone, T.C., Pigosso, D.C.A.: Ökodesign. In: Bender, B., Gericke, K. (eds.) Pahl/Beitz Konstruktionslehre, pp. 975–1021. Springer, Berlin Heidelberg, Berlin, Heidelberg (2021)
3. Ponn, J., Lindemann, U.: Konzeptentwicklung und Gestaltung technischer Produkte. Optimierte Produkte - systematisch von Anforderungen zu Konzepten. VDI. Springer, Berlin, Heidelberg (2008)
4. Birkhofer, H., Schulze, K., Zhao, S., Sarnes, J.: Umweltgerechtes Konstruieren. In: Rieg, F., Steinhilper, R. (eds.) Handbuch Konstruktion, 2nd edn., pp. 599–620. Carl Hanser Verlag, München (2018)
5. UNEP: Guidelines for social life cycle assessment of products and organizations (2020)
6. Laurent, A., et al.: The role of life cycle engineering (LCE) in meeting the sustainable development goals – report from a consultation of LCE experts. J. Clean. Prod. **230**, 378–382 (2019)
7. Pigosso, D.C.A.: Ecodesign maturity model: a framework to support companies in the selection and implementation of ecodesign practices. São Carlos (2012)
8. Geissdörfer, K.: Total cost of ownership (TCO) und life cycle costing (LCC). Einsatz und Modelle: ein Vergleich zwischen Deutschland und USA. Zugl.: Oestrich-Winkel, Europ. Business School, Diss., 2008, 1st edn. Controlling und Management, Bd. 7. LIT, Berlin, Münster (2009)
9. Deutsches Institut für Normung: Umweltmanagement - Ökobilanz - Grundsätze und Rahmenbedingungen. Beuth Verlag GmbH, Berlin (14040:2021-02) (2021)
10. Huijbregts, M.A.J., et al.: ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. Int. J. Life Cycle Assess. **22**, 138–147 (2017)
11. Kohl, J.L., van der Schoor, M.J., Syré, A.M., Göhlich, D.: Social Sustainability in the Development of Service Robots. Proc. Des. Soc.: Des. Conf. **1**, 1949–1958 (2020)
12. van der Schoor, M.J., Göhlich, D.: Integrating sustainability in the design process of urban service robots. Front. Robot. AI **10**, 1250697 (2023)

Author Index

C. Secchi and L. Marconi (Eds.): ERF 2024, SPAR 33, pp. 455–458, 2024.
https://doi.org/10.1007/978-3-031-76428-8