From Human Perception and Action Recognition to Causal Understanding of Human-Robot Interaction in Industrial Environments [*]

Stefano Ghidoni[1], **Matteo Terreran**[1], **Daniele Evangelista**[1], **Emanuele Menegatti**[1],
Christian Eitzinger[2], **Enrico Villagrossi**[3], **Nicola Pedrocchi**[3], **Nicola Castaman**[4], **Marcin Malecha**[5]
Sariah Mghames[6], **Luca Castri**[6], **Marc Hanheide**[6], **Nicola Bellotto**[1,6]
[1]Dept. of Information Engineering, University of Padova, Italy, [2]Profactor GmbH, Austria,
[3]STIIMA-CNR, Italy, [4]IT+Robotics srl, Italy, [5]Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR),
Germany, [6]Lincoln Centre for Autonomous Systems, United Kingdom
{stefano.ghidoni,matteo.terreran,daniele.evangelista,emanuele.menegatti,nicola.bellotto}@unipd.it,
christian.eitzinger@profactor.at, {enrico.villagrossi,nicola.pedrocchi}@stiima.cnr.it,
nicola.castaman@it-robotics.it, marcin.malecha@dlr.de,
{smghames,lcastri,mhanheide,nbellotto}@lincoln.ac.uk

Abstract

Human-robot collaboration is migrating from lightweight robots in laboratory environments to industrial applications, where heavy tasks and powerful robots are more common. In this scenario, a reliable perception of the humans involved in the process and related intentions and behaviors is fundamental. This paper presents two projects investigating the use of robots in relevant industrial scenarios, providing an overview of how industrial human-robot collaborative tasks can be successfully addressed.

1 Introduction

Human-robot interaction plays a fundamental role in industrial applications, where it is a key element in the Industry 4.0 paradigm. This is an enabling technology for mass customization and special production processes, because the human experience and dexterity can be combined with the power of a robot to perform heavy duty tasks.

This paper presents two projects that are strongly based on human-robot collaboration. The DrapeBot project, detailed in Section 2, focuses on the co-manipulation of large carbon fiber parts, lead by the human operator and supported by the robot. The cooperation is based on the AI-driven detection and understanding of the human activity observed by a camera network equipping the workcell. The DARKO project, presented in Section 3, proposes a new generation of mobile robot manipulators for intralogistics applications. The robots are equipped with advanced human-robot interaction capabilities including predictive planning, mutual intention communication, context-based modelling, and causal reasoning. The goal is for the robot to understand and anticipate the dynamics of potentially unknown environments to enhance safety and operational efficiency.

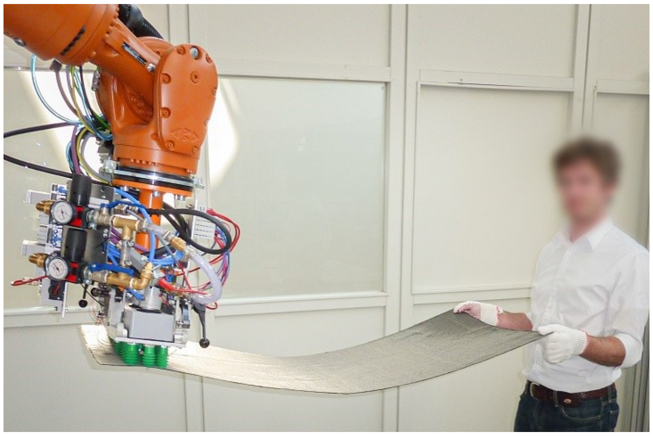

Figure 1: A human operator cooperates with a robot for carbon fiber transportation.

2 The DrapeBot Project

Draping is one of the most complex operations in carbon fiber manufacturing. It is carried out by transporting the carbon fiber fabric onto the preform and adapting its shape. The success of this process depends on the manual ability of the human operator to correctly place the fabric and model the shape: this means that a high degree of manual ability is needed, that can hardly be achieved using an automatic process. This process is usually performed by human operators. When it comes to large components, usually two or more highly skilled operators are needed—one usually controlling the process, and the other(s) acting as a support. This is, at the same time, expensive and unneeded, because one single skilled operator could manage the draping process, while the task of the additional operators could be carried out by an automatic system capable of following the indications provided, as exemplified in Fig. 1. This means to better exploit the skills of the human operator who acts as the leader of the draping process, guiding the whole process and assessing its quality, while, at the same time, employing robots for executing the

---

[*] This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No. 101006732 and No. 101017274.

Figure 2: Example of body pose estimation system, a key element of the smart workcell.

more trivial but heavy duties.

2.1 Human-Robot Interaction for Carbon Fiber Parts Manufacturing

The European project DrapeBot (www.drapebot.eu) has the goal of developing a novel human-robot interaction technology capable of supporting one single human operator handling large carbon fiber parts. This is based on cutting edge AI tools, both at the perception and the action stage. In such context, a thorough understanding of the scene is needed to properly feed the AI systems that are in charge of determining the robot action and motion in real-time.

2.2 The Smart Workcell

DrapeBot aims at developing a smart workcell, that is, a workcell that is aware of the processes that take place inside the workspace, thanks to the sensors and the artificial intelligence modules that equip the cell. Understanding human behavior starts from perceiving the human body posture, involving Action Recognition [1], people re-identification [2] and human-computer interaction—a typical output of a body pose estimation system is shown in Fig. 2. In recent years, several efforts have been done to obtain fast and reliable human motion estimates in real-time. Several works, like [3], focused specifically on employing image processing techniques to gain knowledge about the human pose in real-time. In some cases, filtering has been also applied to provide smoother results [4].

The current status of the workcell is another key element. For example, similar gestures may have different meanings depending on the current status of the workcell. To sense the environment in the most flexible way, semantic segmentation is employed–this is a per-pixel classification task based on a set of classes representing important elements in the workcell.

2.3 Workcell Environment Understanding

Focusing on the perception stage, DrapeBot aims at enhancing the state of the art and the industrial practice by combining visual sensors with inertial sensors to obtain a fast and accurate perception of how human workers move in the workcell. This will be accompanied by a semantic interpretation of the environment. In this way, the system will ensure a smooth and reliable interaction with the human(s), that will not be blocked nor slowed down due to temporary perception failures caused, for example, by an occlusion preventing the vision system to perceive the human operator. Such high-end perception systems will be employed to provide a thorough understanding of the working environment and the body pose of the person(s) present in the working area of the robot. A probabilistic model for human body pose estimation will be derived, capable of taking advantage of both visual and inertial sensors, enhancing the detection when both are available but capable of working when only one sensory system is working. An important element of the project will be the definition of the sensing accuracy needed to meet the safety requirements of the robotic workcell.

The semantic environment perception equipping the workcell will be a key element for detecting in detail the current status of the workcell, because it can provide a high-level description of the components of the environment, enabling automatic evaluation of the current status of the production process, possible misalignments, and possible risks when workers are in proximity of dangerous moving elements.

The DrapeBot project will enhance the state of the art and the industrial practice by deploying an integrated layered framework where the Task and Motion Planning (T&MP) and the low-level control will be jointly designed to work in full synergy. The T&MP will manage human and context information to do both: reason on the user model and interpret several sensorial information. Sensor data will come from the robotic system and the human perception system, that will include an AI-based human action and intention understanding module. This will enable optimization of task sharing between humans and robots and selecting the processing areas to minimize human interference. The low-level control will exploit the estimated human intention control dealing with the compensation of the model inaccuracies, and allowing a continuous adaptation on the best task and motion plan computed by the T&MP module.

Considering the T&MP layer, the core advancement will consist in the deployment of modules designed for efficient on-line computation of a sub-plan after performing each action. Thus, exploiting at each iteration updated sensory readings, it implicitly takes into account unexpected changes in the estimated state due to previous sensors and actuators failures, or environment re-configurations. This framework will integrate a human-aware motion planner that will generate a collision-free trajectory provided in real-time leveraging on a data-driven approach (e.g, a Deep Reinforcement Learning framework) to improve the planning effectiveness.

3 The DARKO Project

Mobile robot manipulators can play a crucial role in future agile production processes. In order to be effective though, they must demonstrate high levels of technical capability, adaptability to dynamic environments, and advanced human-robot interaction skills. Despite significant progress in recent years, the current state-of-the-art does not meet yet the requirements of tomorrow's industry.

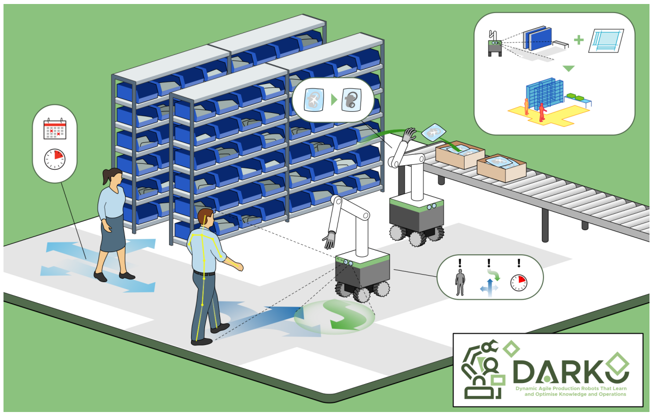

Figure 3: Concept model of the DARKO project. Mobile robot manipulators operate in intralogistic environments exploiting predictive planning and safe human-robot spatial interaction.

The goal of the H2020 project DARKO[1] – Dynamic Agile production Robots that learn and optimise Knowledge and Operations – is to develop a new generation of production robots for highly dynamic manipulation tasks, which are able to navigate safely and efficiently thanks to new predictive planning and human-robot interaction capabilities (Fig. 3). In particular, the project aims to answer scientific and technical questions about dynamic manipulation (e.g. throwing an object) enriched by context- and human-aware navigation in intralogistic settings.

3.1 Human-Robot Spatial Interaction

One of the key objectives of DARKO is efficient and safe human–robot co-production, which is addressed by advancing the state-of-the-art in human-robot *spatial* interaction (HRSI), i.e. the way robot and human move with respect to each other. This includes new methods to learn, predict, and exploit activity patterns in shared environments; communicate mutual navigation intents; and create a framework for risk-aware robot planning and coordination.

The project will extend previous work of the partners on human detection, tracking, and motion prediction. It will further investigate the use of spatial augmented reality and anthropomorphic gestures for robot's intents communication. Finally, new models for context-aware representation of HRSI will be created, including novel solutions for causal understanding of the environment's dynamics, so that the robot can operate safely and efficiently even in case of unexpected situations (i.e. no training data).

3.2 Context-aware Representations of HRSIs

Human spatial interactions, defined as the mutual influence of motion behaviours between two or more people, depend on both human activities (e.g. speed and destination) and objects or constraints (e.g. nearby door, narrow corridor, etc). Similarly, HRSIs are influenced by the robot's goal and nearby humans, but also other factors outside the direct control of the interacting agents (e.g. narrow corridor in Fig. 4). In many real-world scenarios (e.g warehouse, hospitals, etc), the environment is highly dynamic, with sudden variations of agents and potential interactions. The success of a production line is directly linked to each worker's tasks, whether these are done collaboratively or individually. The efficient completion of such tasks on the ability of the agents to communicate their mutual intents, so that pre-defined plans can be adapted and optimised. A requirements for intent communication the ability of the robot to predict and implement the most opportune spatial interactions with the surrounding agents.

Figure 4: Example of HRSI where the robot stops to let the operator pass in a narrow corridor.

In order to better capture the nature of HRSIs in realistic scenarios, as the logistic ones, and therefore enable safer and more socially acceptable robot motion behaviours, DARKO will create new context-aware representations of HRSI embedding the relative motion with respect to other relevant features of typical production environments (e.g. workcells, storage areas, forklifts, etc.). However, given the complexity of real-world dynamic environments, a continuous or purely quantitative description of such relations is computationally unfeasible for the operational requirements of a production robot. Therefore, a neuro-symbolic representation of HRSIs is adopted, which is based on a theoretically-sound Qualitative Trajectory Calculus (QTC) for moving objects [5]. This representations will extend previous work on QTC-based HRSI to explicitly model and predict the relative motion of the robot with respect to nearby humans and other static (e.g. doors, pallets) or dynamic (e.g. forklifts) objects. The increased complexity of these models, driven by the number of people co-existing in a workplace, requires new algorithmic solutions for dealing with space dimensionality, uncertainties, and possible lack of information (e.g partially observable space).

The proposed qualitative models will provide an interface between the classic sensor-level perception of objects and humans in the robot's environment, and the higher-level in-

---

[1] https://darko-project.eu/

terpretation required to understand the causal relations between its agents, focusing in particular on human-human and human-robot spatial interactions. These models can also be used to infer the safest robot motion behaviour in proximity to people and to provide a human-readable description of HRSIs for post-hoc analyses and risk assessment.

Current work in that direction includes the development of predictive models for qualitative interactions between agents in highly dynamic scenarios, considering robot, humans, and their neighboring objects (e.g doors). To this end, some of the available open-source datasets in public spaces (e.g. JackRabbot[2]) are being considered. To assess the computational complexity and efficiency of algorithms that can adapt to domain shifts for different environments and scenarios, a performance comparison is being performed between different hybrid symbolic/deep learning-based architectures for predicting qualitative interactions between human-human and human-robot spatial interactions.

3.3 Causal Reasoning for Safe HRSI

Safety is a key requirement of any human-robot interaction. New methods for the analysis of different safety aspects are needed to increase this requirement. In DARKO, safety will be intensively put at test since the robot could face many different complex situations, including working and sharing the environment with humans, the behaviour of whom can be affected by other agents and objects in the environment. Dealing with such situations implies a deeper understanding of human motion intentions and possible factors that influence them, besides the robot's presence. Causal Inference, i.e. the "science" of understanding cause and effect [6], could represent a turning point in how robots cooperate with humans flexibly and, most importantly, interact with them safely.

One of DARKO's goals is to make the robot able to learn the effect of its behaviours by observing how humans react to its actions. This ability will require the robot to discover and reason on cause-effect relations and, given a desired motion plan, choose the most effective HRSI, taking into account both safety and task performance. The project therefore will develop new algorithms to learn and reason on causal models of HRSI, enabling the robot to compute intervention (e.g. "what happens if I move and get close to that human?") and counterfactual probabilities (e.g. "what would have happened if I remained still instead of moving?"), as illustrated in Fig. 5. By using these probabilities, the robot can choose the most appropriate behaviour in proximity of humans, increasing safety and efficiency of HRSIs.

To this end, causal discovery algorithms are being investigated with application to real-world scenarios similar to DARKO (e.g. THÖR[3] dataset), in order to study the behaviours of human agents in warehouse environments and to reconstruct the underlying causal model which describes them.

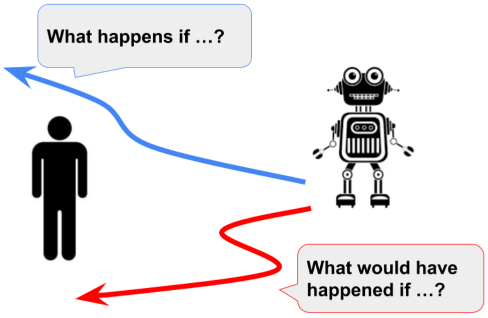

Figure 5: Example of intervention (blue trajectory) and counterfactual (red trajectory) in HRSI.

4 Conclusions

This paper presented two projects investigating the human-robot interaction scenarios and applications in industrial contexts, DrapeBot and DARKO. The former proposes the design of a smart workcell that can actively support the human operators in a production process. The sensory system installed in the workcell can understand human behavior and intention thanks to a wide use of AI systems, that are integrated with the robot task and motion controller. The latter focuses on the development of a new generation of mobile robot manipulators that use advanced AI methods for modelling and predicting human-robot spatial interactions, including algorithms for causal reasoning to improve safety and operational efficiency in dynamic environments.

References

[1] F. Han, X. Yang, C. Reardon, Y. Zhang, and H. Zhang, "Simultaneous Feature and Body-Part Learning for real-time robot awareness of human behaviors", 2017 IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 2621-2628.

[2] S. Ghidoni, and M. Munaro, "A multi-viewpoint feature-based re-identification system driven by skeleton keypoints", Robotics and Autonomous Systems, 90, 45-54, 2017.

[3] Z. Cao, T. Simon, S.E. Wei, and T. Sheikh, "Real-time multi-person 2D pose estimation using part affinity fields", in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 7291-7299, 2017.

[4] A. Malaguti, M. Carraro, M. Guidolin, L. Tagliapietra, E. Menegatti, and S. Ghidoni, "Real-time tracking-by-detection of human motion in RGB-D camera networks", in 2019 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), pp. 3198-3204.

[5] Delafontaine, M., Chavoshi, S. H., Cohn, A. G., Van de Weghe, N. (2012). A qualitative trajectory calculus to reason about moving point objects. In Qualitative spatio-temporal representation and reasoning: Trends and future directions (pp. 147-167). IGI Global.

[6] Pearl, J., Mackenzie, D. (2019). The book of why. Penguin Books.

---

[2] http://svl.stanford.edu/projects/jackrabbot/

[3] http://thor.oru.se/