Rudenko, A., Zhu, Y., Rodrigues de Almeida, T., Schreiter, T., Castri, L. et al. (2025)
Hierarchical System to Predict Human Motion and Intentions for Efficient and Safe Human-Robot Interaction in Industrial Environments
In: *1st German Robotics Conference*

Hierarchical System to Predict Human Motion and Intentions for Efficient and Safe Human-Robot Interaction in Industrial Environments

Andrey Rudenko[1], Yufei Zhu[2], Tiago Rodrigues de Almeida,[2], Tim Schreiter[3], Luca Castri[4], Nicola Bellotto[4,5], Timm Linder[1], Narunas Vaskevicius[1], Luigi Palmieri[1], Martin Magnusson[2], Achim J. Lilienthal[2,3]

***Abstract*— In this paper we present a hierarchical motion and intent prediction system prototype, designed to efficiently operate in complex environments while safely handling risks arising from diverse and uncertain human motion and activities. Our system uses an array of advanced cues to describe human motion and activities, including generalized motion patterns, full-body poses, heterogeneous agent types and causal contextual factors that influence human behavior.**

I. Introduction

Predicting the motion of humans and other agents is a fundamental ability for advanced autonomous systems, in particular if they operate in complex, highly dynamic and safety-critical environments [1]. Contrary to this notion, many existing systems lack advanced human motion perception and prediction tools, and therefore need to operate strictly conservatively to achieve safety. This approach may work in simple cases, but eventually poses risks, delays and distractions in the environment, and prevents the robots from engaging in more complex interactive activities.

We advocate for the notion that the successful adoption of mobile robots in industry is deeply intertwined with their collaboration with human labor. Robot systems should not only integrate smoothly into existing operations but also facilitate effective and safe cooperation between robots and human workers. This way, businesses of varying scales can integrate intelligent automation systems without the need to erect entirely new, fully automated production.

To that end, as part of the EU project DARKO[1], we developed a hierarchical system to predict human motion and intentions with multiple components operating with a short- and long-term perspective. We have designed the system considering the following requirements:

- In order to accomplish its tasks, the robot needs to move over long distances in topologically complex environments. To efficiently plan the global routes, we need to consider long-term trajectory predictions of other moving agents and generalized dynamics patterns in the environment.
- The robot shares the environment not only with walking pedestrians, but also with other robots, as well as people occupied in diverse tasks, potentially operating material handling equipment such as forklifts and delivery vehicles. The robot should be aware of the distinct dynamics patterns of those agents.
- For safe close-proximity interactions in narrow environments, the robot should be able to read and interpret full-body cues of the nearby people.
- As human behavior is a product of many task- and environment-related factors, we need to learn the influence of those factors and use it to refine our predictions.

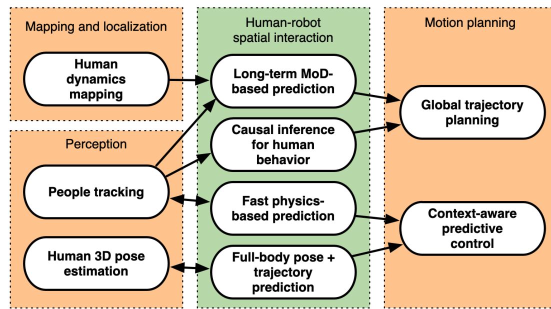

Fig. 1. Overview of the prediction architecture.

The human motion prediction system developed and implemented in the DARKO project hosts an array of methods with increasing levels of context-awareness and high-level cues of human behavior. It is designed to provide the required levels of prediction accuracy and operation speed to the downstream components. A summary is given in Fig. I. In this paper we present the individual components of the system, the novel human motion data we use for training and evaluation, as well as the downstream methods.

II. DARKO prediction system architecture

A. Fast physics-based prediction

For short-term prediction, we deploy fast physics-based methods [2] that are used by the robot controller for fast reaction and collision avoidance. These methods can also be used to improve the temporal association for people tracking. With appropriate filter calibration, we measured $0.15 - 0.3\,\mathrm{m}$ average displacement error of the Constant Velocity Model (CVM) when making $1.6\,\mathrm{s}$ predictions.

B. Long-term trajectory prediction using Maps of Dynamics

For a more far-reaching outlook, we deploy long-term trajectory prediction methods [3], [4]. Our solution is based on using information about previously observed motion,

[1] Bosch Corporate Research, Germany, andrey.rudenko@de.bosch.com
[2] Örebro University, Sweden
[3] Technical University of Munich, MIRMI, Chair of Perception for Intelligent Systems, Germany
[4] University of Lincoln, UK
[5] University of Padua, Italy

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

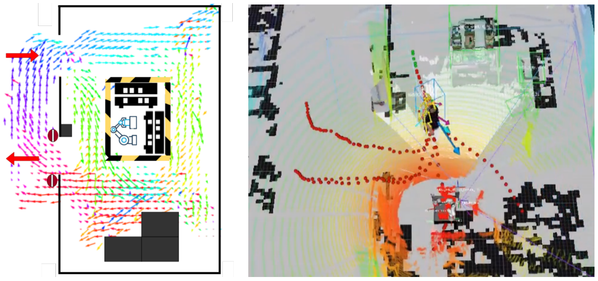

Fig. 2. **Left:** Map of Dynamics, generalized from $5\,\mathrm{min}$ of motion observed in the environment. **Right:** Multi-modal trajectory prediction for a walking person.

specifically represented as a CLiFF map, which encodes motion patterns in each location as a mixture of orientation-velocity components [5]. We bias the CVM with samples from the CLiFF-map to generate multi-modal trajectory predictions (see Fig. 2). We found the predictions to be up to 45% more accurate than the prior art, measuring on average $1.4\,\mathrm{m}$ displacement on a $12\,\mathrm{s}$ horizon. In topologically complex large indoor spaces with multiple robots or stationary sensors, long-term predictions can be used by a global path planner to coordinate the robot with the ongoing activities in the environment and proactively plan a path of least disturbance.

C. Trajectory prediction for heterogeneous agents

We aim to improve the prediction accuracy and reduce the forecast uncertainty in environments shared by heterogeneous agents with diverse dynamics. To that end, we extended the MoD-based predictions with agent classes, as well as train class-conditioned versions of several popular prediction methods based on deep learning (LSTM, VAE, GAN and transformer). We show that considering agent classes can improve prediction accuracy in some settings [6], but future work should concentrate on the rarely occurring and unsupervised classes [7].

D. Full-body pose + trajectory prediction

Jointly with the 2D coordinates, we propose to predict the short-term 3D full-body pose sequence using a novel transformer architecture [8] (see Fig. 3). This method can improve full-body pose tracking accuracy in the events of sensor noise, partial or full occlusions, and inform the robot controller in narrow and crowded locations where the geometric representation of the person's position as a point on the plane may overly constrain the robot. On average, we measured $0.15\,\mathrm{m}$ trajectory displacement on $2\,\mathrm{s}$ horizon and $0.43\,\mathrm{m}$ for individual joints of walking people.

E. Causal inference for modeling human behavior

To gain a deeper understanding of human behavior and the contextual factors that influence it—such as robot proximity, task type, and timing—we employ causal models. Specifically, we developed ROS-Causal [9], a framework designed to collect trajectory data for humans and robots, enabling onboard causal discovery and reasoning. Causal discovery is conducted using F-PCMCI [10] and CAnDOIT [11], algorithms specifically tailored for robotics applications. This approach reconstructs the causal model underlying human behavior, providing a solid foundation for explaining and enhancing human motion prediction.

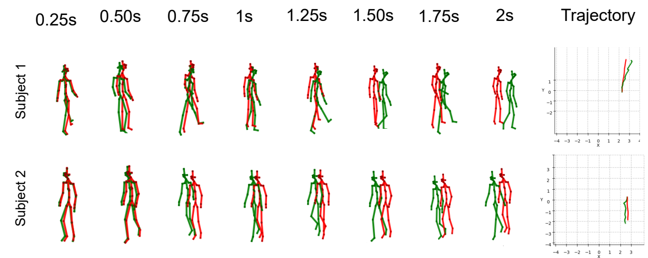

Fig. 3. Full-body pose and trajectory prediction over 2 s horizon at 16Hz. Prediction is shown in **red**, ground truth in **green**.

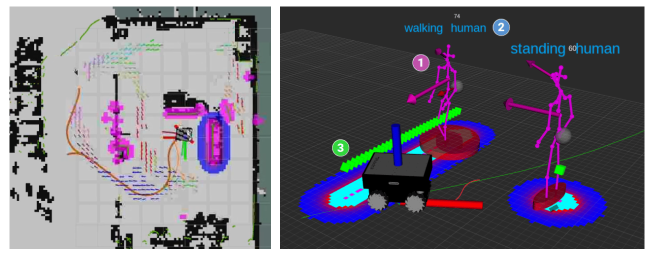

Fig. 4. **Left:** Flow-aware path planning, complying to the direction people usually walk in this environment. **Right:** Context-aware collision avoidance, considering (1) full-body poses, (2) detected activities, and (3) short-term motion predictions. The robot proactively clears the path of a walking human, while preparing to bypass a standing person.

III. THÖR-MAGNI: Human trajectory dataset with advanced movement cues

Development of advanced prediction methods is challenging due to the lack of diverse motion and interaction data with the relevant contextual cues. To support the development of the DARKO components, we designed and recorded the THÖR-MAGNI motion capture dataset [12]. THÖR-MAGNI consists of 52 four-minute recordings of participants performing various industry-inspired activities related to navigating alone and in groups, searching and transporting small and large objects, and interacting with the DARKO robot. THÖR-MAGNI includes over $3.5\,\mathrm{h}$ of 6D trajectory and head orientation data for 40 participants. On top of that, eye-tracking data and fully labeled micro-actions are available for 16 participants [13].

IV. Integration

In order to achieve long-term human-aware operation of the DARKO robot, we propose a planning architecture [14] that benefits from the components of the hierarchical prediction stack. Specifically, we use sampling-based motion planning [15] with added costs coming from the learned dynamics patterns (Maps of Dynamics) to plan flow-compliant path through the environment. To guide the robot along the global path, we propose an efficient Model Predictive Control (MPC) formulation that considers short-term 2D motion predictions, 3D poses and current activities [16]. We showcase these components in Fig. 4 and describe the planning stack in [14].

REFERENCES

[1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[2] A. Rudenko, L. Palmieri, W. Huang, A. J. Lilienthal, and K. O. Arras, "The atlas benchmark: An automated evaluation framework for human motion prediction," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 636–643.

[3] Y. Zhu, A. Rudenko, T. P. Kucner, A. J. Lilienthal, and M. Magnusson, "A data-efficient approach for long-term human motion prediction using maps of dynamics," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), Workshop on Long-term Human Motion Prediction*, 2023.

[4] Y. Zhu, A. Rudenko, T. P. Kucner, L. Palmieri, K. O. Arras, A. J. Lilienthal, and M. Magnusson, "CLiFF-LHMP: Using spatial dynamics patterns for long-term human motion prediction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 3795–3802.

[5] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1093–1100, 2017.

[6] T. R. de Almeida, Y. Zhu, A. Rudenko, T. P. Kucner, J. A. Stork, M. Magnusson, and A. J. Lilienthal, "Trajectory prediction for heterogeneous agents: A performance analysis on small and imbalanced datasets," *IEEE Robotics and Automation Letters*, 2024.

[7] T. R. de Almeida and O. M. Mozos, "Likely, light, and accurate context-free clusters-based trajectory prediction," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 1269–1276.

[8] N. N. Chandregowda, A. Rudenko, and T. Linder, "Unified 3d human pose dynamics and trajectory prediction for robotic applications," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2025.

[9] L. Castri, G. Beraldo, S. Mghames, M. Hanheide, and N. Bellotto, "Experimental evaluation of ros-causal in real-world human-robot spatial interaction scenarios," in *IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN)*, 2024.

[10] L. Castri, S. Mghames, M. Hanheide, and N. Bellotto, "Enhancing causal discovery from robot sensor data in dynamic scenarios," in *Conference on Causal Learning and Reasoning*. PMLR, 2023, pp. 243–258.

[11] ——, "Candoit: Causal discovery with observational and interventional data from time series," *Advanced Intelligent Systems*, vol. 6, no. 12, p. 2400181, 2024. [Online]. Available: https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202400181

[12] T. Schreiter, T. R. de Almeida, Y. Zhu, E. G. Maestro, L. Morillo-Mendez, A. Rudenko, L. Palmieri, T. P. Kucner, M. Magnusson, and A. J. Lilienthal, "THÖR-MAGNI: A Large-scale Indoor Motion Capture Recording of Human Movement and Robot Interaction," *Int. J. of Robotics Research*, 2024.

[13] T. R. de Almeida, T. Schreiter, A. Rudenko, L. Palmieiri, J. A. Stork, and A. J. Lilienthal, "THÖR-MAGNI Act: Actions for human motion modeling in robot-shared industrial spaces," in *Proc. of the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI) (short contribution)*, 2024.

[14] E. Stracca, A. Rudenko, L. Palmieri, P. Salaris, L. Castri, N. Mazzi, V. Rakcevic, N. Vaskevicius, T. Linder, N. Bellotto, T. Schreiter, Y. Zhu, M. C. Quero, O. Napolitano, E. Stefanini, L. Heuer, M. Magnusson, A. Swikir, and A. J. Lilienthal, "DARKO-Nav: Hierarchical risk- and context-aware robot navigation in complex intralogistic environments," in *European Robotics Forum*, 2025.

[15] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on gaussian mixture fields," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6176–6181.

[16] E. Stefanini, L. Palmieri, A. Rudenko, T. Hielscher, T. Linder, and L. Pallottino, "Efficient context-aware model predictive control for human-aware navigation," in *IEEE Robotics and Automation Letters*, 2024.