Knowledge-based Programming by Demonstration using semantic action models for industrial assembly

Junsheng Ding*, Haifan Zhang*, Weihang Li*, Liangwei Zhou*, Alexander Perzylo*

* fortiss - Research Institute of the Free State of Bavaria, Munich, Germany

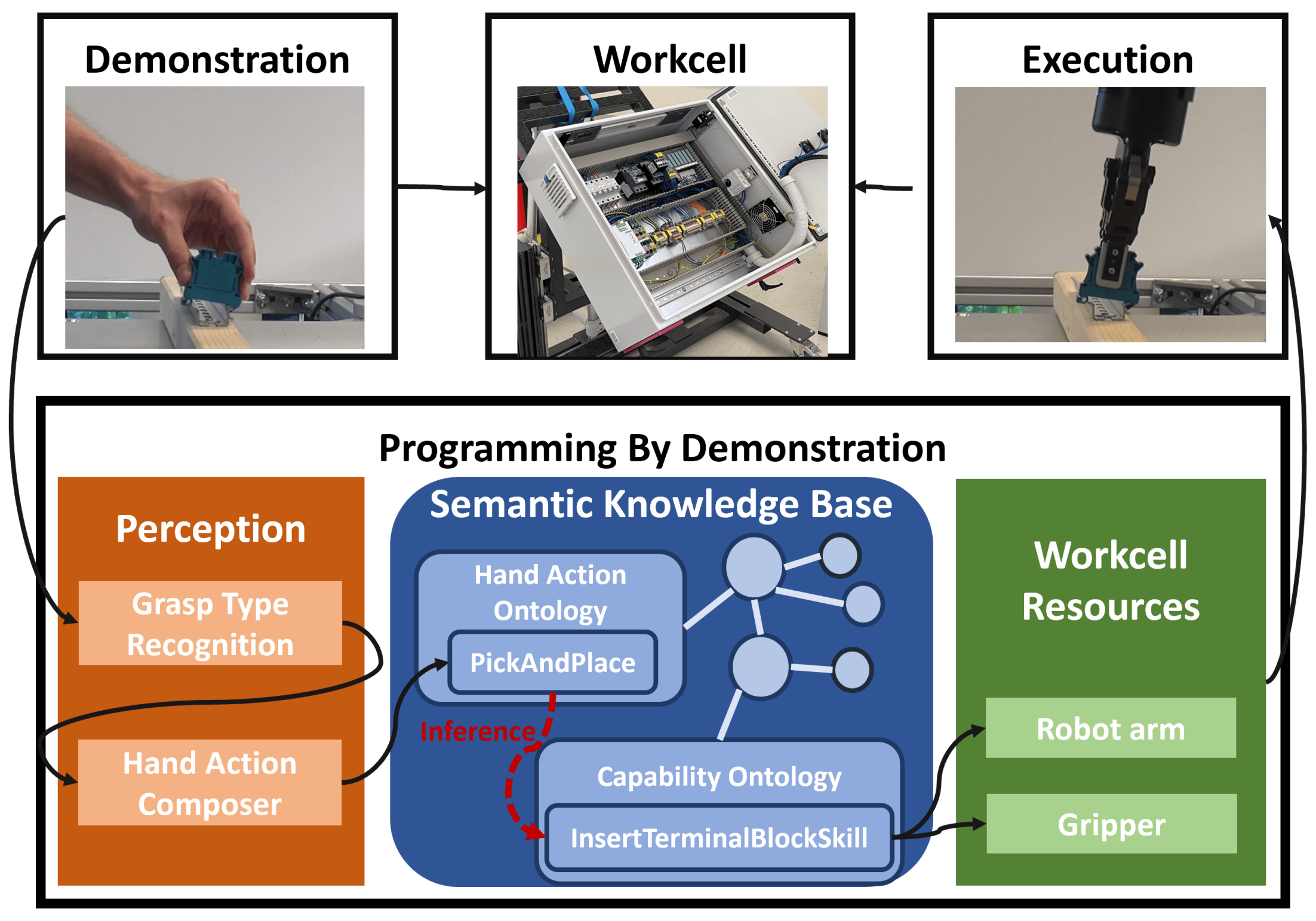

Fig. 1: Concept of the proposed kb-PbD approach.

Concept

Robot assembly processes in small and medium-sized enterprises (SMEs) often need frequent reprogramming due to product changes. Such assemly processes are typically constructed as sequences of robot tasks on the symbolic level.

Programming by Demonstration (**PbD**) with passive obeservation enables robots to learn these processes from the humans, which requires the correct recognition of the product-specific assembly actions. However, vision-based methods for action recognition face challenges due to the lack of a comprehensive dataset in the industrial domain.

In this work, we propose a **knowledge-based PbD** (**kb-PbD**) approach that porpulates a knowledge graph, encoded in Web Ontology Language (OWL), with basic hand action properties. OWL rules are contructed for automatic reasoning on the specific assembly action. Robot tasks for reproduction are then constructed using pairwise matching based on skill representations.

Method

Perception

We utilze grasp type[1] and hand velocity as the general action features to parse **PrimitiveActions** as shown in Fig. 1.

Similar to [2], the Mediapipe Hand [3] is used to locate the hand position and generate 21 hand joints, which are then fed into an LSTM for grasp type recognition. Upon recognizing a new action, we also porpulate the KB with action properties of **grasp type**, **hand pose** $\{t_x, t_y, t_z, r_x, r_y, , r_z\} \in \mathbb{R}^6$, and the **interacted objects**.

Semantic hand action model

The ontology to describe human assembly action is shown as Fig. 2. The product specific actions are represented as Defined Classes, which enables the automatic classification by the OWL reasoner when the necessary and sufficient conditions are met.

For example, the **InsertTerminalBlock** action in the use case as in Fig. 4 is defined on the restriction on the **interacted objects** with the following expression:

$$\begin{aligned}&\mathit{InsertTerminalBlock} :-\\ &\exists\, \mathit{hasSuperClass(PickAndPlace)}\\ &\wedge \exists\, \mathit{hasPickObject(TerminalBlock)}\\ &\wedge \exists\, \mathit{hasPlaceObject(DinRail)}\end{aligned}$$

Reproduction with robot tasks

For reproducing human assembly processes, each action are translated into a robot task regarding the skill descriptions, reusing the parameters of, e.g. the involved object.

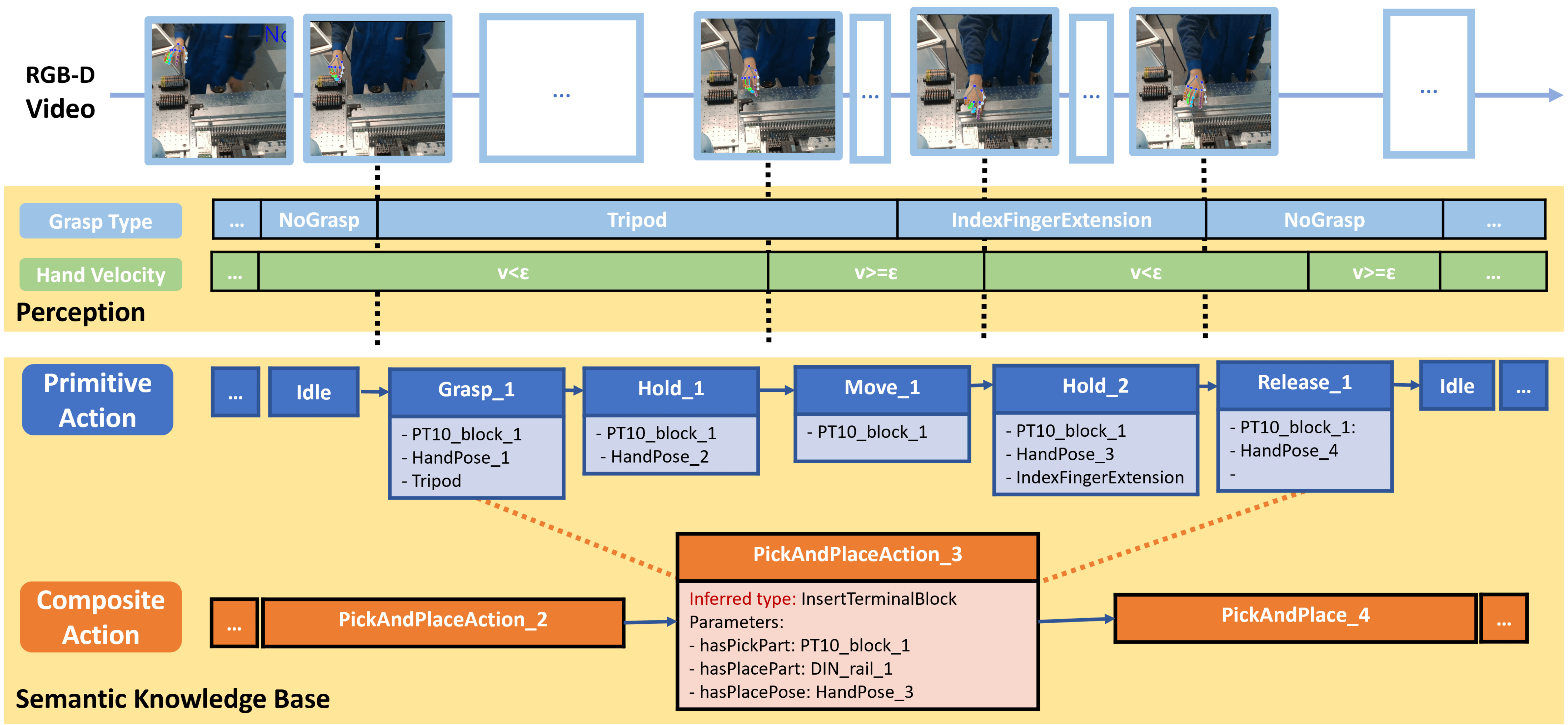

Fig. 2: An example of video processing for human assembly action recognition. Firstly, the *PrimitiveActions* are parsed and cartegorized based on 2 general action properties: hand velocity and grasp type.
The ***PrimitiveActions*** futher aggregate to form ***CompositeActions***, e.g. a ***Grasp*** and a ***Release*** define the start and the end of a **PickAndPlace**.

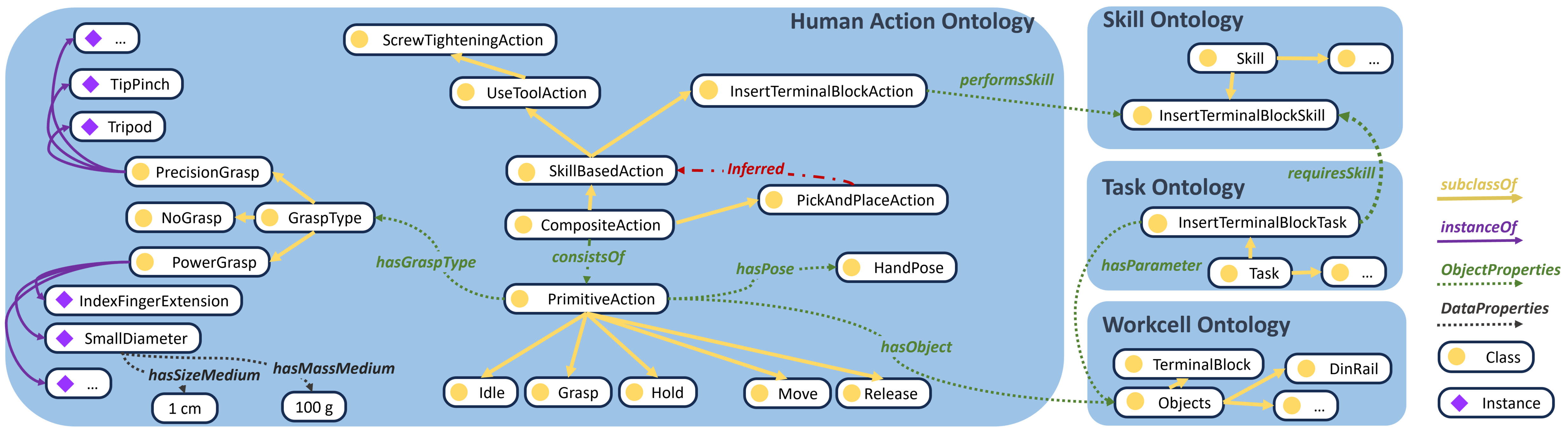

Fig. 3: Overview of partial ontologies within our semantic KB. The semantic linking between human actions and the robot tasks are modeled in the Skill Ontology, which represent their functionality for the assembly.

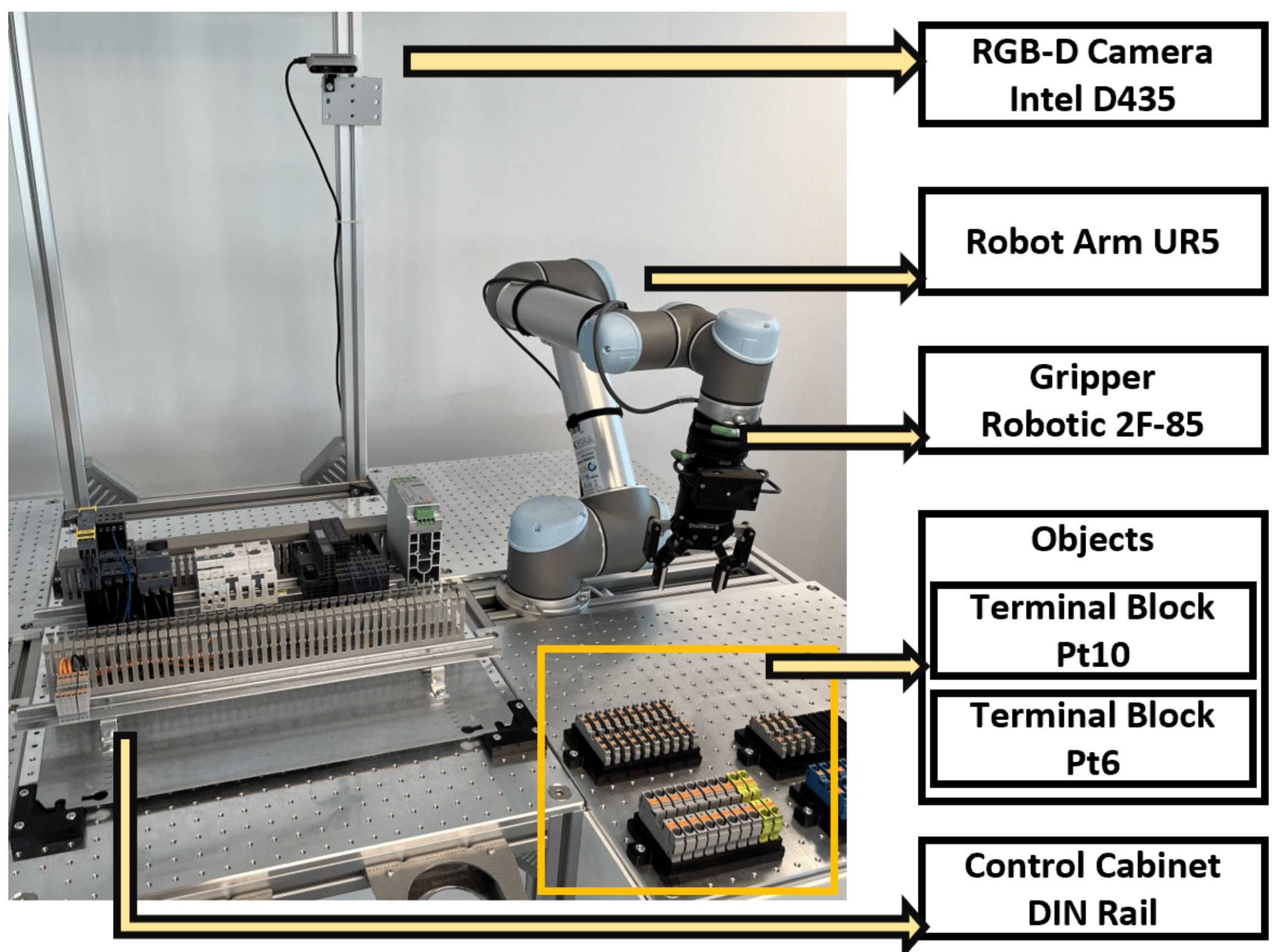

Fig. 4: Robot workcell for control cabinet assembly used during evluation.

Experiment

To demonstrate the effectiveness of the proposed kb-PbD approach, experiments were conducted on a real robot cell for control cabinet assembly.

For this type of production, 3 product-specific actions are required: InsertTerminalBlock, PickAndPlace, and ScrewTightening.

A classical vision-based action recognition method with CNN+LSTM [4] was used for baseline comparison.

In an assembly process consisting of 15 tasks, our approach achieved an average recognition rate of 89.95%[1], while the CNN+LSTM achived an recognition rate on 80.94%.

Additionally, our approach only require reusable dataset for grasp types, while the cnn+lstm method requires extra dataset collection process.

[1] Frame-wise recognition rate. Recognition rate on the CompositeAction using OWL reasoner: 100% .

Conclusion

This work proposed a kb-PbD approach for robot programming in industrial assembly. The human assembly sequence are firstly parsed with the general feartures, while a semantic knowledge base is composed to store the action properties.

The KB classifies product-specific actions based on OWL rules, reducing the need for extensive dataset collection.

The action information stored in the KB are also efficiently used for symbolic-level robot program generation.

For future work, we plan to explore the use of more general action features, such as contact events from hand-object interactions.

Additionally, we aim to investigate how large language models (LLMs) can be utilized to automate OWL rule generation for adding new action types, as these are still manually generated in the current approach.

Reference

[1] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," IEEE Transactions on human-machine systems, 2015.

[2] N. Elangovan, R. V. Godoy, F. Sanches, K. Wang, T. White, P. Jarvis, and M. Liarokapis, "On human grasping and manipulation in kitchens: Automated annotation, insights, and metrics for effective data collection," in 2023 IEEE International Conference on Robotics and Automation (ICRA).

[3] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," arXiv preprint arXiv:2006.10214, 2020.

[4] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

Contact

Junsheng Ding
ding@fortiss.org