# Knowledge-Based Digital Twin for Predicting Interactions in Human-Robot Collaboration

Tadele Belay Tuli
Chair for Manufacturing Automation and Assembly (FAMS)
*PROTECH-Institute for Production Technology*
*University of Siegen*
Siegen, Germany
tadele-belay.tuli@uni-siegen.de

Linus Kohl
Research Group of Smart and Knowledge-Based Maintenance
*Institute of Management Science*
*TU Wien*
Vienna, Austria
linus.kohl@tuwien.ac.at

Sisay Adugna Chala
*Data Science and AI*
*Fraunhofer FIT*
Sankt Augustin, Germany
chala@fit.fraunhofer.de

Martin Manns
Chair for Manufacturing Automation and Assembly (FAMS)
*PROTECH- Institute for Production Technology*
*University of Siegen*
Siegen, Germany
martin.manns@uni-siegen.de

Fazel Ansari
Research Group of Smart and Knowledge-Based Maintenance
*Institute of Management Science*
*TU Wien*
Vienna, Austria
fazel.ansari@tuwien.ac.at

***Abstract*— Semantic representation of motions in a human-robot collaborative environment is essential for agile design and development of digital twins (DT) towards ensuring efficient collaboration between humans and robots in hybrid work systems, e.g., in assembly operations. Dividing activities into actions helps to further conceptualize motion models for predicting what a human intends to do in a hybrid work system. However, it is not straightforward to identify human intentions in collaborative operations for robots to understand and collaborate. This paper presents a concept for semantic representation of human actions and intention prediction using a flexible task ontology interface in the semantic data hub stored in a domain knowledge base. This semantic data hub enables the construction of a DT with corresponding reasoning and simulation algorithms. Furthermore, a knowledge-based DT concept is used to analyze and verify the presented use-case of Human-Robot Collaboration in assembly operations. The preliminary evaluation showed a promising reduction of time for assembly tasks, which identifies the potential to i) improve efficiency reflected by reducing costs and errors and ultimately ii) assist human workers in improving decision making. Thus the contribution of the current work involves a marriage of machine learning, robotics, and ontology engineering into DT to improve human-robot interaction and productivity in a collaborative production environment.**

***Keywords—human-robot interaction, digital twin, human action models, ontology, machine learning***

## I. Introduction

Semantic representation of the human action model in a human-robot collaborative environment is essential for agile planning and building digital twins (DT) in a hybrid workplace. Action level motion description and recognizing anticipatory actions have considered semantic knowledge for robot behavior development [1]. This semantic knowledge may endow collaborative robots with the cognitive capability to assist human workers by learning and adapting to assembly activities. Additional functionalities with cognitive capabilities may further increase the market for collaborative robots (cobots). Statistics from [2] regarding human-robot collaboration (HRC) show the global market's projection from USD 981 million in 2020 to USD 7972 million in 2026. It is a 41.8% compound annual growth rate (CAGR) during 2020-2026. The emerging of affordable and low-cost cobots leads to greater adoption in various industries such as automotive and electronics. The interact analysis report also predicted more than 15% growth in the installation of cobots until 2028 [3]. Future achievements should address accuracy, productivity, efficiency, performance, safety, and autonomy [4].

Agile assembly planning is one of the critical elements in a DT for assisting the decision-making during the process development, e.g., assembly operation or maintenance process [5]. However, further investigations should be conducted in the context of HRC that may require efficient models, which can be contextualized based on human motion behaviors. In this regard, interpreting and predicting human actions during an assembly task can help to create a powerful collaboration (e.g., joint task handling). This may lead to efficient and adaptable robot programming using a semantic representation of human and robot models.

The DT that is made possible by the advancing digitalization of manufacturing is changing the boundaries and possibilities of production systems. Digital human models are now being simulated to represent realistic motions using wearable sensors in production systems. In this regard, wearable human motion capturing methods are employed to gather detailed human motion data, including eye, finger, and body movements, in given assembly tasks. However, describing the motion behaviors that are collected from these sensors in a meaningful and interpretable way is necessary for developing efficient models that can predict human intentions in the production environment. The analysis of epistemological and ontological aspects of HRC can be essential for semantic representation of human's activities. Semantic representation in the field of HRC serves not only to divide up the activities but also to ensure efficient collaboration. An HRC ontology can be considered to conceptualize models for motion representation, with benefits such as interaction data management, adopting model-driven approaches, and behavior reasoning with existing or new knowledge [6], [7].

The power of machine learning (ML) and artificial intelligence (AI) techniques for modeling and predicting human intentions in an industrial environment has been exploited for transferring human activities to, e.g., assistant

978-1-7281-2989-1/21/$31.00 ©2021 IEEE

robots [8]. In particular, for understanding human motion behaviors for teaching autonomous machines such as robots. This activity helps robots understand what human operators intend to do and how they perform – using sensor data.

Various demonstrations have been presented among the scientific communities regarding applying human models' representation in an HRC context. For instance, the works of [8]–[10] can be considered to observe the opportunities and challenges of intention prediction. Identifying the human intention using a semantic model could allow robots to overtake assembly tasks in the cases of uncertainties or errors of motion capture data. Implementing human and robot interfaces [11] for offline and online simulations is also considered for evaluating the models and methods' applicability and acceptability. However, the performance and efficiency of the models are inherent challenges, especially in task sharing and allocation [12].

The current paper investigates the opportunities and challenges for the semantic representation of human models in the operation phase of cyber-physical production systems (CPPS), specifically in hybrid assembly systems. It investigates the research question on how to create a methodology to automate task-sharing in HRC using a DT enabled by deep learning and semantic technologies. Therefore, the paper aims to present a concept based on the review of significant current and future challenges of human-robot interaction in various application domains such as AR/VR, multi-modal interaction, HRC, and big data demonstration for human-robot interaction in the production environment. The outcome of this review will set the foundation for the development of a knowledge-based DT concept of human-robot interaction, e.g., in hybrid assembly systems.

The remaining part of this article is organized as follows: Section II explores the aspects of predicting human-robot interaction, namely, ontology (semantic representation), DT, motion modeling and simulation, action and attention recognition, modeling human intentions, and attention, and prediction of human intentions using convolutional neural network (CNN). Section III presents and describes the proposed method for semantic representation of the human action models for predicting human-robot interaction in hybrid assembly systems. Finally, sections IV and V discuss the results and outline the conclusion and future work, respectively.

## II. ASPECTS OF PREDICTING HUMAN-ROBOT INTERACTION

This section explores the aspects that determine the effectiveness of predicting human intentions for human-robot interaction, namely: a) Ontology, b) DT, c) motion modeling and simulation, as well as d) action and attention recognition for attention prediction.

### *A. Ontology, semantic representation in Industry 4.0*

Industry 4.0 further develops the interactions between humans and technology, especially between humans and collaborative robots in HRC. In computer science, ontologies are referred to an explicit specification of a shared conceptualization in a machine-readable format [13] using description logic derived from first-order logic. The concepts, relations, and individuals of an ontology are stored in its knowledge-base. The integration of humans in CPPS and the necessary exchange of information, integration, and reciprocal learning between the participants is based on a framework called cyber-physical-socio space [7]. The aforementioned framework defines the criteria for optimal collaboration between humans and/or CPPS [7]. Frameworks such as AJAN focus on the evaluation of critical properties in dynamic HRC environments [14]. In the context of Industry 4.0, a landscape for human-machine and machine-machine cooperation is needed [15], in order to enable a flexible work organization. Therefore, an ontology-based system should be established that fulfills the criteria for cooperative manufacturing [16], [17]. Such a system must be based on a flexible knowledge base for the allocation of tasks in HRC as identified by [18]. This knowledge-base is necessary to superpose preset rules over probabilistic policies derived from optimization problems or ML algorithms [18]. This flexible, task-based ontology can be used in coordinating collaboration on multiple levels (system, work-cell, machine, worker) [19]. However, such a flexible task ontology must be filled with corresponding tasks. For this purpose, RoboEarth [20] and CORA [21] provided a knowledge base that allows robots to access existing tasks, extend them by learning, and find new solutions for existing problems. The task ontology then can be integrated into an Industry 4.0 domain ontology [22] with a corresponding sensors ontology [23] and connected to digital humans from human motion analysis and modeling. This integration allows reasoning over an extended knowledge-base in order to connect all parts of the process and to discover hidden knowledge.

### *B. Digital Twin*

Historically DT is centered at a simulation-based understanding which in recent years evolved further. Today, the German Industry 4.0 platform defines a DT as a digital representation of a product within a single or across multiple life-cycles using models, information, and data [24]. The virtual models on which the DT is based, are created from physical objects and digitally linked to them to simulate their behavior in real environments [25]. This understanding is also reflected in the three components of the DT [26], i) the physical entities in the real world, ii) the virtual models in the cyber world, and iii) the connected data that link the two worlds. This understanding can be extended to 9 dimensions, see Table 1, for the use of DT in HRC [27]–[31]. In HRC, a special focus lies on the dimension's physical environment, virtual environment, as well as task division.

TABLE 1. APPLICATION DIMENSIONS OF THE DIGITAL TWIN, BASED ON [27]–[31]

| Dimensions | Characteristics | | | |
|---|---|---|---|---|
| ***Goals*** | Information Retrieval | Information Analysis | Decision & Action Selection | Action implementation |
| ***User focus*** | Single User | Multi User | | |
| ***Physical Environment*** | Human | Collaborative | Machine | |
| ***Virtual Environment*** | Visualization | Interactions | Context | |
| ***Life Cycle Focus*** | One Phase | Multi Phase | | |
| ***Task division*** | independent | sequential | collaborative | |
| ***System Focus*** | Component | Subsystem | System | System of Systems |
| ***Data Sources*** | Measurements | Virtual Data | Knowledge | |
| ***Level of data integration*** | Manual | Semi-automatic | Fully automatic | |

A core concept of DTs in HRC is the fusion of the physical and virtual space over the planned distribution of tasks. The application areas of a DT range from use in i) product design [32], especially in simulation [33], to ii) production, to iii) maintenance [34], iv) networking and communication of plant manufacturers and operators [35] and v) HRC [27], as well as factory and plant optimization [36]. In the manufacturing domain, the DT consists of data on machine operators, material, equipment, tools, environment, behaviors, rules, and dynamics models, which can be optimized by the DT [37]. Therefore, digital and seamless information flows are fundamental. This requires a vertical integration across the machine life-cycle phases and horizontal integration within different manufacturing and cooperation steps. In predictive maintenance [34] and assembly, the collection of static and dynamic machine information (e.g., master data, dynamic task planning, and adaptive robot control) is particularly complex. This can result in wrong results in known analysis models due to the missing possibility to combine static, dynamic machine, and collaboration data. Therefore, it is essential to combine both current data and information from technical documentation (e.g., machine documentation, work safety regulations) with own analysis models [33]. This is especially important with regard to the requirements of a flexible HRC environment [38]. The sharing of the DT in shop floor control and the necessary communication between the worker and the robot are critical parts of a DT [38].

### C. *Motion modeling and simulation*

Human motion modeling and simulation are broad research domains in various application areas such as gaming, sport, health, human factors, and production systems. Human motion simulation has been used in production systems for virtual planning, verification, and ergonomic analysis. A motion modeling unit (MMU) representing various basic motions has been presented for the assembly of products in the MOSIM project [39], in which an open human motion modeling framework is developed and standardized. This framework helps to plan human motion sequences and tasks using agents. However, the framework does not include methods for integrating robot systems into the simulation loop.

In the context of MMUs, various techniques have been proposed for modeling human motions in manual and hybrid assembly environments. Statistical methods: e.g., functional principal component analysis (FPCA) [40], Gaussian mixture models (GMM) [17], and learning methods such as generative adversarial network (GAN) [41], and deep-learning or convolutional neural networks (CNN) [42] are some of the common approaches. Commercially available software such as AnyBody [43], and RAMSIS [44] also can be mentioned regarding human motion simulations in production systems. Furthermore, open-source tools such as Gazebo [45] and Unity3D [46] have been implemented with ROS interface.

### D. *Action and attention recognition for intention prediction*

Uncertainties in a fast-changing manufacturing system require operators to comprehend their behavior and assess their performance in near-real-time operations. [47] presented action recognition in manufacturing assembly using a multi-modal sensor fusion. In this context, recognition for human action is intended for self-performance improvement. In computer vision applications, e.g., surveillance or monitoring, motion capture data has been used to recognize human actions [48]. However, human action and attention recognition using computer vision are challenging due to occlusions and image quality [48]. In this regard, wearable motion tracking systems such as Xsens® have been employed. However, a robust model for discriminating joint spatial variation and temporal representation should be further investigated, particularly for human-robot hybrid working space. Collaborative human-robot interaction is bringing together Robotics, HCI, and Machine Learning to enable robot programming with knowledge from human experts. This involves collecting, integrating, and processing data from multi-sources (e.g., finger tracking, eye tracking, and body tracking) for intention modeling.

Once the human operator acts on the allocated task, the action performed can be extracted and presumed as the operator's intention. With the attention captured in the previous stage, these intentions are fed to the machine learning algorithm to build a model. The machine learning algorithm builds from a series of attention-intention order pairs and then ported to the robot. The robot behaves the same way the human operator behaved for a specific stimulus that resembles any attention. Modeling human intention has two dimensions: 1) behavioral modeling, where the cognitive aspect of human behavior is handled as it affects intention prediction from observed attention, and 2) data-intensive modeling of a causal relationship between intention for decision-based on observed attentions and motions.

#### 1) *Behavioral Modeling*

Multiple human factors must be considered to predict intentions rightfully from the observed human attentions or motions. A mixed environment of human movement for non-task issues, manufacturing equipment, and the robot will make human-robot communication a challenge. What cues/signals can be used to estimate/infer human intent, decision-making, and behavior reliably? How consistent are these cues across the human population, and what factors must be considered regarding human-machine interface heterogeneity? How are human cognition and behavior modulated by the environment (e.g., moving parts of the machine, moving robots)? One potential agenda for human-robot collaborative interactions research/engineering is to explore more profoundly human cognition and behaviors relevant in mixed-production environments and develop human state monitoring systems that can track signals that facilitate human-robot coordination.

#### 2) *Modeling Relationships of Attention and Intention*

Modeling relationships between attention and the intended task involves identifying tools and methodologies that enable HRC with the outcome of robot programming from human expertise, followed by evaluating and verifying the methodology on a selected domain of assembly/maintenance use-case. The methodology includes proposing a machine learning approach that utilizes CNN. The CNN enables prediction of intention for a given observation of attention in order that a model could learn the patterns that enable it to derive a relationship between the observed attention and the expected intention. Evaluation verifies the prediction model's effectiveness as a basis for developing the simple explanation for making forecasts using past data of human expertise.

Based on the gathered data inputs from human motion captures and use case stories, the human intention are modeled by investigating a suitable artificial intelligence or machine learning technique, CNN in this case. This approach has to answer what the impact of this task is concerning generalizability, repeatability, accuracy, and reliability.

*3) Human-Robot Interface modeling*

A user-friendly interface is crucial for improving usability, information flow, and human confidence for using robots while working next to the robot. The task descriptions and definitions of actions define the role of a robot based on the context of what humans intend to do.

Robot programming using a human expert is achieved through CNN in the prediction of human intention from human attention information obtained from eye tracking visual input, human action from external camera or sensors attached to nodes in production or assembly lines together with feedback data (intention) from external camera or sensors connected to nodes in production or assembly lines as well as more heuristic data from inputs of experts (human in the loop).

## III. METHODS FOR SEMANTIC REPRESENTATION AND INTENTION PREDICTION IN HRC

The presented methodology uses a knowledge-based DT which combines human motion interpretation and representation with CNN and semantic technologies. As seen in Fig. 1, the methodology consists of three core parts, the human user, the cobot, and the DT.

The core concept is based on HRC leveraging AI in order to use data feeds from sensors installed on humans that track the behavior of the human user. The DT will then use this sensor data in order to predict the intention of the human user from the actions associated with attention. Fig. 1 shows how robots and humans can collaborate on a task in such a way that when the human focuses on a particular component, the robot will predict the focus location based on where this component is placed. Then it will reason over in its knowledge-base, consisting of domain ontologies or semantic representations of tasks, in order to select the best matching one and execute that task. For example, for picking a part, its location and orientation should be known to decide what actions should follow, which can be moving to assembly location, aligning with the correct orientation, and fixing. The knowledge-based DT is therefore used to predict movements optimized to the user behavior by using the user's attention, location and descriptions of the task. By optaining the location of the parts and their target placement the DT can derive the necessary actions that are associated with these locations from the knowledge-base and execute the associated command on the robot to perform the desired action. The knowledge-based DT consists of 3 main parts.

### A. *The knowledge-base*

The knowledge base is based on the CORA ontology. The *RobotMotion* class defines the movements of the robots. This class can be further specified by using types of motions needed (e.g., *RobotCarrying, RobotPick, RobotHold,* and others.) [49]. Further sensor inputs can be taken from the Sensor Ontology. The connection of CORA with the Sensor Ontology [23] can be achieved by linking the *robotSensingPart* class with the Sensor class from the Sensor ontology. A human can be modeled with the *Agent* class in the CORA ontology. The connected ontologies, including real-time sensor data and assembly/maintenance workflows, enable the discovery of hidden knowledge in combination with an AI reasoning algorithm. This knowledge can then be used further in a DT environment.

### B. *Human motion generation and intention prediction in HRC*

A wearable and IMU-based human motion capture system is employed to capture human activity during the hybrid assembly processes. In this regard, motions from the full-

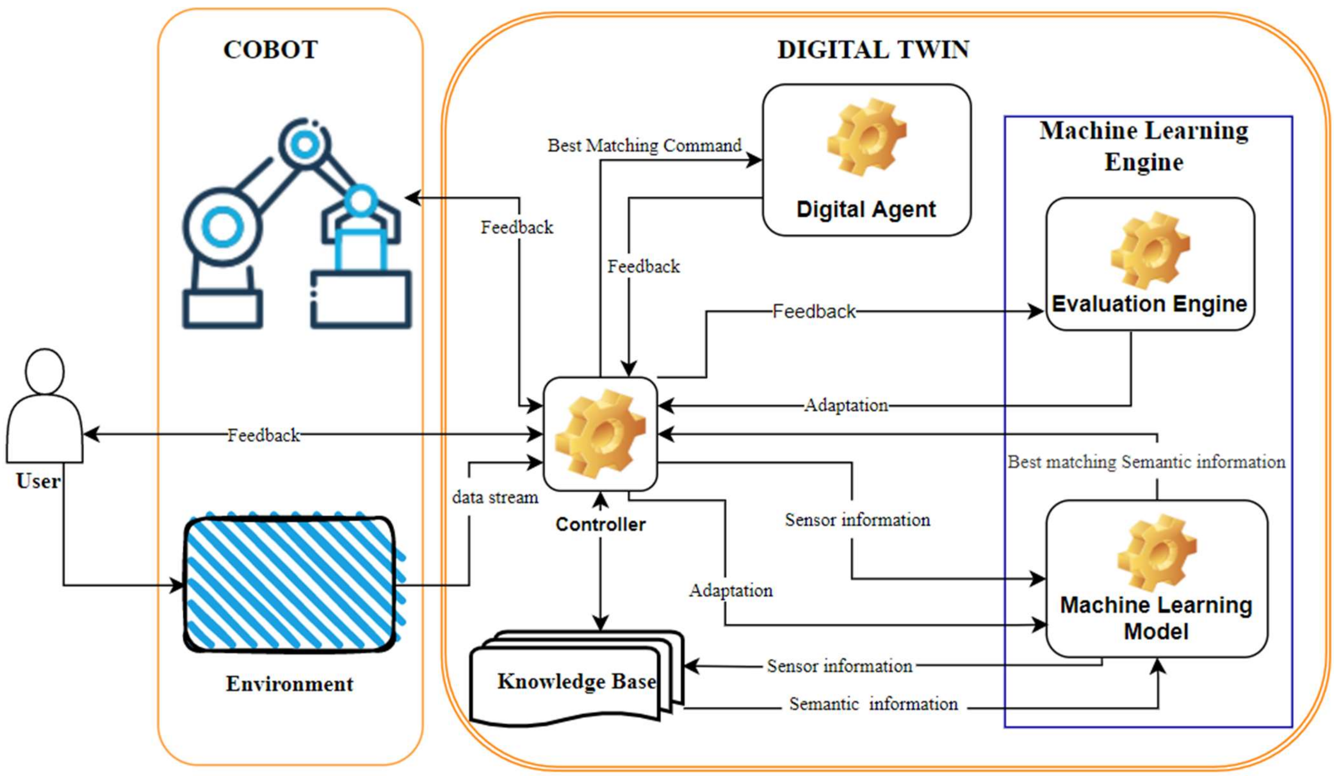


Fig. 1. Knowledge-based digital twin concept of the ontological human intention prediction for HRC.

body, finger, and eye focus are targeted by employing multi systems. The generated motion comprises vector positions and quaternion rotations for human body joints. However, in the eye-tracking system, surface tracking using markers (e.g., *AprilTags*) is required to identify an object's position and orientation with respect to a fixed reference frame. The motion capture systems provide real-time streaming for joint positions and orientations. This poses data (i.e., positions and rotations) are used as an input for intention prediction. The intention prediction model applies AI models to compute the possible motion sequences and ontologically defined task descriptions in the context of HRC.

### C. *Knowledge-based Digital twin*

All described parts are then combined in the knowledge-based DT framework. The knowledge-base itself is designed as a flexible task ontology via a Python interface in the semantic knowledge-base using the Python library *Owlready 2*. The knowledge-base enables the construction of a DT with corresponding reasoning and simulation algorithms. These reasoning and cognitive capabilities are defined using the ML models. The ML component computes sensor information and best matches semantic information by comparing computation results with the corresponding semantic information from the knowledge-base. The controller components act as a broker between the environment that feeds data stream to the digital twin, machine learning and digital agent in such a way that it feeds sensor information scanned from the environment to the ML component, retrieves best matching semantic information from the ML component and forwards it to the digital agent so that it executes the task corresponding to the demand it receives. In other words, the digital agent is a virtual equivalent of the physical cobot. Moreover, the controller also routes the feedback and adaptation information among components and between the human and digital twin. The MOSIM interface is applied to simulate the digital human model in a virtual environment (e.g., Unity3D) based on motion capture data (e.g., for real-time interface) and physics models (e.g., for simulation). The Unity3D environment simulates both humans' and robots' motion. In this case, the ROS interface controls the robot's motion. The semantic representation in a knowledge-base, e.g., object IDs, motion types, geometry constraints, joint types, and other parameters, is utilized by the DT.

## IV. RESULTS AND DISCUSSION

The assembly system based on Bosch TS/2 conveyors at the University of Siegen serves as a use-case of hybrid working space (see Fig, 3). Here, the methodology of a knowledge-based DT is simulated and compared to real industry settings with the help of a cobot and an existing assembly and maintenance demonstrator. In the case of the assembly use-case, the worker assembles a frame consisting of aluminum profiles in collaboration with a cobot. The movements performed by the worker are: Looking at the object, picking it, moving it, and placing it at the assembly location. Only hand and gaze motions are considered to describe human activities during an assembly operation in the current investigation. The complete description of sensors and motion data metrics is shown in Table 2.

The system interface for human motion data capturing requires calibration of each system before they are combined in the DT. Gaze tracker can calibrate itself using the AI technology automatically, which makes it ready for use immediately. In contrast, the body motion tracking system requires multiple trials to obtain high-quality motion data. In this aspect, the MVN analyze tool provides the quality confidence of the system calibration. Therefore, high-quality calibration is achieved during the motion capture. Thus, the sensing system is employed for motion capturing. A user has to pay attention to sensor attachments to the body parts to avoid joint data loss or motion artifacts.

TABLE 2 SUMMARY OF MOTION DATA METRICS

| **Parameter** | ***Sensor type*** | ***Specification (Update rate)*** |
|---|---|---|
| Gaze tracking | Pupil Invisible | 200Hz |
| Body joint tracking | Xsens Awinda | 60Hz |
| Finger tracking | Manus VR Prime II | 90Hz |
| Robot | Universal robot UR5 | 50Hz |

The assembly operation is performed in the lab environment, and motion capture systems are employed to capture and annotate the desired motion types such as look, reach, pick, move and assemble or place. These motions are decomposed into motion clips, and ontological representations are assigned from the *RobotMotion* class. These specific motions that are defined in the *RobotMotion* class are used as input for the intention prediction model.

In the simulation of the use-case, based on the knowledge-based DT, parts position and type are identified by using markers (Fig. 2.c). The gaze data observations are fixations of eye focus on a two-dimensional image (pixel-wise). Assuming human attention is on the given object, the ML model utilizes this information to combine and refine it with the knowledge-base and reasoning for recognizing the position and orientation of the target object. The worker’s focus could be on one of the marker-labeled assembly parts. The model utilizes these labels to predict the action that should be performed, e.g., reach, move, or focus. The prediction model analyzes the attention and action of the worker. The worker's intention is the best matching semantic information based on the reasoning over the knowledge-base. The intention is then further used in the Digital Agent. These semantic descriptions are the human-understandable contexts that simplify the understandability of DT processes.

In DT, human motion simulation is implemented using the MOSIM framework [39], and the actual motion is captured in the real environment (see Fig. 2 and Fig. 3). Key performance indicators such as decision and action time are measured for both simulation and actual experiments. This measurement is helpful for the preliminary investigation of time reductions which can be extended further for knowledge evolution. The activity sequences are hierarchically controlled according to [50]. Accordingly, the sequence of operations is to gaze at an object – reach for picking and move to an assembly position. The posture blending service [51] is applied for motion transition from one activity to the other. The actual motion is measured using a temporal scale. The process of identifying object type, position, and orientation configuration is considered a decision.

Fig. 2 illustrate the process of measuring the effectiveness of the human operator and cognitive workloads in both virtual and real physical environment. In order to test if the approach is in principle able to provide realistic simulated times for pick and place shop floor activities, simulated and measured times (i.e., time difference between starting and end of an action) have been compared for the assembly setup depicted in Fig. 3.

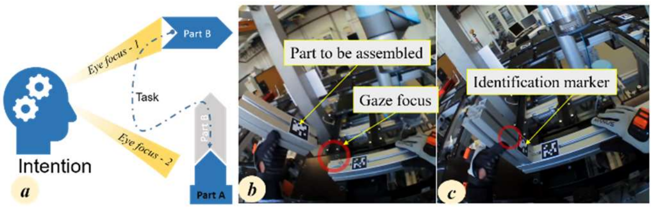


Fig. 2. Process illustration for motion capture-based (e.g., gaze) on intention prediction scenario and proof of the concept. In (a), the human looks at part B with *eye focus-1* and then finds the target position by moving into *eye focus 2*. Then decides the assembly position and orientation. Finally, the robot picks the assembled product and places it in mobile storage.

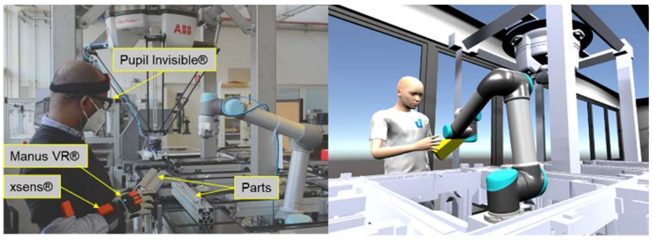


Fig. 3. Demonstration of the digital hub for ontological human intention prediction in HRC.

In the simulation process, the decision and action process is defined using reasoning and knowledge-base. However, the actual motion is adaptable to the situation, which changes for each cycle.

According to the simulation and measurements from the lab (see Table 3), the simulation time, in general, is faster than the actual motion captured by the motion capture.

TABLE 3 PERFORMANCE EVALUATION

| **Activities** | ***Measured time[ms]*** | ***Simulation time [ms]*** |
|---|---|---|
| Look at | 550 | 166 |
| Reach | 1284 | 633 |
| Decision of part type | 2550 | 360 |
| Move to assemble | 4183 | 267 |
| Move to home | 1283 | 66 |

For example, looking at an object is 30.18%, and reaching an object is 49.29% faster than the actual motion. Similarly, the time to retrieve reasoning logic from an existing knowledge base is by far faster in the simulation than in the measurement. The discrepancy is especially apparent for the decision activity. It is assumed that the model for the decision action deviates too much from reality, so that identifying the orientation and specific assembly position takes more time for the human worker than the knowledge-based DT simulation. Therefore, it is essential to model the human attention-intention relationship for interpreting human action by semantic representations. Human motion behaviors that are acquired from gaze and body movements serve as an input for predicting the next sequences of operations. This sequence of operations can be picking up a part, identifying the placement target, moving the object to the target location, and placing or assembling the object in the target position as it is depicted in Fig. 2. In this regard, based on the predicted action, the robot is designated to execute predefined motions such as pick and place. In the proof of concept, a robot is expected to understand what the human expert wants to accomplish during an assembly operation and then complements the worker.

In general, the semantic representation of basic motions requires further investigations for enabling evolutionary knowledge acquiring and reasoning capabilities in the HRC context. This approach could benefit users in cost reduction, human acceptance, enhancing safety, improving efficiency, and reducing errors that are caused due to fatigues on repetitive tasks.

## V. CONCLUSION AND FUTURE RESEARCH AGENDA

This paper provides an AI-enhanced approach and knowledge-based DT, which explores the interaction of human users with robots in a collaborative production environment so that the actions are derived from human attention predicted from attention captured via sensors and their associated information that is mined from semantic ontologies. The preliminary evaluation for pick-and-place showed a promising reduction of time for assembly tasks. The future research in HRC using DT involves more indicators such as effectiveness, decision time, and action time analytically. Future work needs a focus on human understandability and explainability to enable the evolution of the knowledge base and comply with the EU guidelines on trustworthy AI [52]. In subsequent development steps, the methodology, in particular the knowledge base, should be expanded to include a possibility for humans and machines for reciprocal learning in order to learn new tasks on the job. This also needs the consideration of individual competence levels of human workers in the DT and the cooperation process. The other aspect is the workforce's mobility, leading to human experts' operations from diverse cultural backgrounds as attention manifestation differs across cultures. An extension to this could be "reciprocal learning," in which other robots (or new employees) of a company could learn from robots.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support by the Federal Ministry of Education and Research of Germany within the ITEA3 project MOSIM (grant number: 01IS18060AH), by the European Regional Development Fund (EFRE) within the project SMAPS (grant number: 0200545), and by Horizon 2020 Programme of the European Commission within the project KnowlEdge (grant No. 957331).

## DECLARATION

The authors declare that the human motion data is stored and processed with the consent of the participants only for the purpose of the study.

## REFERENCES

[1] K. Ramirez-Amaro, Y. Yang, and G. Cheng, “A survey on semantic-based methods for the understanding of human movements,” *Robot. Auton. Syst.*, vol. 119, pp. 31–50, Sep. 2019, doi: 10.1016/j.robot.2019.05.013.

[2] MarketsandMarkets™ Research Private Ltd, “Collaborative Robot (Cobot) Market by Payload, Component (End Effectors, Controllers), Application (Handling, Assembling & Disassembling, Dispensing,

Processing), Industry (Electronics, Furniture & Equipment), and Geography – Global Forecast to 2026," *MarketsandMarkets$^{TM}$ Research Private Ltd*. https://www.marketsandmarkets.com/Market-Reports/collaborative-robot-market-194541294.html (accessed Jan. 19, 2021).

[3] Interact analysis, "The Collaborative Robot (Cobot) Market - Infographic - Interact Analysis." https://www.interactanalysis.com/the-collaborative-robot-cobot-market-infographic/ (accessed Apr. 23, 2021).

[4] L. D. Evjemo, T. Gjerstad, E. I. Grøtli, and G. Sziebig, "Trends in Smart Manufacturing: Role of Humans and Industrial Robots in Smart Factories," *Curr. Robot. Rep.*, vol. 1, no. 2, pp. 35–41, Jun. 2020, doi: 10.1007/s43154-020-00006-5.

[5] P. Burggräf, M. Dannapfel, T. Adlon, and C. Fölling, "Hybrid approach to agile assembly planning – Empirical evaluation of the industrial practice," *Procedia CIRP*, vol. 93, pp. 1170–1175, Jan. 2020, doi: 10.1016/j.procir.2020.02.140.

[6] N. Köster, S. Wrede, and P. Cimiano, "An Ontology for Modelling Human Machine Interaction in Smart Environments," in *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Cham, 2018, pp. 338–350. doi: 10.1007/978-3-319-56991-8_25.

[7] F. Ansari, M. Khobreh, U. Seidenberg, and W. Sihn, "A problem-solving ontology for human-centered cyber physical production systems," *CIRP J. Manuf. Sci. Technol.*, vol. 22, pp. 91–106, Aug. 2018, doi: 10.1016/j.cirpj.2018.06.002.

[8] D. Aarno and D. Kragic, "Motion intention recognition in robot assisted applications," *Robot. Auton. Syst.*, vol. 56, no. 8, pp. 692–705, Aug. 2008, doi: 10.1016/j.robot.2007.11.005.

[9] S. Li, L. Zhang, and X. Diao, "Deep-Learning-Based Human Intention Prediction Using RGB Images and Optical Flow," *J. Intell. Robot. Syst.*, vol. 97, no. 1, pp. 95–107, Jan. 2020, doi: 10.1007/s10846-019-01049-3.

[10] O. Dermy, F. Charpillet, and S. Ivaldi, "Multi-modal Intention Prediction with Probabilistic Movement Primitives," in *Human Friendly Robotics*, 2019, pp. 181–196.

[11] T. B. Ionescu and S. Schlund, "Programming cobots by voice: A human-centered, web-based approach," *Procedia CIRP*, vol. 97, pp. 123–129, Jan. 2021, doi: 10.1016/j.procir.2020.05.213.

[12] C. Schmidbauer, S. Schlund, T. B. Ionescu, and B. Hader, "Adaptive Task Sharing in Human-Robot Interaction in Assembly," in *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Dec. 2020, pp. 546–550. doi: 10.1109/IEEM45057.2020.9309971.

[13] R. Studer, R. Benjamins, and D. Fensel, "Knowledge Engineering: Principles and Methods," *Data Knowl Eng*, 1998, doi: 10.1016/S0169-023X(97)00056-6.

[14] A. Antakli *et al.*, "Agent-based Web Supported Simulation of Human-robot Collaboration:," in *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, Vienna, Austria, 2019, pp. 88–99. doi: 10.5220/0008163000880099.

[15] I. Grangel-González *et al.*, "The industry 4.0 standards landscape from a semantic integration perspective," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sep. 2017, pp. 1–8. doi: 10.1109/ETFA.2017.8247584.

[16] A. R. Sadik and B. Urban, "An Ontology-Based Approach to Enable Knowledge Representation and Reasoning in Worker–Cobot Agile Manufacturing," *Future Internet*, vol. 9, no. 4, Art. no. 4, Dec. 2017, doi: 10.3390/fi9040090.

[17] M. Manns, K. Fischer, H. Du, P. Slusallek, and K. Alexopoulos, "A new approach to plan manual assembly," *Int. J. Comput. Integr. Manuf.*, pp. 1–14, Apr. 2018, doi: 10.1080/0951192X.2018.1466396.

[18] S. El Zaatari, M. Marei, W. Li, and Z. Usman, "Cobot programming for collaborative industrial tasks: An overview," *Robot. Auton. Syst.*, vol. 116, pp. 162–180, Jun. 2019, doi: 10.1016/j.robot.2019.03.003.

[19] A. M. Djuric, R. J. Urbanic, and J. L. Rickli, "A Framework for Collaborative Robot (CoBot) Integration in Advanced Manufacturing Systems," *SAE Int. J. Mater. Manuf.*, vol. 9, no. 2, Art. no. 2016-01–0337, Apr. 2016, doi: 10.4271/2016-01-0337.

[20] M. Tenorth, A. C. Perzylo, R. Lafrenz, and M. Beetz, "Representation and Exchange of Knowledge About Actions, Objects, and Environments in the RoboEarth Framework," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 643–651, Jul. 2013, doi: 10.1109/TASE.2013.2244883.

[21] S. Balakirsky *et al.*, "Towards a Robot Task Ontology Standard," presented at the ASME 2017 12th International Manufacturing Science and Engineering Conference collocated with the JSME/ASME 2017 6th International Conference on Materials and Processing, Jul. 2017. doi: 10.1115/MSEC2017-2783.

[22] V. R. S. Kumar *et al.*, "Ontologies for Industry 4.0," *Knowl. Eng. Rev.*, vol. 34, ed 2019, doi: 10.1017/S0269888919000109.

[23] A. Sheth, C. Henson, and S. S. Sahoo, "Semantic Sensor Web," *IEEE Internet Comput.*, vol. 12, no. 4, pp. 78–83, Jul. 2008, doi: 10.1109/MIC.2008.87.

[24] H. Hirsch-Kreinsen, U. Kubach, and Uwe Rainer Stark, "Key themes of Industrie 4.0." https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/acatech-keythemes-industrie-4-0.html (accessed Mar. 24, 2021).

[25] J. D. Hochhalter, W. P. Leser, J. A. Newman, E. H. Glaessgen, V. K. Gupta, and V. Yamakov, "Coupling Damage-Sensing Particles to the Digitial Twin Concept," *Natl. Aeronaut. Space Adm. Langley Res. Cent.*, p. 15, 2014.

[26] M. Grieves, "Digital Twin: Manufacturing Excellence through Virtual Factory Replication," Mar. 2015.

[27] J.-F. Uhlenkamp, K. Hribernik, S. Wellsandt, and K. Thoben, "Digital Twin Applications : A first systemization of their dimensions," *2019 IEEE Int. Conf. Eng. Technol. Innov. ICEITMC*, 2019, doi: 10.1109/ICE.2019.8792579.

[28] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Trans. Ind.*

*Inform.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019, doi: 10.1109/TII.2018.2873186.

[29] X. Ma, F. Tao, M. Zhang, T. Wang, and Y. Zuo, "Digital twin enhanced human-machine interaction in product lifecycle," *Procedia CIRP*, vol. 83, pp. 789–793, Jan. 2019, doi: 10.1016/j.procir.2019.04.330.

[30] E. Negri, L. Fumagalli, and M. Macchi, "A Review of the Roles of Digital Twin in CPS-based Production Systems," *Procedia Manuf.*, vol. 11, pp. 939–948, Jan. 2017, doi: 10.1016/j.promfg.2017.07.198.

[31] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital Twin in manufacturing: A categorical literature review and classification," *IFAC-Pap.*, vol. 51, no. 11, pp. 1016–1022, Jan. 2018, doi: 10.1016/j.ifacol.2018.08.474.

[32] S. Hagemann, A. Sünnetcioglu, and R. Stark, "Hybrid Artificial Intelligence System for the Design of Highly-Automated Production Systems," *Procedia Manuf.*, vol. 28, pp. 160–166, Jan. 2019, doi: 10.1016/j.promfg.2018.12.026.

[33] A. Barthelmey, E. Lee, R. Hana, and J. Deuse, "Dynamic digital twin for predictive maintenance in flexible production systems," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, Oct. 2019, vol. 1, pp. 4209–4214. doi: 10.1109/IECON.2019.8927397.

[34] F. Ansari, S. Nixdorf, and W. Sihn, "Insurability of Cyber Physical Production Systems: How Does Digital Twin Improve Predictability of Failure Risk?," *IFAC-Pap.*, vol. 53, no. 3, pp. 295–300, Jan. 2020, doi: 10.1016/j.ifacol.2020.11.063.

[35] M. Dietz, B. Putz, and G. Pernul, "A Distributed Ledger Approach to Digital Twin Secure Data Sharing," in *Data and Applications Security and Privacy XXXIII*, Cham, 2019, pp. 281–300. doi: 10.1007/978-3-030-22479-0_15.

[36] Y. Zheng, S. Yang, and H. Cheng, "An application framework of digital twin and its case study," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 3, pp. 1141–1153, Mar. 2019, doi: 10.1007/s12652-018-0911-3.

[37] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About The Importance of Autonomy and Digital Twins for the Future of Manufacturing," *IFAC-Pap.*, vol. 48, no. 3, pp. 567–572, Jan. 2015, doi: 10.1016/j.ifacol.2015.06.141.

[38] A. A. Malik and A. Bilberg, "Digital twins of human robot collaboration in a production setting," *Procedia Manuf.*, vol. 17, pp. 278–285, Jan. 2018, doi: 10.1016/j.promfg.2018.10.047.

[39] MOSIM, *Github repository codes of MOSIM framework*. Daimler Group, 2021. Accessed: Feb. 23, 2021. [Online]. Available: https://github.com/Daimler/MOSIM

[40] M. Manns, M. Otto, and M. Mauer, "Measuring motion capture data quality for data driven human motion synthesis," *Procedia CIRP*, vol. 41, pp. 945–950, 2016.

[41] L.-Y. Gui, K. Zhang, Y.-X. Wang, X. Liang, J. M. F. Moura, and M. Veloso, "Teaching Robots to Predict Human Motion," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Oct. 2018, pp. 562–567. doi: 10.1109/IROS.2018.8594452.

[42] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Trans. Graph. TOG*, vol. 35, no. 4, p. 138, 2016.

[43] M. Peters and S. Wischniewski, "Evaluation of Physical Workload Using Simulation Versus Motion Capturing Data for Biomechanical Models," *DHM2020*, pp. 380–387, 2020, doi: 10.3233/ATDE200046.

[44] "HS Group - Produkte Mobility - RAMSIS Automotive - Effiziente Innenraumentwicklung – ohne Nachbessern." http://www.human-solutions.com/mobility/front_content.php?idcat=252 (accessed Oct. 18, 2017).

[45] L. He, P. Glogowski, K. Lemmerz, B. Kuhlenkötter, and W. Zhang, "Method to Integrate Human Simulation into Gazebo for Human-robot Collaboration," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 825, p. 012006, May 2020, doi: 10.1088/1757-899X/825/1/012006.

[46] T. B. Tuli and M. Manns, "Real-Time Motion Tracking for Humans and Robots in a Collaborative Assembly Task," *Proceedings*, vol. 42, no. 1, Art. no. 1, 2019, doi: 10.3390/ecsa-6-06636.

[47] Md. Al-Amin *et al.*, "Action Recognition in Manufacturing Assembly using Multimodal Sensor Fusion," *Procedia Manuf.*, vol. 39, pp. 158–167, Jan. 2019, doi: 10.1016/j.promfg.2020.01.288.

[48] S. Vantigodi and R. V. Babu, "Real-time human action recognition from motion capture data," in *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Dec. 2013, pp. 1–4. doi: 10.1109/NCVPRIPG.2013.6776204.

[49] J. I. Olszewska *et al.*, "Ontology for autonomous robotics," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2017, pp. 189–194. doi: 10.1109/ROMAN.2017.8172300.

[50] F. Gaisbauer, J. Lehwald, P. Agethen, J. Sues, and E. Rukzio, "Proposing a Co-simulation Model for Coupling Heterogeneous Character Animation Systems:," in *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Prague, Czech Republic, 2019, pp. 65–76. doi: 10.5220/0007356400650076.

[51] F. Gaisbauer, J. Lehwald, J. Sprenger, and E. Rukzio, "Natural Posture Blending Using Deep Neural Networks," in *Motion, Interaction and Games*, Newcastle upon Tyne United Kingdom, Oct. 2019, pp. 1–6. doi: 10.1145/3359566.3360052.

[52] E. LEMONNE, "Ethics Guidelines for Trustworthy AI," *FUTURIUM - European Commission*, Dec. 17, 2018. https://ec.europa.eu/futurium/en/ai-alliance-consultation (accessed Apr. 23, 2021).