

Intention Recognition and Communication for Human-Robot Collaboration

Hadi Banaee^{1,*}, Franziska Klügl¹, Fjollë Novakazi¹ and Stephanie Lowry¹

¹Centre for Applied Autonomous Sensor Systems (AASS), Örebro University, Fakultetsgatan 1, 701 82 Örebro, Sweden

Abstract

Human-robot collaboration follows rigid processes, in order to ensure safe interactions. In case of deviations from predetermined tasks, typically, processes come to a halt. This position paper proposes a conceptual framework for intention recognition and communication, enabling a higher granularity of understanding of intentions to facilitate more efficient and safe human-robot collaboration, especially in events of deviations from expected behaviour.

Keywords

Intention recognition, intention granularity, human-robot collaboration, human-robot communication

1. Introduction

The promise of Industry 4.0 is a paradigm shift towards interconnected manufacturing systems that leverage advanced technologies such as artificial intelligence and automated technologies to optimise production processes. As removing humans from manufacturing is not a viable option [1], the focus is instead on finding ways to integrate humans and machines to work collaboratively and efficiently. The reason for the development of mixed human-robot teams in various application areas, such as assembly and transportation, is to combine the flexibility, adaptability, and problem-solving skills of humans with the precision and efficiency of robots. In the past, this has been accomplished by organising activities in highly controlled environments where robots and humans are kept apart by enforcing safeguards, such as maintaining spatial or temporal distances between them and assigning tasks or objectives to each agent, whether they are human or robotic.

For a successful transition into an industry 4.0 and subsequently 5.0 setting, robots need to have the ability to coexist and interact with humans in both physical and social settings. This entails creating a safe environment for humans where they can perceive, interpret, and respond to the actions and intentions of robots, and vice versa. This, though, becomes a challenge in the event of deviations from the established processes or assigned tasks. While a simple intention recognition (IR) approach places the responsibility exclusively on the robotic agent to detect

HHAI-WS 2024: Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 10–14, 2024, Malmö, Sweden

*Corresponding author.

✉ hadi.banaee@oru.se (H. Banaee); franziska.klugl@oru.se (F. Klügl); fjolle.novakazi@oru.se (F. Novakazi); stephanie.lowry@oru.se (S. Lowry)

🆔 0000-0002-9607-9504 (H. Banaee); 0000-0002-1470-6288 (F. Klügl); 0000-0001-6381-2346 (F. Novakazi); 0000-0003-3788-499X (S. Lowry)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

actions and adapt their behaviour accordingly, a more effective approach would integrate a hierarchical IR with communication strategies to detect and clarify the reasons for deviations. This would allow the sharing of responsibilities to ensure a seamless continuation of teamwork and ultimately, productivity in the evolving industry landscape.

To achieve this goal, we argue that there is a need to identify the appropriate level of granularity, sequence, and abstraction of tasks as intentions for an IR and communication framework in human-robot teams. Therefore, this position paper explores alternative solutions driven by an intention recognition approach to managing a human-robot collaboration (HRC), or mixed-agent teams, to enable more effective and seamless interactions between humans and robots that do not require rigid safety constraints, but instead rely on communication strategies to enable the handling of deviations and enhanced reactions.

2. Related Work

For HRC to be successful, the robot must analyse and understand human intentions, as well as effectively convey its own goals. Intention refers to the mental state or attitude of aiming to do or achieve something [2]. It involves a conscious decision or plan to act, often driven by a purpose or goal. Hence, understanding intent involves identifying the occurring activity, inferring the objectives of the task, and predicting the next actions. In an HRC context, IR refers to the process of identifying the intentions of agents, whether they are human or robotic, by examining their sequences of actions and/or analysing the impact of their actions on the state of the environment [3]. In other words, IR can be defined as the process of inferring the intentions of an agent by analysing their behaviour [4]. Hence, in this paper, we define IR as the capacity to identify the particular goal being pursued by precisely discerning the exact course of action that is taken [5].

The current research body presents various methodologies to address this challenge. One of the driving ideas is that robots need to be more proactive in their interactions with humans, which puts IR at the centre of possible solutions. For example, Tong and colleagues [6] explored a method for proactive human IR based on context changes and triggers, utilising vision-based technologies. However, a single feature reliance can make such a system unreliable. Other research looked at how two-way IR and communication affect HRC to figure out what people are going to do, like pick up an object, and help team members work together better [7]. In yet another attempt to explore task sequences as a source for IR, the trajectory of human actions was analysed to predict the subsequent actions [8], thus breaking down intentions into smaller actions. A different study utilised an inverse planning approach to IR, putting forward a logic-based approach for fully observable systems. This approach inferred the human's goal by observing a sequence of actions and still allowed for small deviations in the sequence where the human perform their actions in a different order [9].

These approaches lack addressing granularity to recognise intermediate intentions hierarchically. Moreover, deviations are merely studied by moving from a limited set of actions to a single intention as the overall goal. This creates a challenge for a complex HRC since inferring only the final goal as an intention from the atomic actions might not lead to enhanced decision-making.

Typically, in HRC, communication (whether explicit or implicit) is often used to convey an

intention from the robot to the human [10]. However, the framework presented here aims to facilitate purposeful communication in the event of a deviation. This allows the robot to adjust its decision-making and subsequent reaction when interacting with a human agent.

3. Conceptualisation of IR Framework

In the context of mixed human-robot teams, intention recognition is crucial to achieving two main challenges: 1) safety enhancement by avoiding potential harm or hazards, and 2) efficiency by providing proper support and assistance from the agents. Our conceptual framework for intention recognition and communication addresses these challenges by focusing on aspects of the appropriate level of granularity, sequentiality of actions, and deviation in performing the tasks. To formalise the conceptual framework, first, we clarify the input of the framework in such a context. Then, we specify in which situation the proposed intention recognition framework will be triggered.

In a human-robot environment, the agents (i.e., humans and/or robots) are asked to fulfill a certain goal by performing a sequence of predetermined tasks. Each of the predetermined tasks might be either at the level of primitive observable actions (e.g., pick a tool) or at a higher level of abstraction (e.g., repair the machine). The agents should have a shared understanding of the tasks and the overall team goal. However, situations may arise when an agent deviates from the shared plan. In such a situation, the other agents need to recognise the intentions behind the deviation and react accordingly, to avoid unnecessary interruptions to the workflow.

3.1. Conceptual Model

Our proposed conceptual model takes into account three following aspects:

Temporal sequence of actions: An intermediate intention may not be directly observable by a single action. Therefore it is important to consider the sequence of actions to infer the intermediate intention behind them. For example, in the kitting task below, a robot follows the sequence of the move–pick–move–place actions to fulfill the intermediate intention of “collecting one item”.

Granularity of intentions in a hierarchical structure: One can consider primitive actions as the first level of intentions. The combination of these intentions can then lead to inferring a higher abstraction level of intermediate intentions. But these intermediate intentions can also be combined to infer even higher levels of abstractions as intermediate intentions. The highest level of abstraction will be then the shared intention of the team as the overall goal. Note that the granularity of the intentions is not fixed and can be adjusted based on the context and the requirements of the task.

Deviation from the predetermined tasks: In the process of inferring the intermediate intentions toward achieving the goal, the agents should be able to detect a deviation from the shared plan. The deviation is either detected by observing an unexpected sequence of primitive

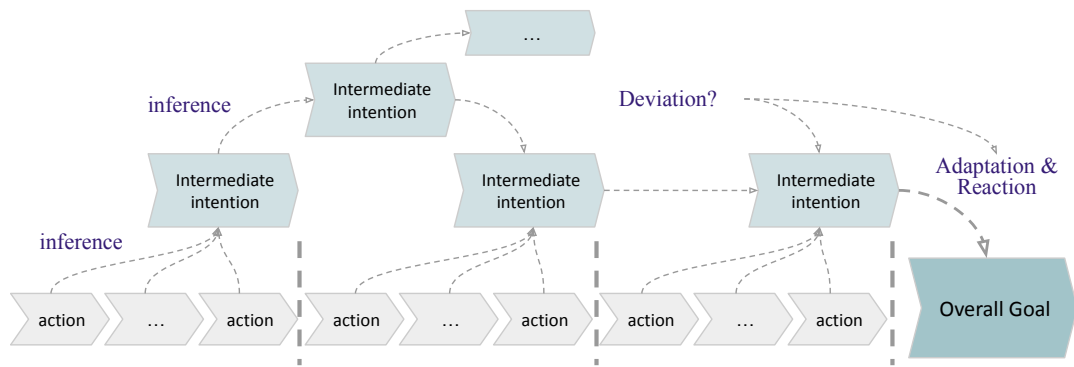


Figure 1: Conceptual model of the aspects behind the framework.

actions or is identified after inference of intentions, i.e., as a deviation in the intermediate intentions.

Figure 1 illustrates the conceptual model of the aspects behind the framework, considering the sequence of actions, the hierarchy of intermediate intentions, and detection.

3.2. Framework Processes

Based on the concepts illustrated in Fig 1, the following processes are required for the overall adaptive intention recognition and communication framework.

1) Observation and Context Analysis: The agents follow the shared tasks and observe the primitive actions of the other agents, considering the context and the goal of the task. These observable actions will be the ingredients of the IR process.

2) Intention recognition: The sequence of observed actions is analysed to infer the intermediate intentions of the other agents, considering the temporality of the actions and the granularity of the lower-level intentions. These inferences can be made in higher levels of intention abstractions until leading to meaningful decision-making.

3) Deviation Detection: The agents recognise the deviation from the shared tasks by either observing the sequence of actions performed by the other agent, or at a higher level of abstraction, by recognising the change in the intermediate intentions.

4) Adaptation & Reaction: Based on the deviation detection, and the expected intermediate intentions based on the shared tasks, the agents can make a decision based on the higher-level abstract intention behind the deviation and react accordingly to the situation. The proper reaction depends on the level of abstraction of the intention and the context of the task. This component can either lead to a direct execution of the reaction or may infer further adaptation based on the complexity of the situation, which leads to the next step (i.e., communication).

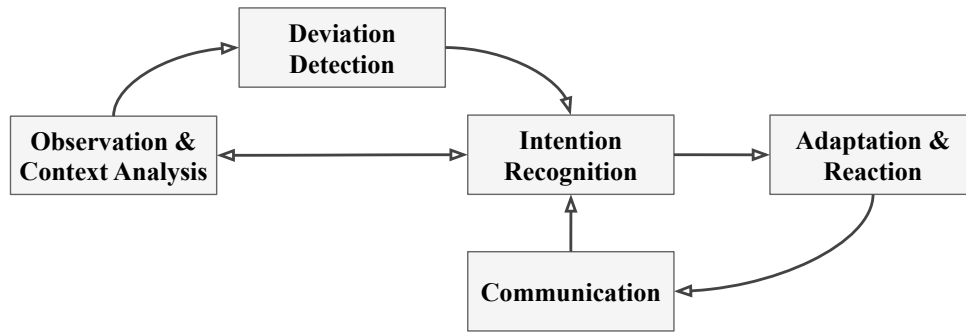


Figure 2: The components of the framework for intention recognition and communication.

5) Communication: In most cases, before deciding whether a reaction is proper to be executed, further adaptations are needed. To do so, the agents communicate the reasons behind the deviation, to ensure a shared understanding of the task and the goal. This component along with others can be seen as a cycle, as the new adaptation driven by the performed communication can lead to further iterations of intention recognition.

These components as the main building blocks of the proposed framework create a loop of intention recognition and communication, which ensures the shared understanding of the task and the goal, and also the safety and efficiency of the agents in a mixed human-robot environment. Figure 2 illustrates the components of the proposed framework and the loop of intention recognition and communication.

4. Illustrative Scenario: Large Sale Kitting Task

To demonstrate the proposed framework, we consider a large-scale kitting task in an industrial environment. The overall goal of the kitting task is to collect items from storage racks around a kitting area and place them on a central table. Figure 3 illustrates a simple representation of the kitting task scenario with a human and a robot in the environment.

In this scenario, we assume there are two agents – a human and a robot – working collaboratively to complete this task. Each agent has access to its designated half of the storage racks and the kitting table, and the predetermined plan they are following ensures that – to minimize collisions and disruptions – they should not enter each other’s designated zones.

The agents have a shared understanding of the tasks and the goal and they can observe the actions of each other within the environment. The robot is equipped with the proposed intention recognition framework: it can observe human actions, infer human intentions, detect deviations in various levels of abstraction, and react accordingly.

When a deviation from the shared tasks is detected, the robot adapts its reactions based on the inferred intentions and communicates with the human if necessary. The robot then can make a decision based on the higher abstract intention behind the deviation and react accordingly to the situation. We emphasise on the differences between the possible reactions: 1) intrinsic reaction where the robot directly reacts to the deviation by only considering the observed actions, or 2)

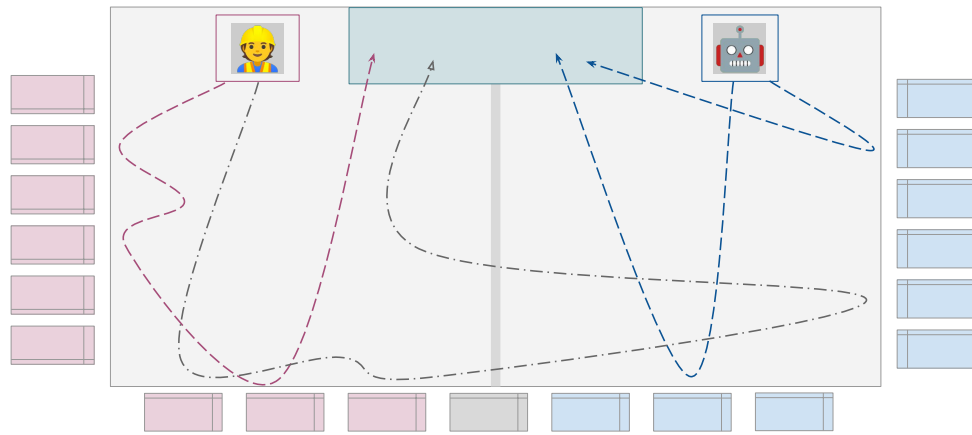


Figure 3: Illustration of the kitting task with a human and a robot in the environment. The goal is to retrieve items from the storage boxes around the edges of the kitting area and place them on the kitting table (top centre). Under the predetermined plan, the human should only collect items from the pink boxes on the left-hand side and the robot should collect items from the blue boxes on the right-hand side. However, sometimes deviations may occur from the plan, and the human may cross the central dividing line to enter the robot's work area.

enhanced reaction where the robot infers the intentions and communicates with the human.

We demonstrate two examples of deviations that may occur during the execution of the kitting task and discuss how the proposed framework can be applied to handle these situations with enhanced reactions based on inferred intentions and communication in comparison with the intrinsic reaction.

Scenario 1: The robot fails to pick up an item because the items are placed in the box in an unfortunate way. The intrinsic reaction of the robot is to give an alarm signal and then wait for the human to come and help pick up the difficult item. According to the intrinsic reaction, the robot must remain stationary while the human is within the robot's designated zone to avoid any possible collisions.

However, an enhanced reaction is for the robot, once it has informed the human of the picking failure, to deviate from the predetermined plan and continue to collect other items from the rack while the human retrieves the difficult item. The robot will re-plan to ensure that there is no collision with the human – that is, it must only collect items that do not interfere with the human's ability to retrieve the difficult item. The IR framework will facilitate the robot to recognise that the human has deviated from the predetermined plan to help the robot retrieve the difficult item, and allow the robot to communicate its own new intentions while the human is within the robot's designated zone. This enhanced reaction can potentially avoid the collision and improve the efficiency of the task.

Scenario 2: The robot has lost an item on its way to the kitting table, and the human crosses into the robot's designated zone to help retrieve the item. The intrinsic reaction of the robot is to stop to avoid collisions with humans. An enhanced reaction could be to recognise that

the human comes to help and communicate about what to do with the lost item and/or who brings it to the table. This enhanced reaction can potentially improve the efficiency of the task by resolving the ambiguities in the plan execution.

5. The Role of Communication for IR

In our proposed framework, the inferences driven by IR enable humans and robots to adapt and react safely and appropriately to deviations in a predetermined workflow. When deviations occur, humans and robots need to communicate and negotiate to enable reasonable decision-making, to react appropriately, and to allow the continuation of the workflow in a safe and efficient manner. The proposed framework introduces a hierarchy to inferring the intentions, by introducing semantically meaningful knowledge on a higher level, which enables more precise communication on any of the intermediate intention levels. It is a tool to reduce ambiguity, and consequently, uncertainty in the intention recognition framework.

Communication is an essential component of the intention recognition – adaptation/reaction – communication cycle. Through more purposeful understanding and recognition of each other's intentions, empowering more expressive communication strategies, the team becomes more of a peer-to-peer system, assisting each other to clarify the inferred intentions, aligning the team's mental model, thereby enhancing the process instead of stopping operations in case of deviations.

Acknowledgments

This work was supported by the Swedish Knowledge Foundation in the TeamRob Synergy Project (contract number 20210016).

References

- [1] A. Kolbeinsson, E. Lagerstedt, J. Lindblom, Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing, *Production & Manufacturing Research* 7 (2019) 448–471.
- [2] Merriam-Webster, Intention, n.d. URL: <https://www.merriam-webster.com/dictionary/intention>, accessed 16 Apr. 2024.
- [3] F. Sadri, Logic-based approaches to intention recognition, in: *Handbook of research on ambient intelligence and smart environments: Trends and perspectives*, IGI Global, 2011, pp. 346–375.
- [4] G. B. Smith, V. Belle, R. P. Petrick, Intention recognition with problog, *Frontiers in Artificial Intelligence* 5 (2022) 806262.
- [5] J. Porteous, A. Lindsay, F. Charles, Communicating agent intentions for human-agent decision making under uncertainty, in: *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS '23*, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2023, p. 290–298.

- [6] T. Tong, R. Setchi, Y. Hicks, Context change and triggers for human intention recognition, *Procedia Computer Science* 207 (2022) 3826–3835.
- [7] M. L. Chang, R. A. Gutierrez, P. Khante, E. S. Short, A. L. Thomaz, Effects of integrated intent recognition and communication on human-robot collaboration, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 3381–3386.
- [8] J. Zhang, H. Liu, Q. Chang, L. Wang, R. X. Gao, Recurrent neural network for motion trajectory prediction in human-robot collaborative assembly, *CIRP annals* 69 (2020) 9–12.
- [9] S. Buyukgoz, J. Grosinger, M. Chetouani, A. Saffiotti, Two ways to make your robot proactive: Reasoning about human intentions or reasoning about possible futures, *Frontiers in Robotics and AI* 9 (2022) 929267.
- [10] R. Salehzadeh, J. Gong, N. Jalili, Purposeful communication in human–robot collaboration: A review of modern approaches in manufacturing, *IEEE Access* 10 (2022) 129344–129361.