Towards Intention Recognition for Human-Interacting Agricultural Robots

Alexander Gabriel
Lincoln Centre for Autonomous Systems
University of Lincoln
Lincoln, United Kingdom
agabriel@lincoln.ac.uk

Paul Baxter
Lincoln Centre for Autonomous Systems
University of Lincoln
Lincoln, United Kingdom
pbaxter@lincoln.ac.uk

***Abstract*—Robots sharing a common working space with humans and interacting with them to accomplish some task should not only optimise task efficiency, but also consider the safety and comfort of their human collaborators. This requires the recognition of human intentions in order for the robot to anticipate behaviour and act accordingly. In this paper we propose a robot behavioural controller that incorporates both human behaviour and environment information as the basis of reasoning over the appropriate responses. Applied to Human-Robot Interaction in an agricultural context, we demonstrate in a series of simulations how this proposed method leads to the production of appropriate robot behaviour in a range of interaction scenarios. This work lays the foundation for the wider consideration of contextual intention recognition for the generation of interactive robot behaviour.**

***Index Terms*—Agricultural Robotics, Human-Robot Interaction, Intention Recognition, Belief-Desire-Intention System**

I. Introduction

Introducing robots into a human working space can increase efficiency but should not come at the cost of comfort or safety. To achieve this balance in a challenging setting like agriculture, a robot needs to understand the intentions behind their coworkers' behaviour and basic communication. Gestures form an ideal medium to maintain reliability in adverse circumstances but are limited to situations where the human has their hands free. Additional clues from the environment as well as behaviour analysis can be used to estimate their state. Our interpretation of intentions [5] sees them as the meaning [4] of, explanation [7] for, or idea [9] behind an action, plan or utterance. In our agricultural setting, workers pick berries into crates in a poly-tunnel environment. The robot is acting in a supporting role, supplying the human with empty crates, taking away full crates and staying out of the way the rest of the time. To facilitate the robot's autonomy, we created in integrated sensor data processing pipeline and Belief-Desire-Intention (BDI) [1] agent system. The general motivation for this system is that in order to 'understand' the intentions of their human interaction partner (from observable behaviour) and to generate appropriate responses, the robot should consider both the environmental context but also its own goals (or 'desires'): this supports our use of a BDI architecture. This work summarizes a first evaluation of the system's performance in a range of scenarios. We evaluated the system in simulation, using sensor readings as well as statistical data gathered in real-world poly-tunnels.

Supported by the RASberry project (https://rasberryproject.com), and the CTP for Fruit Crop Research (www.ctp-fcr.org).

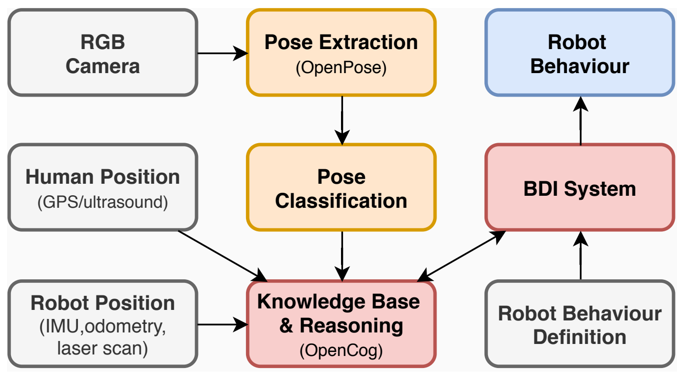

Fig. 1. HRI System Representation **Grey:** Datasources, **Orange:** Pre-Processing **Red:** Memory, Reasoning, and Agent **Blue:** Output

II. Human-Robot Interaction (HRI) System

A. Data Processing: The robot perceives its environment through a stereo RGB-D camera, a thermal camera, 2D and 3D LIDAR, as well as differential GPS and odometry. It can also receive its coworkers' location either provided by GPS or ultrasonic localization and is supplied with predefined topological and laser maps. During this simulation, the robot determines its position using simulated laser scans and odometry. The location of the human is supplied by a picker-simulation engine and abstracted using Qualitative Trajectory Calculus (QTC) [8].

As shown in Fig. 1, the video is first pre-processed using OpenPose [2] to extract joint positions. Those are further processed to extract joint angles before both are fed into a naive classifier that produces pose labels for each frame individually, based on predefined prototype poses. Series of frames are classified using voting rounds where each frame contributes a single vote towards a pose. Whichever pose first wins 10 votes, labels the round. The movement samples used in this evaluation are part of a new dataset for Action Recognition in agri-robotics. It contains samples of behaviours, such as picking of berries and carrying of crates, and samples of gestures to communicate with the robot. The samples were recorded from 10 different subjects between morning and early afternoon in a poly-tunnel environment featuring ripe strawberry plants.

B. Belief-Desire-Intention Agent: The BDI system chooses intentions (plans to reach a goal) from its desires (abstract goals) based on its beliefs as captured in the Knowledge Base (KB)[1]. This separates reasoning about which goals to achieve from managing the execution of said goals. This allows us to consider more contextual information when deciding which goal to follow and leads to an aesthetic analog to our idea of human motivation, intention and action on the robot.

In our system, plans are represented as ordered tree structures with executable actions as leafs. Actions have a set of preconditions and expected consequences, which combine to form the preconditions and expected consequences of a plan. When the agent decides which of its desires are applicable in a given situation, it searches the KB for patterns of beliefs that match its desire's preconditions. If successful, it produces a corresponding intention. When idle, the robot chooses from the possible next actions defined by its current intentions, based on utility and expected time requirements.

III. Evaluation

In the context of interactions between a robot and a human in an agricultural context, we explore three scenarios (Fig. 2). They cover different situations in the work environment: starting to work (crate delivery), moving around (evading the human), and resupply (crate exchange). Both the delivery and exchange scenarios are initiated by the human gesturing to the robot, but require a different response. The robot can make the distinction based on prior observed human behavior (whether the person has been picking berries). The evasion scenario is triggered by human behaviour (approaching) without any conscious interaction.

[1]We are using OpenCog's [6] AtomSpace and Pattern matcher for knowledge representation and reasoning.

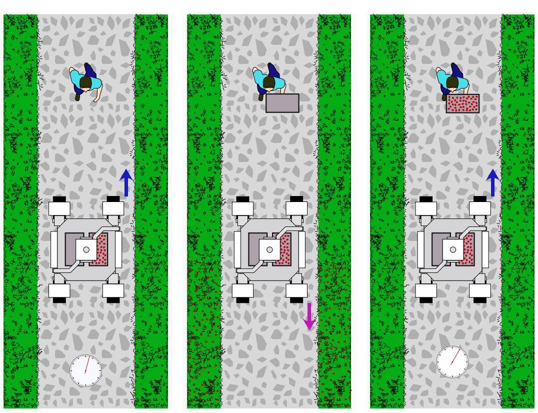

Fig. 2. **Left:** Delivery Scenario: Meet the human, wait for 2.5 seconds, leave. **Middle:** Evasion Scenario: On human approach, move to the next waypoint. **Right:** Exchange Scenario: Meet the human, wait for 5 seconds, leave.

TABLE I
EXPERIMENT RESULTS

Scenario	Success Rate	Meeting Distance [m]		Time to Service [s]	
		μ	σ	μ	σ
Delivery	0.99	0.37	0.007	12.65	22.552
Evasion	1.00	N/A		10.06	1.706
Exchange	0.99	0.37	0.006	11.83	0.634

The recorded human behaviours in poly-tunnels (Section II), are passed as input to the proposed system (Fig. 1), acting on a simulated environment. Given 10 recorded subjects, 20 simulations per subject are performed (given a stochastic simulation), resulting in 200 simulations per scenario.

Table I shows for each scenario the mean (μ) and standard deviation (σ) of the three metrics of evaluation: success rate, meeting distance and time to service. Success rate is defined as the share of experiment runs that ended with the robot successfully interpreting the situation and performing the expected actions. Meeting distance is the distance between human and robot at which the robot decided to halt to facilitate the delivery or exchange of a crate. The variance for this metric can be interpreted as an indicator of how much reasoning affects the agent's reaction time (the robot and human don't meet in the Evasion case). Time to Service is the time between the human displaying behaviour that should trigger a change in robot behaviour, and the time at which the robot performed the expected action (delivered or exchanged a crate, or moved away from the human to the next waypoint). This time consists mainly of the time it takes to detect the behavior, the time it takes to meet the human (~$4s$), and the time for the delivery ($2.5s$) or exchange ($5s$) of crates. The large variance for the Delivery scenario stems from the robot's failure to detect one subject's behavior correctly. Without these cases, the variance for the Delivery scenario is 0.298.

IV. Conclusion

Our evaluation shows benefits compared to a system that is unaware of its human coworkers except for their location and service calls. However, they do not directly support any claim of advantage over a simpler, reaction-based, human-aware system. The hypothesis is that by using intention recognition to anticipate the human's requirements, there will be observable benefits in both task efficiency and perceived comfort of the interaction, but this requires validation.

The service times of under $15s$ are short in comparison to the $110 \pm 44s$ [3] it takes for the robot to service the picker when stationed outside the field. This points at a potential increase in productivity, achievable by estimating the next time a worker requires the robot's service and arriving in the vicinity early (another source of information to facilitate anticipatory robot behaviour).

Overall, while there remain a number of outstanding challenges, this paper has outlined the fundamentals of our approach: the consideration of both observable human behaviour and the wider environmental context in supporting anticipatory interactive robot behaviour.

REFERENCES

[1] M. Bratman et al. *Intention, plans, and practical reason*, volume 10. Harvard University Press Cambridge, MA, 1987.

[2] Z. Cao et al. Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition*, 2017-Januar:1302–1310, 2017.

[3] G. Das, G. Cielniak, et al. Discrete event simulations for scalability analysis of robotic in-field logistics in agriculture – a case study. In *IEEE International Conference on Robotics and Automation, Workshop on Robotic Vision and Action in Agriculture (WRVAA)*, 2018.

[4] M. Fleischman and D. Roy. Why Verbs are Harder to Learn than Nouns: Initial Insights from a Computational Model of Intention Recognition in Situated Word Learning. In *Proc. of the Annual Meeting of the Cognitive Science Society*, 2005.

[5] A. Gabriel, S. Coşar, N. Bellotto, and P. Baxter. A dataset for action recognition in the wild. In *Annual Conference Towards Autonomous Robotic Systems*, pages 362–374. Springer, 2019.

[6] B. Goertzel, C. Pennachin, and N. Geisweiller. The opencog framework. In *Engineering General Intelligence, Part 2*, pages 3–29. Springer, 2014.

[7] K. A. Tahboub. Intelligent human-machine interaction based on Dynamic Bayesian Networks probabilistic intention recognition. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 45(1):31–52, 2006.

[8] N. Van de Weghe et al. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and cybernetics*, 35(1):97–119, 2006.

[9] S. Youn and K. Oh. Intention recognition using a graph representation. *International Journal of Applied Science, Engineering and Technology*, 4(1):13–18, 2007.