Team ORIon — 2023 Team Description Paper

Samuel Sze, Ricardo Cannizzaro, Matthew Munks, Kim Tien Ly, Shu Ishida, Ana Deligny, Arundathi Shanthini, Tiffany Horter, Lars Kunze, Ioannis Havoutis, Nick Hawes, and Bruno Lacerda

Oxford Robotics Institute, Dept. of Engineering Science, University of Oxford, UK
orion@oxfordrobotics.institute
https://ori.ox.ac.uk/student-teams/team-orion

Abstract. This document outlines the approach *Team ORIon* will take to the RoboCup@Home 2023 (DSPL). Our first experience competing with Toyota's Human Support Robot (HSR) was at the World Robot Summit (WRS) 2018. We have since competed at RoboCup 2019 and 2022 in the DSPL @Home league. We aim to return to the competition in 2023 with a refreshed team and revamped HSR. Our research interests are centred around long-term autonomy; mobility and whole body motion planning; causal reasoning and knowledge representation; and trustworthy and explainable AI. Advances in these directions will enable service robots to interact with humans and complete useful everyday tasks in typical household settings. We aim to demonstrate robust and intelligent autonomous behaviour, that uses experience to learn and refine a growing set of robot skills, on the HSR.

1 Introduction

Team ORIon is a student robotics competition team created in 2018 within the Oxford Robotics Institute (ORI) at the University of Oxford. The team consists of undergraduate and graduate students, and is supported by robotics postdoc researchers, principal investigators, and other faculty members of the ORI. We come from a strong research institute with seminal work in mobile autonomy and machine learning. The ORI has a significant track record in field robotics and real-world trials of autonomous systems. Further, the ORI also has a team of professional hardware and software engineers. This experience and support is leveraged to deliver a strong performance in the RoboCup competition.

1.1 Past Competition Involvements

The RoboCup@Home DSPL league provides a challenging environment and complex tasks to which new and existing ORI research can be applied. Since acquiring our Toyota Human Support Robot (HSR) in 2018, we have competed in the Partner Robot Challenge of WRS; we also competed in the 2019 RoboCup@Home competition, placing 6^{th} out of 10 on our first attempt at the competition . Most recently, we competed in RoboCup@Home 2022 at Bangkok, Thailand (see Figure 1). Although the results from Bangkok was not what we hoped for, competing

Fig. 1. *Left:* Team ORIon's HSR grasping a plant at the World Robot Summit (WRS) 2018 in Tokyo, Japan. *Middle:* Team ORIon at RoboCup@Home 2019 in Sydney, Australia. *Right:* Team ORIon at RoboCup@Home 2022 in Bangkok, Thailand.

Fig. 2. *Left:* Team ORIon demonstrating the HSR to the public at the 2021 Goodwood Festival of Speed. *Middle:* Demonstrating the HSR for HRH Prince William. *Right:* Demonstrating HSR at Dept. of Engineering Science Open day 2022.

in it allowed us to see the shortcomings in our capabilities. We hope to leverage these valuable experience from past competitions and continuously improve our HSR for future participations.

1.2 Public Outreach

We also run robot demonstrations to the public at outreach events (e.g., Engineering Science Open day and the Goodwood Festival of Speed), to promote greater awareness and interest in STEM, robotics, and AI to the wider public. In October 2019, we demonstrated the HSR for His Royal Highness Prince William at the ORI to celebrate the opening of a new college building (see Figure 2).

2 Team Composition

The team is led by ORI Master student Samuel Sze, and is composed into six subteams to manage different robot sub-systems: task-level planning, perception, manipulation, human-robot interaction, semantic mapping and navigation. The core team members for 2022 are Oxford University PhD and undergraduate student sub-team leaders Arundathi Shanthini, Matthew Munks, Kim Tien Ly, Ana Deligny and Tiffany Horter. They are supported by additional PhD and undergraduate team members. The sub-team responsibilities and team members are described on the Team ORIon website[1].

The team is supported by ORI principal investigators Dr. Lars Kunze, who has a strong research focus on scene understanding and semantic mapping, causal reasoning, and explainable and trustworthy AI; Prof. Nick Hawes, who has extensive background in intelligent autonomous robots that can work with or for humans in uncertain environments; and Dr. Ioannis Havoutis, an expert in combining motion planning with machine learning.

3 Capabilities and Goals

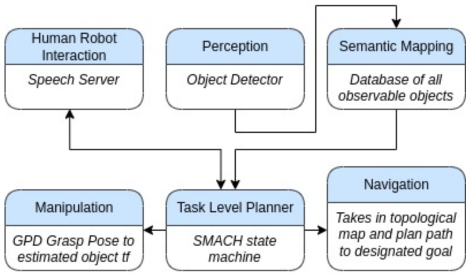

Fig. 3. Overall Team ORIon HSR system architecture

We built our robot's capabilities by leveraging open-source software to create new sub-system functionality and combine them to produce new autonomous behaviours to perform service tasks in typical domestic environments. In the following sub-sections, each of the sub-system shown in Figure 3 will be explained in detail.

3.1 Manipulation

The manipulation stack incorporates grasp synthesis directly on target object point clouds, including efficient point cloud filtering and segmentation. We use Grasp Pose Detection(GPD) [1,2] for grasp pose synthesis, giving us the 6-dimensional pose of approach for the highest quality grasp generated. An overview of the grasp generation and execution process can be seen in Figure 4.

Manipulation actions to be carried out are specified by the task-level planner, which provides the target transformation. At the same time, we use point clouds from the Xtion RGB-D camera for online collision mapping (with *OctoMap* occupancy estimation), detection of horizontal (e.g., table top) and vertical (e.g., door) surfaces, and detection of handles for drawers and doors. Given the object's 3D pose provided by the object detection system, the manipulation system

[1] `https://ori.ox.ac.uk/student-teams/team-orion/meet-the-team`

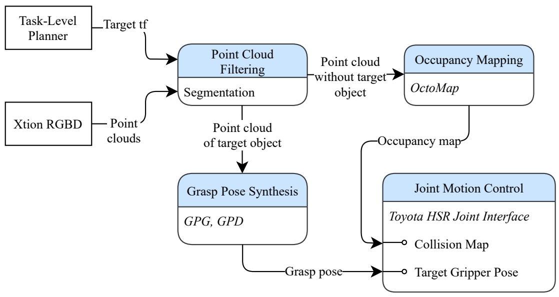

Fig. 4. Manipulation grasp synthesis and execution.

segments the object's point cloud out of the point cloud input and feeds it into the grasp synthesis module. This module then returns a ranked list of feasible candidate poses for the end effector to approach from. Accordingly, the system moves to the highest-ranked end effector pose to attempt to grasp the object while avoiding collisions. Motion control for the HSR arm and gripper is done via the Toyota joint motion control interface.

An observed flaw is that the grasp synthesis pipeline is less reliable for small and flat objects such as cutlery. This is due to GPD lacking knowledge of the gripper's shape and the difficulty of extracting a representative point cloud of these objects. Thus, we implement a custom visual feedback system, using the in-hand camera to align and orient the end effector from above to ensure a successful grasp. We also perform object-dependent selection between the gripper or the vacuum pad to allow the manipulation system to make full use of the HSR's manipulation hardware.

Team ORIon guidance member Lars Kunze's previous work includes knowledge-enabled manipulation [3], demonstrated in a system which could grasp an egg[2] and make a pancake[3]. Further, Team ORIon member Ioannis Havoutis has an extensive background in learning, synthesis and control of complex motions.

3.2 Task-Level Planning

The task-level planning sub-system acts as the executive controller of the robot system. Autonomous behaviours are defined through the use of hierarchical SMACH state machines[4], which combine abilities provided by other sub-systems, accessible via ROS Action servers, to perform useful tasks in a domestic environment (e.g., fetching a bottle of water for a human).Task level planning takes in information from semantic map, navigation, speech,18000 / 80 manipulation, and various intrinsic utility packages and generates the appropriate states for each sub-system. Subsequently, these states help guide the respective sub-system to understand which stage of the task it is on, and what is required to accomplish them.

[2] https://www.youtube.com/watch?v=jLz87H4q3hU
[3] https://www.youtube.com/watch?v=YQs5gRei8k4
[4] http://wiki.ros.org/smach

3.3 Navigation

In terms of global path planning, with reference to figure 5, we use a probabilistic navigation system inspired and modified from the STRANDS [4,5] project structured around a topological map (i.e., discrete locations connected by directed edges). Edges corresponds to a probabilistic navigation action the robot could perform to transition between locations, where a higher probability represents higher success rate. At the same time, navigation takes in movement goals from the task-level planning subsystem. After obtaining the goal destination, navigation package performs a topological search through the map to plot an optimal draft path along the edges. The robot then follows the draft path via the ROS Action server provided by the *tmc_move_base* ROS package.

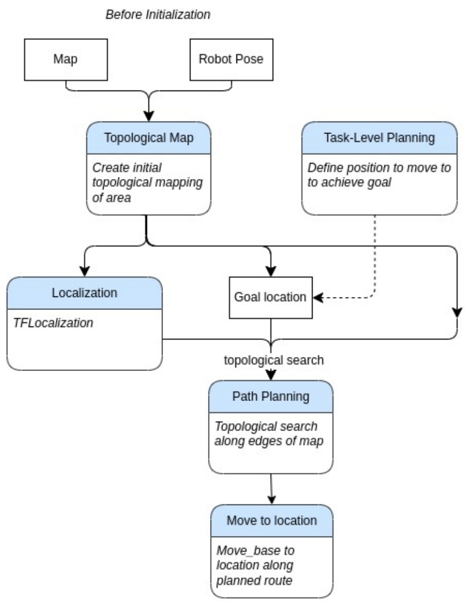

Fig. 5. Global Path Planning

Based on our past experiences, we also seek to define a better local path planning methodology for the HSR to adapt to dynamic or previously unseen obstacles within the topological map. Additionally, it should also correct drifts in the robot's movement due to errors in localization or low level control.

3.4 Perception

Perception allows the robot to use the visual and depth input from its RGB-D cameras to determine the identity and location of objects in the surroundings. Currently, Team ORIon have implemented a Faster R-CNN model with a ResNet-50-FPN backbone [6], which have been pre-trained on the COCO datasets. With reference to figure 6, the current image from the camera being inputted into the model, which in turn outputs a bounding box for an object, the object's identity, and the confidence level. We then process the output by applying non-maximum suppression to the set of outputs to remove duplicate detections, then apply a minimum confidence threshold. Image coordinates of detected objects are transformed into 3D, using ROS *tf*.

Perception is the primary information source in our system. Therefore, future work in this sub-system is centred around providing fast and reliable results. Firstly, we seek to replace the current Fast R-CNN model with Yolov7 [7] in order to improve inference speed by utilizing a smaller and more lightweight network architecture. With more memory to work with by using Yolov7, the second area of focus will be to create a separate facial recognition model that can identify unique traits of people faces, thus allowing for completion of Robocup tasks such as searching for specific people. Lastly, we also seek to implement multi-object tracking using StrongSORT [8] to aid the HSR's ability to classify and track objects better temporally.

3.5 Semantic Mapping

Learning and recognising objects during operation, and reasoning about them, are key tasks for a mobile service robot in human environments. Our semantic mapping sub-system obtains information from the perception sub-system and maintains a semantic object map to track the positions of objects based on their class (e.g., book, water bottle). It is implemented on top of a MongoDB database. One functionality of the semantic object map is to evaluate queries on relative position relationships between objects (e.g., is object A *above* object B?, are there any objects *in the living room*?). The semantic map can also compute how semantically similar two object classes are (e.g., how semantically similar is an apple to a banana?) using the OWL ontology format.

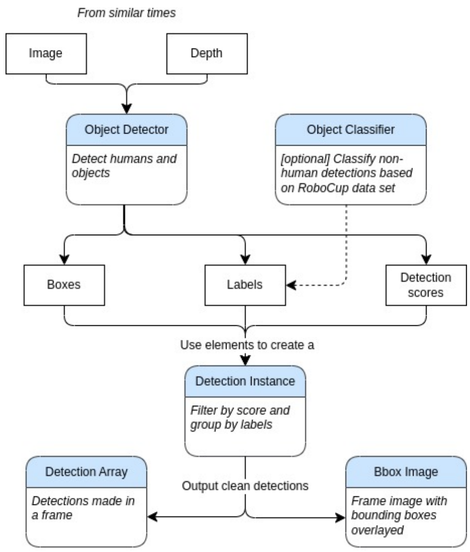

Fig. 6. Perception pipeline

The system has been designed to be very extendable, with ROS message definitions to store/retrieve being stored at runtime. This makes for a very flexible system, one that can be added to very easily. We also have the capability to check for the persistence of an object. If an observation comes in of the same class and at the same rough location as an object already stored, then that object is updated.

3.6 Human Robot Interaction

To allow humans to interact with the robot using natural language, we have developed a sub-system that performs classification of speech commands. This system interfaces with the cloud-based Google Speech-to-Text[5] as well as local

[5] https://cloud.google.com/speech-to-text/

methods such as PocketSphinx[6] and a pre-trained WaveNet [9] when an internet connection is not available. For classifying speech commands, we measure the Levenstein distance with candidate texts, with added semantic robustness by looking up synonymous words in WordNet [10]. This allows an operator to communicate with the robot in predefined dialogues, such as issue commands ("tidy up", "bring me *something*", "move to *location*") and/or respond to requests made by the robot.

4 ORI Original Research

Semantic Affordance Segmentation - Team ORIon member Matthew Munks is currently developing a system for working out how doors and drawers move relative to the world frame just from visual input. This information will then be stored in the semantic mapping system, thus enhancing what the robot has to work from.

Ethical Black Box - Team ORIon member Matthew Munks has developed an Ethical Black Box (EBB) for the HSR with the goal of increasing explainability and trustworthiness of the autonomous robot system. As proposed in [11], an EBB is a system, analogous to an aircraft's flight data recorder, that logs the internal states and perceptions of a robot, as well as any errors or warnings that occur while carrying out its actions. The EBB permits end users and developers to query the behaviour and decisions of the autonomous system and reconstruct a scenario of the robot via simulation. This is particularly useful to generate explanations for unexpected or questionable robot behaviour. A human-robot dialogue system has been developed to support this.

Robust Causal Inference - As part of his PhD research, past ORIon team lead Ricardo Cannizzaro is developing probabilistic causal reasoning methods to permit the HSR to reason about the complex relationships that exist between its environment and itself, including its tasks, capabilities, and components. This allows the HSR to predict likely outcomes of its actions before committing to them, as well as inferring the likely causes for observed outcomes and the likely states of unobserved variables given observations of visible ones.

Motion Planning, Scene Reconstruction & Gaze Control - As part of his PhD research, former Team ORIon member Mark Finean has developed several whole-body motion planning methods for robots operating in dynamic environments [12,13]. His work includes a hybrid mapping and motion planning system to permit simultaneous scene reconstruction and collision avoidance, and a 'smart' gaze controller that achieves effective perception of the environment for obstacle avoidance and motion planning in dynamic and unknown environments. These results were validated on Team ORIon's HSR. Thus, Team ORIon can leverage these research efforts.

[6] `https://github.com/cmusphinx/pocketsphinx`

5 Conclusion

Team ORIon aims to return to the DSPL again in 2023. Competing in the RoboCup@Home event in France will allow us to demonstrate the new and presented capabilities on the HSR in a challenging environment and will provide valuable real-world robotics experience for our team.

References

1. Andreas ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. Grasp Pose Detection in Point Clouds. *The International Journal of Robotics Research*, 36(13-14):1455–1473, dec 2017.
2. Marcus Gualtieri, Andreas ten Pas, Kate Saenko, and Robert Platt. High precision grasp pose detection in dense clutter. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 598–605. IEEE, oct 2016.
3. Lars Kunze and Michael Beetz. Envisioning the qualitative effects of robot manipulation actions using simulation-based projections. *Artificial Intelligence, Special Issue on AI and Robotic*, 2015.
4. Nick Hawes, Chris Burbridge, Ferdian Jovan, and Lars Kunze. et.al. The STRANDS project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine (RAM)*, 24(3), September 2017.
5. Bruno Lacerda, Fatma Faruq, David Parker, and Nick Hawes. Probabilistic planning with formal performance guarantees for mobile service robots. *International Journal of Robotics Research*, 38(9), 2019.
6. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.
7. Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.
8. Yunhao Du, Yang Song, Bo Yang, and Yanyun Zhao. Strongsort: Make deepsort great again, 2022.
9. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
10. George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
11. Alan F. T. Winfield and Marina Jirotka. The case for an ethical black box. In Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou, editors, *Towards Autonomous Robotic Systems*, pages 262–273, Cham, 2017. Springer International Publishing.
12. Mark Nicholas Finean, Wolfgang Merkt, and Ioannis Havoutis. Simultaneous scene reconstruction and whole-body motion planning for safe operation in dynamic environments. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Czech Republic, 2021.
13. Mark Nicholas Finean, Wolfgang Merkt, and Ioannis Havoutis. Where should i look? optimised gaze control for whole-body collision avoidance in dynamic environments. *IEEE Robotics and Automation Letters*, 2021. Under review (arXiv preprint arXiv:2109.04721).

HSR Software and External Devices

We use a standard DSPL HSR robot from *Toyota*. No modifications have been applied.

Fig. 7. Team ORIon's HSR ("*Bamm-Bamm*")

Robot's Software Description

For our robot we are using the following software:

- Platform: Ubuntu 20.04 / ROS Noetic
- Task-level planning: SMACH ROS library
- Navigation: *tmc_move_base* ORI Topological map and ROS package[7]
- Semantic mapping: *semantic_mapping**, MongoDB
- Object recognition and pose estimation: Faster R-CNN model with ResNet-50-FPN backbone (Yolov7 and StrongSORT planned for early 2023)
- 3D occupancy mapping: OctoMap[8]
- Arm control and gripper coordination: Toyota HSR Motion Interface (ROS MoveIT planned for early 2023)
- Grasp pose synthesis: Grasp Pose Detection (GPD)[9], Grasp Pose Generator (GPG)[10]
- Speech recognition: PocketSphinx and Wavenet (offline speech recognition), Levenshtein[11] (semantic similarity checking), WordNet (synonym checking)

Items marked with an asterisk are Team ORIon custom solutions.

External Devices

The HSR robot relies on the following external hardware:

- Alienware laptop in robot mount (for object recognition and pose estimation)

Cloud Services

The HSR connects to the following cloud services:

- Speech recognition: All-purpose recogniser (Google API).

[7] http://github.com/ori-orion/orion_topological
[8] https://octomap.github.io
[9] https://github.com/atenpas/gpd
[10] https://github.com/atenpas/gpg
[11] https://github.com/maxbachmann/Levenshtein