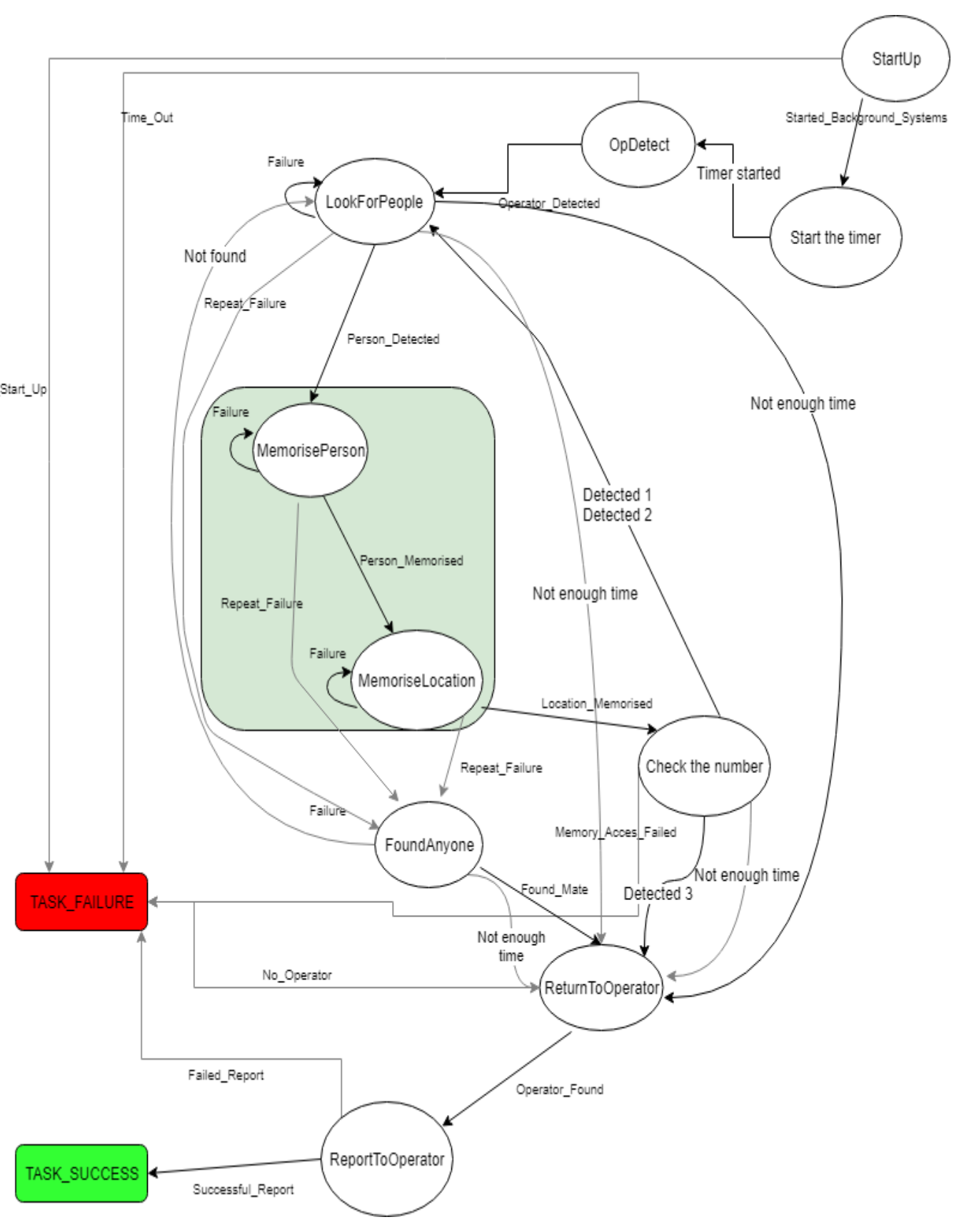
ORI
OXFORD ROBOTICS INSTITUTE

Team ORIon

UNIVERSITY OF OXFORD

TASK LEVEL PLANNING

Low level behaviours are implemented using ROS action servers – SMACH task level architecture. By combining these behaviours into a state machine, we create a robust, and flexible Execution framework. It can be optimized by reusing low-level behaviours. Flexibility is achieved by reusing building blocks.



Block diagram for the task: Find my mate

VISION

The object detector outputs the 3d location of objects. We designed an automatic pipeline for retraining on custom objects. It consists of a photo booth fitted with green felt and a turntable.

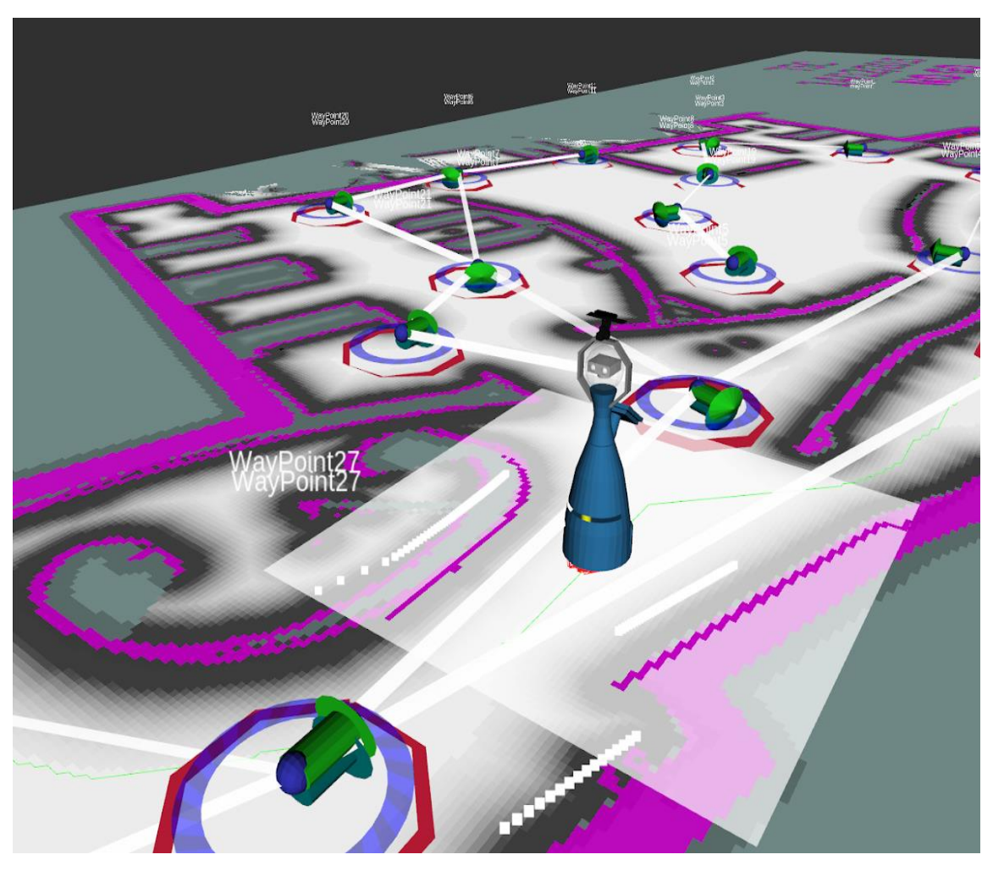
The face detector is pretrained on human faces and it returns the age, gender and emotion.

The human pose detector gives the location of different parts of the body (Cao et al. 2018).

Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." arXiv preprint arXiv:1812.08008 (2018).

NAVIGATION

The robot performs high-level planning over a topological graph. Edges have associated actions to navigate between nodes by opening doors and performing trajectory planning. We model navigation over the graph as a Markov Decision process based on navigation data (Hawes et al. 2017). We use this model to generate optimal policies with probabilistic guarantees over the robot performance metrics (Lacerda et al. 2019).

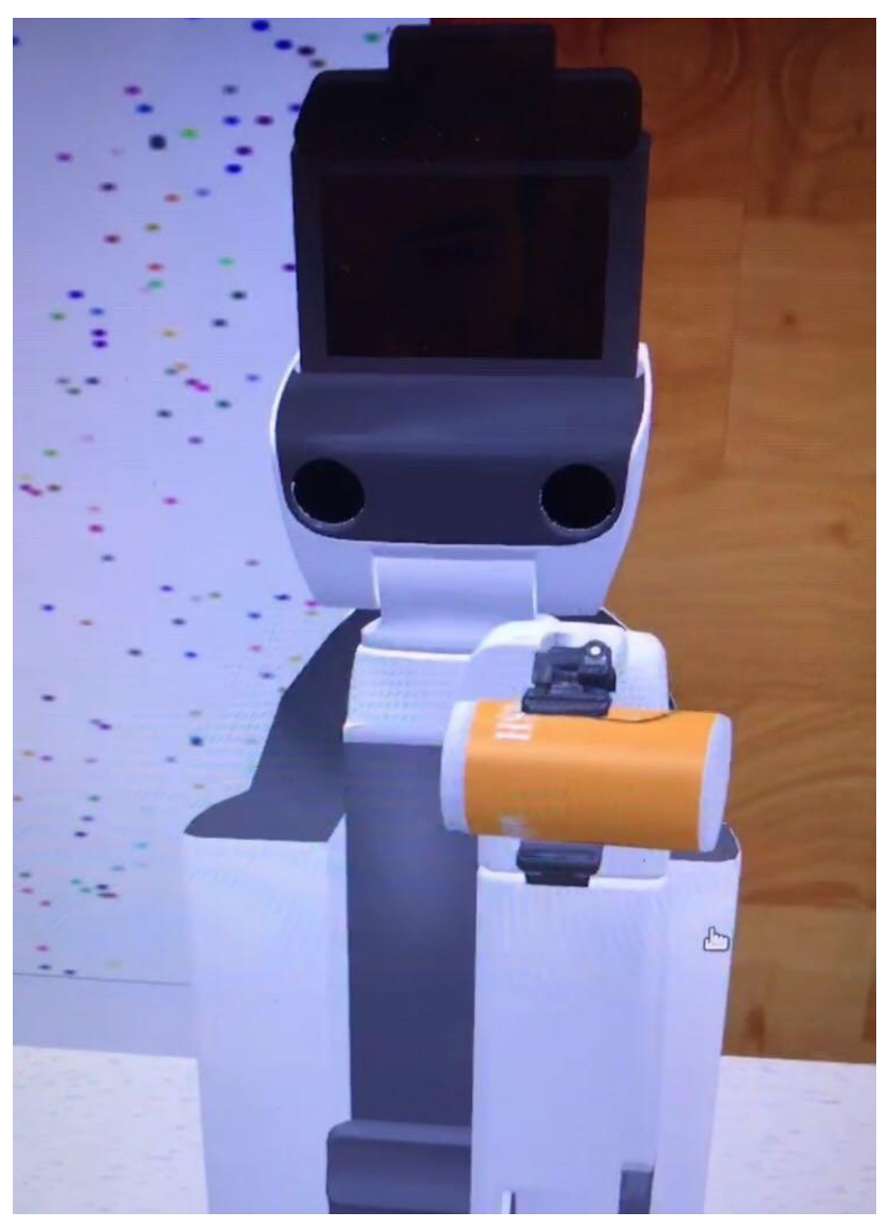


To perform low level trajectory planning we use search algorithms between each of the node regions. We have performed research on improving the efficiency of planning through a series of regions (Ishida et al. 2019).

1. Ishida, Shu, et al. "Robot path planning for multiple target regions'' to appear in IEEE European Conference on Mobile Robots (2019).
2. Hawes, Nick, et al. "The strands project: Long-term autonomy in everyday environments." IEEE Robotics & Automation Magazine 24.3 (2017): 146-156.
3. Lacerda, Bruno, et al. "Probabilistic planning with formal performance guarantees for mobile service robots." The International Journal of Robotics Research (2019): 0278364919856695.

MANIPULATION

Manipulation encompasses all tasks in which the robot needs to interact with an object, ranging from door opening to pick up desired objects. We utilise the DBoW2 implementation of Visual SLAM [1] to create a 3D collision map for obstacle avoidance in the motion planning. In order to robustly grasp objects and handles, we use PCL surface segmentation to extract a point cloud of the object of interest and perform grasp synthesis as described by Pas et al. in [2] which uses a trained CNN (convolutional neural network) to score sampled grasps. We find that this approach is effective for larger objects but fails for small, thin objects such as cutlery. For objects such as these, we use our own implementation of approaching the object from above and using visual feedback from the hand camera in order to optimally grasp the object along its minor axis.

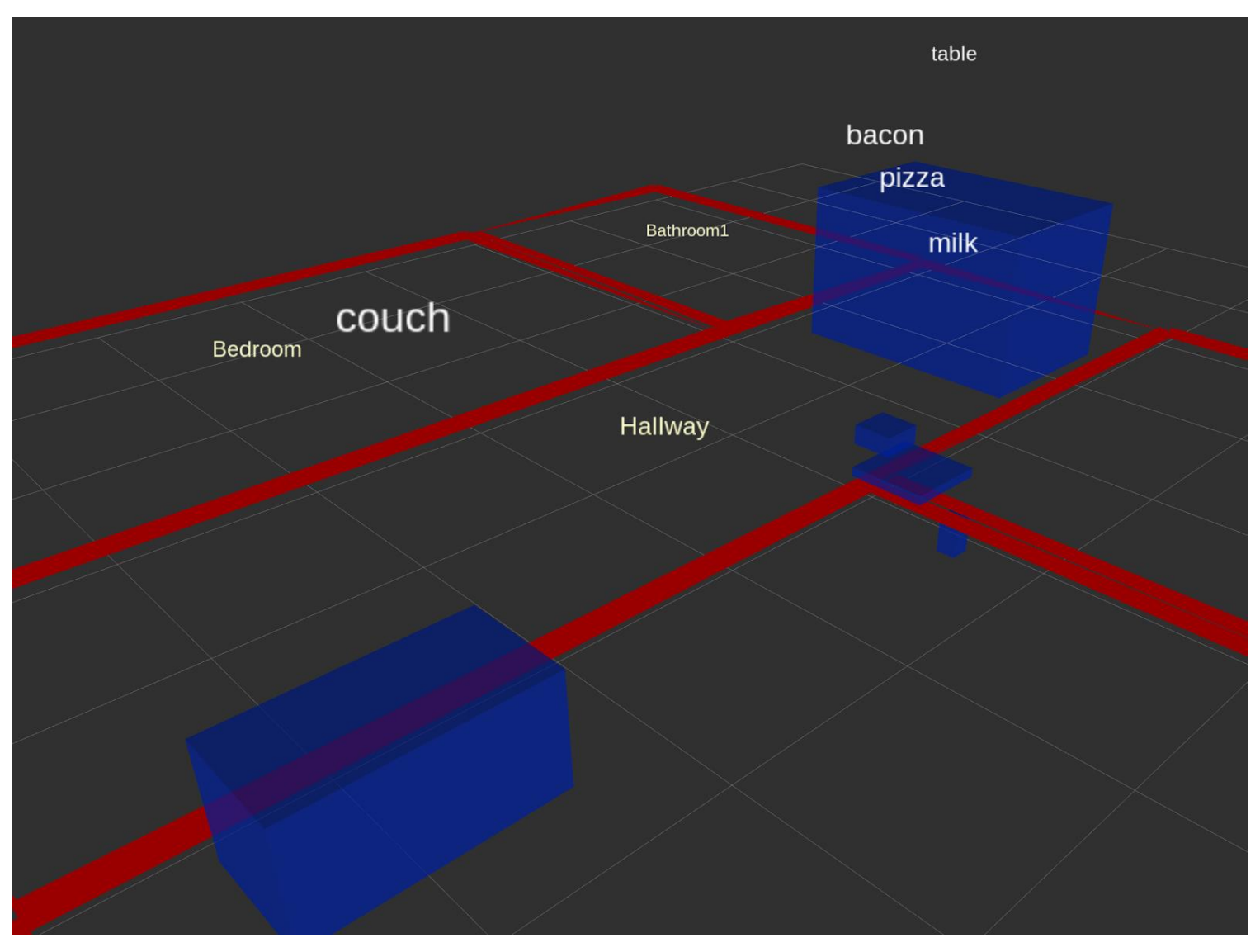


[1] D. Galvez-Lopez, J. D. Tardos, Bags of Binary Words for Fast Place Recognition in Image Sequences, IEEE Transactions on Robotics, 2012

[2] A. Pas, M. Gualtieri, K. Saenko, R. Platt, Grasp Pose Detection in Point Clouds, The International Journal of Robotics Research, Vol 36, Issue 13-14, pp. 1455 - 1473, October 2017

SEMANTIC MAPPING

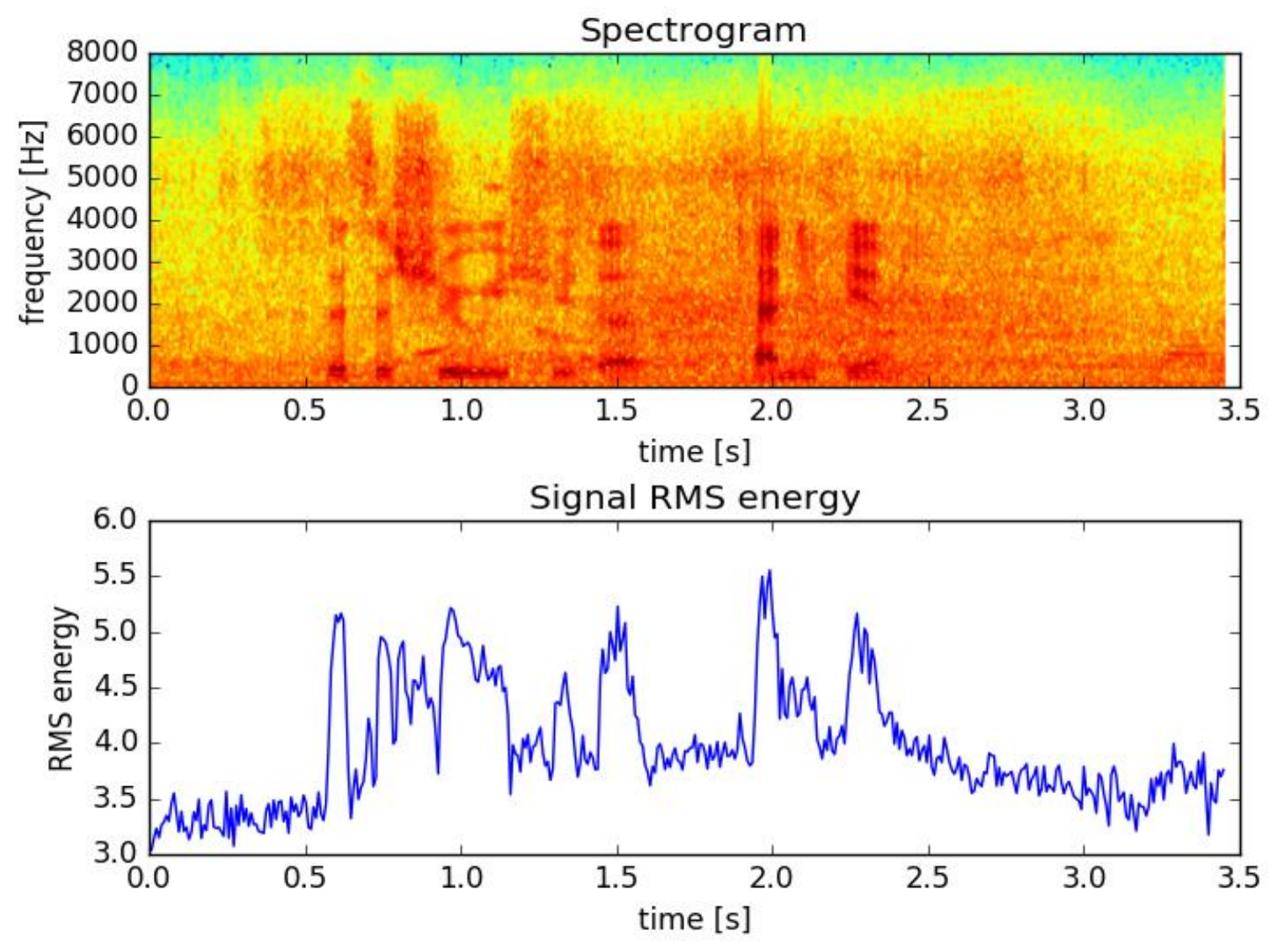
We extended the SOMA framework (Kunze et al. 2018) to utilise prior knowledge about objects, and perform more flexible queries. As the robot detects objects, they populate the semantic map. The semantic map can be queried for objects of a given type (eg. "where are fruit objects?"), or by the spatial relations between objects (eg. "what is the left-most object above the table?"). The environment is segmented into rooms using a GUI, and prior knowledge about the expected locations of objects is integrated with observations to estimate object locations. An ontology of object types is used to understand the relationship between object types, and enable queries of varying levels of specificity. For example, an "apple" object would be returned in a query for "apples", "fruit", or "food".



Kunze, Lars, et al. "SOMA: A Framework for Understanding Change in Everyday Environments Using Semantic Object Maps." AAAI, 2018.

SPEECH RECOGNITION

Speech recording and recognition are performed simultaneously and asynchronously in batches, determined by thresholding the moving average of RMS signal energy. Besides Google Speech to Text, we also have PocketSphinx (Huggins-Daines, David et. al. 2006) and WaveNet (Van Den Oord, Aaron et. al., 2016) as offline fallback alternatives. To overcome the low accuracy of speech to text, Levenshtein distance is used to classify the commands. We also take advantage of offline hotword detection to rapidly identify important commands.



1. Huggins-Daines, David et. al. "Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices", IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 2006
2. van den oord, Aaron et. al. "WaveNet: A Generative Model for Raw Audio", SSW, 2016