UC San Diego 2017 Team Description Paper

Henrik I. Christensen, Laurel D. Riek, Ruffin White, Priyam Parashar, Shengye Wang, Tariq Iqbal, Angelique Taylor, and Darren Chan

Contextual Robotics Institute
University of California, San Diego
9500 Gilman Dr., La Jolla, CA 92093-0436

Abstract. It is challenging to deploy robots in homes. The environment is partly structured and in general the users are not robot experts. Key challenges are long-term autonomy and natural user interaction. We present our research approach on human augmented autonomy for domestic robot applications.

1 Introduction

The home is an important frontier for robotics. The environment is semi-structured, which implies that there are strong expectations about the presence of certain object categories but not their exact location. At the same time, there is a need for a high degree of autonomy as most users will not be experts on the use of robot technology / smart appliances. Achieving a high degree of autonomy in partly/semi-structured environments poses an interesting challenge. It is important to recognize that full autonomy may be too high a bar, but 99% autonomy with human assistance for the remaining 1% may be a much more viable goal at a significantly reduced complexity.

A key aspect to RoboCup @ Home is the ability to take a robot on a tour of an area, to recognize a set of objects identified by a human and later retrieve these objects. This scenario is at the center of our research and this proposal. The key challenges addressed are:

1. Generation of full autonomy through human augmentation
2. Utilization of cloud resources for tasks that are computationally challenging
3. Study of fluency in human-robot interaction

In the following sections we will present the background, approach, and expected output from our efforts.

2 Background

To accomplish the goals of this project, we will build on a body of successful prior and current work in standard and social navigation, object recognition, manipulation, and HRI.

Navigation: A fundamental aspect of any mobile robotics project is the ability to navigate in the environment without getting lost. In a domestic environment it is important to be able to generate a map of an environment on the fly and to use such dynamic maps for automatic localization. This problem is the Simultaneous Localization and Mapping (SLAM) problem. Over the last 30 years, an abundance of research has been carried on SLAM. We have developed the OmniMapper approach [1, 2] that allows for mapping and localization using a wide variety of sensors and it can generate object, feature, and grid based maps of the environment. The software is available in open-source as a ROS package. While we do not plan to perform new research on SLAM for RoboCup-2017, we plan to leverage this prior work. Another aspect of navigation involves moving meaningfully around people. Traditionally in robotics this has been explored as a person-as-obstacle, or environment as non-dynamic. Our work leverages sensing human positional information [3], modeling proxemic and group dynamics [4], and engaging in context-aware navigational paradigms [5, 6]. Navigating in a novel environment and fluent interaction with people is core to the robot-tour scenario.

Object Recognition: When navigating through the environment, a robot needs to be able to recognize a variety of objects for interaction. We have developed a package for object recognition and pose estimation [7]. The package utilizes CAD models for recognition and estimation. The appearance of an object is used for training. Once recognized the pose of the object is estimated using a set of Lie generators to fit the (CAD) model to observed edges. The pose estimate is tracked at a rate of 30 Hz. A more advanced version using RGB-D data has also been developed [8]. Recognition of natural objects on the fly is fundamental for the robot-tour scenario.

Object Manipulation: For manipulation a key aspect is grasp planning. We recently developed a planning model that performs this in a two-stage process: i) fitting a super ellipsoid to a point cloud model of the objects and ii) performing grasp planning using the super ellipsoid to reduce complexity and generate repeatable grasps [9]. The method was evaluated for handling of a variety of tabletop objects.

Human Robot Interaction: We have engaged in several projects to enable robots to understand social context as a means to springboard interaction. We have designed models of context which robots can use to perceive unstructured human environments and activities, and use that information to enable a mobile robot to automatically interact appropriately around people [6, 10]. We are exploring new models for selective attention to help robots to zoom in on important features of the environment. We can speed up existing region proposal algorithms by 30% [11]. We also have designed new non-linear algorithms to inform how robots can synthesize their behavior to cooperate, adapt to, and work with people, including non-linear models of group entrainment [12], and algorithms for robots to sense entrainment in real time and coordinate their activity with people [13]. We are also exploring human-robot adaption algorithms which can work longitudinally in home-based environments [14].

3 Robot

We plan to use the Toyota HSR robot for the contest. We have unfortunately not received the robot yet. Consequently we demonstrate the performance of the person tracking on a Fetch Research Robot. The software transfer to the Toyota HSR without changes. We use a phased array microphone plugged into an audio connector and a USB port for the speaker interface. We will use the Toyota HSR with our ROS modules installed. The robot will require internet access to access the cloud services (Amazon and Garmin 3D World) we are leveraging.

Robot's Software Description

- Platform: Secure Robot Operating System (SROS)
- Navigation, localization and mapping: OmniMapper
- Face recognition: Viola-Jones / FaceTracker
- Speech recognition: Amazon AWS / Alexa
- Speech generation: Amazon AWS / Alexa
- Object recognition: CNN based recognition / See earlier
- Social navigation: Leg Detector and proxemics based tracking/navigation

4 Approaches

In this section we briefly outline some of the key challenges addressed in our research.

4.1 Leveraging Cloud Services

A key challenge is utilization of cloud resources. The repository 3D Warehouse (former Google 3D warehouse), is a great example of models of objects in domestic environments. We will use the 3D Warehouse resources to perform object recognition and pose-estimation. We have evaluated the system on a variety of IKEA furniture [8]. The key challenge here will be scalability to 1000+ objects. We will use Convolutional Neural Nets (CNN) for the object detection. The objects used for the home-tour will be retrieved or added to the 3D Warehouse

For speech interaction with an operator and the audience we use a modified Amazon Echo API. The objective is to evaluate use of cloud based natural language processing (NLP). The underlying NLP dialog will be used both for general information enquiries "what is the weather in Nagoya today?" and for specific robot actions "please take me to the kitchen area". Speech-based dialog is integral to the home-tour scenario.

4.2 Social Navigation

We will use an evolutionary model for navigational behavior around people which gauges time and familiarity with the help of social cues, and can determine how and if the level of familiarity dictates a change in spatial proxemics between the pair. We will be building on our prior work to design socially-aware navigation strategies [3, 15, 4, 16]. Two key aspects of our approach include ego-centric person detection of body position, and leveraging an understanding of group dynamics to enable realistic motion and interaction.

Ego-centric person detection is non-trivial. We are collecting 2D and 3D leg data to tune a cascaded, simple to complex, detector. The 2D detector will be implemented by fitting an ellipse to leg measurements. The 3D detector will use a neural network trained over a window containing leg depths. In execution, the window for 3D detector will be initialized by the 2D detector.

Using this, and face detection, we plan to design an affiliation module to enable a robot to navigate to individuals or a group of people. The module will employ group dynamical principles in real-time, such as proxemics, mirroring, and familiarity, which will integrate with its navigation strategy.

4.3 Adaptive Human-Robot Teaming

In addition to social navigation, robots also need to coordinate their behaviors with people, and adapt to them in real-time during shared activities such as joint motion, shared manipulation, and so on. Building on our prior work [13, 14, 12], we will employ a non-linear approach for robots to engage in temporal adaptation and anticipation. Our method uses computational neuroscience techniques, where a linear timekeeper model can compensate for errors on a cycle-by-cycle basis, and error correction is modeled as a linear autoregressive process. The robot will employ two adaptive processes - phase correct and period correction - where the timing of the next action is adjusted to compensate for asynchrony, and the timing of the next action is modifying accordingly. Furthermore, the robot will be able to use this model to predict timing of future actions.

For example, during a collaborative manipulation task, the robot will learn on the fly to adapt its motion and behavior in a way to enable effective task completion and easy interpretation by end users. Also, since the robot is building models of human task motion, these methods will also be effective for bootstrapping learning tasks for other aspects of our project.

4.4 Autonomy

Unlike structured industrial environments where robots perform predefined tasks, domestic settings require robots to act autonomously. We propose a planning and execution framework that independently plans for the robot's actions, and overcomes difficulties or generates a new plan when the original plan is not achievable.

We divide the system into three layers of control: 1) task-level planner that generates plans for the task at hand, 2) an executor that tried to implement the plan, and 3) a reactive set of primitive actions.

Planning and Re-planning The task planner considers present state, the current mission / task and plans a sequence of actions that are encoded as a control flow graph (CFG). The post-conditions for each atomic action enable on-the-fly error detection. The planner is implemented as a D$*$ search over possible action in the task space.

Execution The control flow graph contains nodes and edges: Nodes represent primitive actions (such as moving to a particular position) and general control flow actions like an "if" statement, while edges denote transfers of control. Every action also specifies post-conditions. In the event of errors re-planning is initiated.

Diagnosis and Recovery We have several fault-handling mechanisms: 1) The plan itself contains exception detection and handlers used for common exception conditions. 2) The planner would re-plan with the updated context when the original plan is not feasible. 3) When the planner cannot find alternative approaches the user is asked for assistance to resolve the challenge. The user interaction is performed using a natural language dialog.

4.5 Integration

System components will be broken into existing frameworks. ROS provides the necessary motion planning and flexible computational graphs will be used for low level robotic control. The SROS implementations [17] is used to secure the ROS application layer. Networking and cluster access for deliberation and task coordination will use cloud APIs such as Amazon's AWS IoT.

5 Results

The results from the entry into RoboCup 2017 from UC San Diego will be made available in open source as ROS packages. Our past software has also been released into open source through our GitHub repository - `http://github.org/CognitiveRobotics`.

OmniMapper The OmniMapper system is a package that contains three components: a set of modules for feature extraction from lines and planes to RGB-D based object detection. Detected features are used for data-association. Features matched across space and time the result are integrated into a map, represented as a factor graph.

Cloud HRI We will have a few standard ROS APIs for cloud-based HRI across speech and vision. This includes a Google-based "Vision API" that does image recognition; Amazon's "Alexa Skills Kit" for voice interaction using Echo hardware; ConceptNet providing basic knowledge for the world. We will make open source ROS-wrappers with uniformed interfaces.

SROS Cybersecurity is becoming a pervasive issue as robots become ubiquitous within society. Personal robots integrating with the Internet Of Things could become targets for breaches in privacy and sources of identity theft. SROS is an addition to the ROS API and ecosystem to support cryptography and security [17].

Social Navigation The package will consist of a face-recognition system, which can distinguish people. We will provide a cascaded leg-detector. The package will provide an evolutionary navigation system that infer proxemics settings for different people depending upon familiarity levels and interaction history.

References

1. A.J.B. Trevor, J.G. Rogers, and H.I. Christensen. Omnimapper: A modular multimodal mapping framework. In *IEEE ICRA*. IEEE, 2014.
2. S. Choudhary, L. Carlone, C. Nieto, J. Rogers, Z. Liu, H. I. Christensen, and F. Dellaert. Multi robot object-based SLAM. In *Intl. Symp. on Experimental Robotics*, Tokyo, JP, Oct 2016. IFRR.
3. A. Cosgun, A. Sisbot, and H. I. Christensen. Anticipatory robot path planning in human environments. In *The 25th IEEE International Symposium on Robot and Human Interactive Communication*, New York, NY, Aug 2016. IEEE.
4. A. Taylor and L. D. Riek. Robot perception of human groups in the real world: State of the art. *In Proc. of the AAAI Fall Symposium on Artificial Intelligence in Human-Robot Interaction (AI-HRI)*, 2016.
5. P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H.I. Christensen. Navigation for human-robot interaction tasks. In *IEEE ICRA*, New Orleans, 2004.
6. A. Nigam and L.D. Riek. Social context perception for mobile robots. In *IEEE Intelligent Robots and Systems (IROS)*, pages 3621–3627, 2015.
7. C. Choi and H. I. Christensen. Robust 3d visual tracking using particle filtering on the special Euclidean group: A combined approach of keypoint and edge features. *Intl. Jour. of Robotics Research*, 31(4):498–519, 2012.
8. C. Choi and H.I. Christensen. Rgb-d object pose estimation in unstructured environments. *Robotics and Autonomous Systems*, 75(1):595–613, January 2016.
9. A. Huaman, H. Ben-Amor, and H. I. Christensen. Combining arm and hand metrics for sensible grasp modeling. In *Conf. on Automation Science and Engineering (CASE)*, Austin, TX, Aug 2016. IEEE.
10. M.F. O'Connor and L.D. Riek. Detecting social context: A method for social event classification using naturalistic multimodal data. *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, 3:1–7, 2015.
11. D. Chan, A. Taylor, and L.D. Riek. An RGB-D region proposal generator for faster robot perception. *In review*, 2017.
12. T. Iqbal and L. D Riek. A method for automatic detection of psychomotor entrainment. *IEEE Trans. on Affective Computing*, 7(1):3–16, 2016.
13. T. Iqbal, S. Rack, and L. D. Riek. Movement coordination in human-robot teams: A dynamical systems approach. *IEEE Trans. on Robotics*, 32(4), 2016.
14. T. Iqbal, M. Moosaei, and L.D Riek. Tempo adaptation and anticipation methods for human-robot teams. *Robotics, Science and Systems (RSS) 2016 Workshop on Planning for Human-Robot Interaction*, 2016.

15. G. Ferrer, A.G. Zulueta, F.H. Cotarelo, and A. Sanfeliu. Robot social-aware navigation framework to accompany people walking side-by-side. *Autonomous Robots*, pages 1–19, 2016.
16. A. Taylor and L. D. Riek. Robot affiliation perception for social interaction. *In Review*, 2017.
17. R. White, M. Quigley, and H.I. Christensen. SROS: Securing ROS over the wire, in the graph, and through the kernel. In *Humanoids Workshop: Towards Humanoid Robots OS*. Cancun, Mexico, 2016.
18. F. Bonsignorio and A.P. del Pobil. Toward Replicable and Measurable Robotics Research. *IEEE Robotics & Automation Magazine*, 22(3):32–35, 2015.
19. E. Guglielmelli. Research Reproducibility and Performance Evaluation for Dependable Robots. *IEEE Robotics & Automation Magazine*, 22(3):4–4, 2015.
20. E. Pacchierotti, H. I. Christensen, and P. Jensfelt. Design of an office-guide robot for social interaction studies. In *Intl Conf on Intelligent Robots and Systems (IROS)*, Beijing, China, October 2006. RSJ/IEEE.
21. A. Cosgun, D.A. Florencio, and H.I. Christensen. Autonomous person following for telepresence robots. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 4335–4342. IEEE, 2013.
22. F. Dellaert. Factor graphs and gtsam: A hands-on introduction. Technical report, Georgia Institute of Technology, 2012.