

# The Robotic Busboy: Steps Towards Developing a Mobile Robotic Home Assistant

Siddhartha SRINIVASA <sup>a</sup>, Dave FERGUSON <sup>a</sup>, Mike VANDE WEGHE <sup>b</sup>,  
Rosen DIANKOV <sup>b</sup>, Dmitry BERENSON <sup>b</sup>,  
Casey HELFRICH <sup>a</sup>, and Hauke STRASDAT <sup>c,1</sup>  
<sup>a</sup> *Intel Research Pittsburgh, Pittsburgh, USA*  
<sup>b</sup> *Carnegie Mellon University, Pittsburgh, USA*  
<sup>c</sup> *University of Freiburg, Freiburg, Germany*

**Abstract.** We present an autonomous multi-robot system that can collect objects from indoor environments and load them into a dishwasher rack. We discuss each component of the system in detail and highlight the perception, navigation, and manipulation algorithms employed. We present results from several public demonstrations, including one in which the system was run for several hours and interacted with several hundred people.

**Keywords.** Manipulation, Navigation, Grasping, Autonomous Systems

## Introduction

One of the long-term goals of robotics is to develop a general purpose platform that can co-exist with and provide assistance to people. Substantial progress has been made toward creating the *physical* components of such an agent, resulting in a wide variety of both wheeled and humanoid robots that possess amazing potential for dexterity and finesse. However, the development of robots that can act autonomously in unstructured and inhabited environments is still an open problem, due to the inherent difficulty of the associated perception, navigation, and manipulation problems.

To usefully interact in human environments, a robot must be able to detect and recognize both the environment itself and common objects within it, as well as its own position within the environment. Robust and safe navigation approaches are required to effectively move through populated environments; a robot must differentiate between moving objects (e.g. people) and static objects, and know when it is appropriate to avoid or interact with people. Finally, complex manipulation techniques are required to interact with household objects in the environment; a robot must cope with the complexity of objects as well as errors in perception and execution. Most importantly, to function

---

<sup>1</sup>Corresponding Authors: Siddhartha Srinivasa and Dave Ferguson, Intel Research Pittsburgh, E-mail: {siddhartha.srinivasa, dave.ferguson}@intel.com



**Figure 1.** The Robotic Busboy composed of a Segway RMP200 and a Barrett WAM arm, with a snapshot from simulation and reality.

seamlessly in dynamic human environments, all of these actions must be planned and executed quickly, at human speeds.

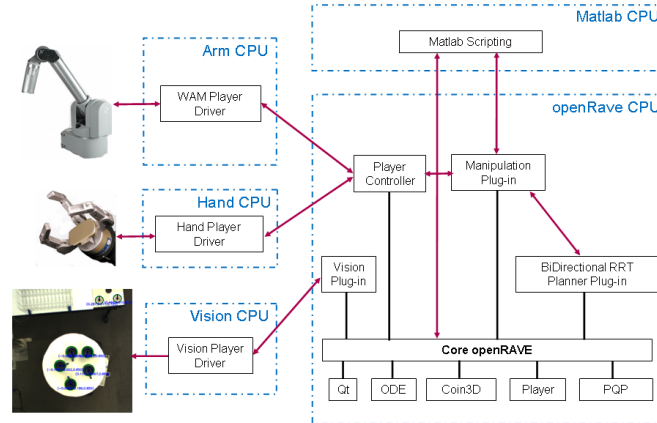
Currently, although several researchers are working on robotic systems to satisfy some portion of these requirements for specific tasks, very few [1,2,3] have tried to tackle all of them in conjunction or achieved real autonomy. We focus our current work toward achieving this goal.

The Robotic Busboy is a project intended to encompass several of the challenges inherent in developing a useful robotic assistant while restricting the scope of the tasks performed. We concentrate on a dishwasher loading scenario in which people moving around an indoor environment place empty cups on a mobile robot, which then drives to a kitchen area to load the mugs into a dishwasher rack (Figure 1). This specialized task requires robust solutions to several of the challenges associated with general robotic assistants and we have found it to be a valuable domain for providing new research problems.

In this paper, we describe a multi-robot system to tackle this problem. We first introduce the architecture of the system and the interaction between the various components. We then discuss each component in detail and highlight the perception, navigation, and manipulation algorithms employed. We present results from several public demonstrations, including one in which the system was run for several hours and interacted with several hundred people, and provide a number of directions for future research.

## 1. System Architecture

In our approach to the Robotic Busboy task, we have a Segway mobile robot navigating through the environment collecting empty mugs from people. The Segway then transports the collected mugs to the vicinity of a Barrett WAM robotic arm which, in turn, detects and removes the cups from the mobile robot and loads them into a dishwasher rack. The overall system is divided into three main components: Segway navigation, vision-based mug and Segway detection, and grasp planning and arm motion planning.



**Figure 2.** System Architecture Diagram.

We have developed a plugin-based architecture called OpenRAVE, released publicly on Sourceforge [4], that provides both basic services like collision detection and physics simulation, as well as a novel scripting environment that allows for the seamless interaction of many components to perform a combined task. Furthermore, the simulation environment provided by openRAVE allows for the testing of all subsystems via virtual controllers.

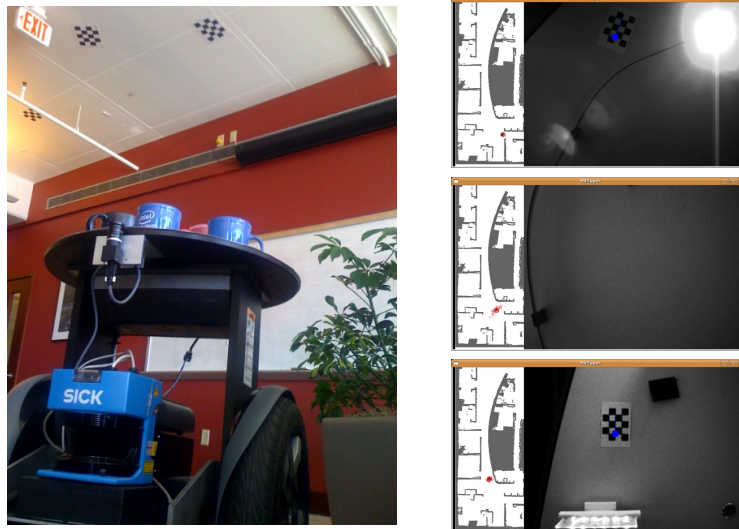
Figure 2 illustrates the high-level architecture of the manipulation portion of our system. Feedback and control of the arm and hand is achieved by a controller plugin which interfaces with Player/Stage [5] drivers for each hardware component. The vision plugin updates the scene with the pose of the Segway and the mugs. The manipulation plugin oversees the grasping and motion planning tasks and instructs the planner plugin to plan appropriate collision-free motions for the arm.

The following sections describe in detail the algorithms implemented in the individual subsystems as well as their interactions with each other.

## 2. Mobile Robot Navigation

To reliably navigate through highly populated indoor environments, the Segway needs to localize itself accurately within the environment and detect and avoid people and other obstacles that may be present. To facilitate this, the Segway is equipped with both a laser range finder and an upwards-pointing monocular camera.

For localization, we first build offline a 2D map of the (unpopulated) environment using the laser range finder (see Figure 4, left image for an example such map). This provides an accurate representation of the static elements in the environment and can be used effectively for determining the pose of the Segway during navigation in static environments (see, for instance, [6]). However, in highly populated environments such an approach is of limited accuracy, since the a priori 2D map doesn't accurately represent the dynamic elements encountered by the robot in the environment and these elements can significantly restrict the amount of the static environment that can be observed at any time.



**Figure 3.** Camera-based indoor localization using ceiling markers. On the left is a view of the Segway with the upward-pointing camera and the ceiling markers. Images on the right show snapshots from localization performed using the ceiling markers. As the vehicle moves between the markers its position error accrues (particles from its particle filter-based localization are shown in red), but collapses upon reaching a marker.

To improve upon this accuracy, in populated environments we use a vision-based localization approach, exploiting an area of the environment that is never populated and thus never changing: the ceiling [7,8]. To do this, we place unique markers, in the form of checkerboard patterns, at various locations on the ceiling and store the pose of each marker with respect to the 2D map constructed offline<sup>2</sup>. We then use an upwards-pointing camera on the Segway to detect these markers and thus determine the pose of the Segway in the world (see Figure 3). This provides very good global pose information when in the vicinity of one of the markers, and in combination with odometry-based movement estimation provides a reliable method for determining the pose of the Segway in all areas of the environment.

For detecting and avoiding people and other objects that may not have existed in our prior 2D map, we use the laser range-finder to provide local obstacle information. This information is then combined with the prior 2D map and used when planning paths for the Segway. For planning we use the open-source Player implementations of a global wavefront planner and a local vector-field histogram planner [9]. The global planner is used to provide a nominal path to the goal, or value function, for the Segway and the local planner is used to track this nominal path by generating dynamically feasible, collision-free trajectories for the Segway to execute. Together, the global and local planners enable the Segway to reliably navigate through large populated indoor environments and find its way to the manipulator to deliver its collection of mugs.

<sup>2</sup>It is also possible to use ceiling-based localization without requiring markers, but for our application we found the marker-based approach to be easy to set up and very reliable.



**Figure 4.** Laser-based obstacle avoidance. A person stands between the segway and its desired location. The person is detected by the Segway’s laser range finder and is avoided.

### 3. Robotic Manipulation

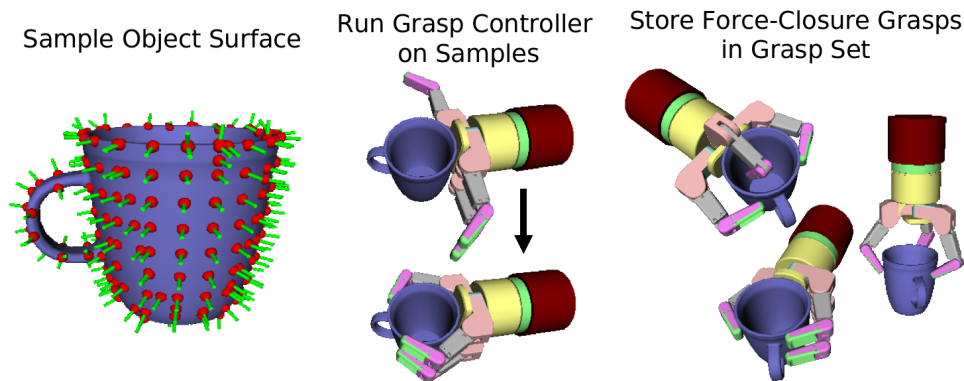
After successfully navigating to the arm, the Segway and the mugs are registered by the robot using a camera mounted on the ceiling. Once registered, the arm unloads the mugs onto a dishwasher rack. The remainder of this section details our algorithms for detecting and manipulating the mugs and the tradeoffs between planning and execution speed.

The vision system tracks the 2D position and orientation of the Segway, mugs, and table in real-time while allowing users to add, remove or move mugs at any point during the demonstration. We used normalized cross-correlation template matching to detect the center of each mug and to find the orientation of the mugs we then search in an annulus around the center of each mug for a color peak corresponding to the mug’s handle.

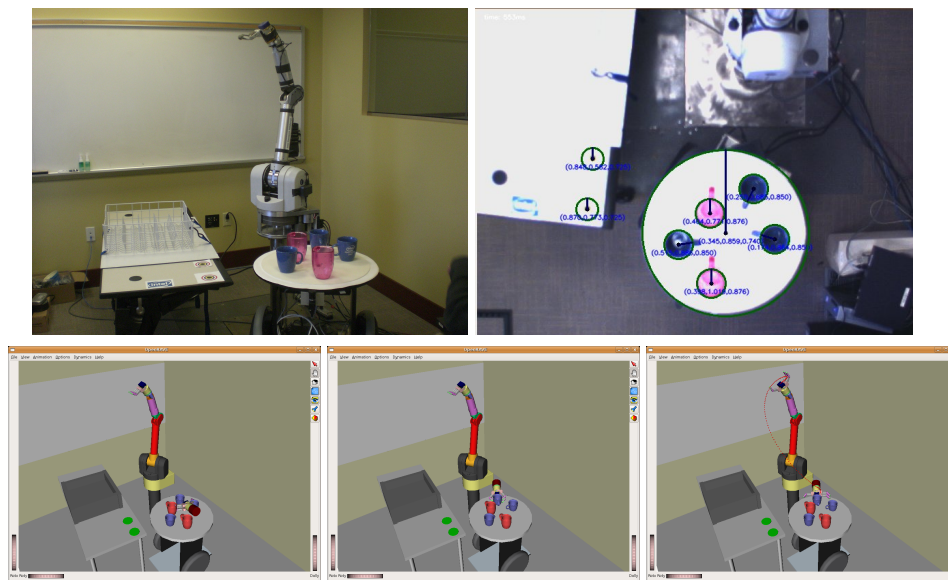
We have developed a grasping framework that has been successfully tested on several robot platforms [10]. Our goal for the grasp planner is to load the mugs with minimal planning delay. To achieve this, we perform most of the heavy computation offline by computing a grasp set comprising of hundreds of valid grasps, in the absence of any obstacles (Figure 5). Once a scene is registered with multiple mugs, we first select a candidate mug based on the proximity to the arm. The planner then efficiently sorts the mug’s grasp set based on various feasibility criteria like reachability and collision avoidance, and tests the sorted grasps. Once a grasp is chosen, the arm plans to an empty position in the dish rack using a bi-directional RRT, and releases the mug. An illustration of the entire algorithm is shown in Figure 6.

Generated paths were smoothed and communicated across Player to the WAM. The grasp controller simply closed each finger until it exerted a minimum torque, securing the grasp. Once a cup is grasped, the configuration of the current hand is compared to the expected configuration for validation of a successful grasp. If the two configurations are substantially different, the arm replans the trajectory.

We added several heuristics specific to our problem. For aesthetic appeal, we prioritized loading the mugs face-down in the rack. The planner switched to the next best mug if it was unable to find a collision-free path within a given time. If the planner fails to repeatedly find grasps that put a mug face down in the rack, it would remove the face-down constraint and search for grasps regardless of final orientation. When the planner is no longer able to find any valid spot in the dish rack to place a mug, it requests for the dishwasher rack to be unloaded.



**Figure 5.** Offline grasp generation. Grasps are generated offline for a known object and stored for online evaluation.



**Figure 6.** The real scene on the top left is registered by the ceiling camera on the top right and fed to OpenRAVE. The first grasp on the bottom left is invalidated by collision but the second one succeeds, triggering the arm planner which plans a collision-free path for the arm.

#### 4. Results

The Robotic Busboy system has been developed and tested incrementally for more than 8 months. It has operated in populated environments dozens of times, often for several hours at a time. The largest demonstration, both in terms of duration and audience, was at the 2007 Intel Open House in October where the system operated continuously for four hours and interacted with hundreds of people. Figure 7 shows snapshots of the arm removing cups from the Segway and loading them into the dishwasher rack.

During the Open House demonstration the robot dropped or failed to pick up about 20 mugs and was successful about 200 times in picking up a mug and loading it into the

dish rack. In most of the failures, the robot realized that it had not picked up the mug and indicated failure. In a more structured experiment with different placements of 4 mugs, we measured 19 out of 20 successes. We observed that failure to pick up a mug could be due to incorrect calibration of the arm, inaccuracies due to the loosening of cables that drive the arm, mis-registration of the mug handle by the vision system, and when someone moved the mugs when the arm was in motion to pick one up.

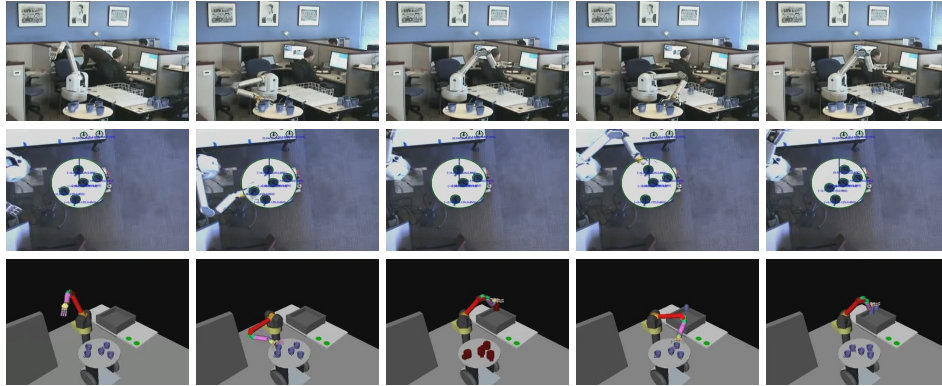
In our structured experiment, the average total time from giving the order to pickup a cup to the cup being released in the dishrack was measured to be 51 seconds, with a standard deviation of 4.9 seconds. We noted that execution times remained approximately consistent regardless of the number of mugs the arm had to consider. We believe that this is a testament to the dexterity of our arm and our grasp planner. We observed that the most challenging arrangement is with the mugs close together with the handles pointing outward.

As is the case for most complex autonomous systems, it was not until the entire system was functional and stress-tested that many of the key challenges to grasping and failure recovery were discovered. For example, if the arm fails to grasp a cup, it has to restart its grasping script gracefully without any human intervention. This was achieved by constantly comparing the hand's predicted encoder values vs the real encoder values. Using grasps that are robust against errors in the mug pose and arm configuration really helped increase the success rate. We picked these grasps by randomly moving the cups in simulation and testing whether force closure still persists. This ability to recover from error was very important for successful continuous operation, particularly in the presence of people endeavoring to test the system in challenging scenarios. The same technique is used to detect, and recover from, the case when a person has physically removed the cup from the hand of the robot.

## **5. Conclusions**

We have presented an autonomous multi-robot system designed to play the role of a robotic busboy. In our system, a mobile robot navigates through a populated environment receiving empty cups from people, then brings these cups to a robotic manipulator on a fixed platform. The manipulator detects the cups, picks them up off the mobile robot, and loads them into a dishwasher rack. Although a specialized task, this problem requires robust solutions to several of the challenges associated with general robotic assistants and we have found it to be a valuable domain for providing new research problems and general purpose perception and planning algorithms. Our system is entirely autonomous and has interacted with people in dozens of public demonstrations, including one in which it was run for several hours and interacted with several hundred people.

We are currently working on adding compliance to the arm to enable smoother interaction with unexpected obstacles and to allow for humans to work in the arm's workspace and interact with the arm comfortably. While the Segway and the arm currently function as one coordinated robot system, we are also working on integrating them physically into a mobile manipulator and on migrating all of the sensing onboard. We believe that these improvements will enable us to perform complex mobile manipulation tasks like opening doors and cabinets, and interacting with people even more closely in unstructured human environments.



**Figure 7.** Removing cups from the Segway and loading them into a dishwasher rack during the Intel Open House Demonstration. The top images show the environment, the center images show the result from the ceiling camera vision system, and the bottom images show the corresponding OpenRAVE model. Note that while the system is running, users have added extra cups to the Segway and the system simply updates its model and continues.

We believe that the ability to plan robustly under uncertainty is a compelling challenge. We are currently working on active vision for information gain, kinesthetic sensing on the hand and the arm for fine motion control while grasping, and on long-range people prediction and tracking for autonomous navigation in indoor spaces.

### Acknowledgements

We are grateful for the contributions of James Kuffner, Ali Rahimi, and Chris Atkeson, and for the technical support of Barrett Technologies and Will Pong of Segway Inc.

### References

- [1] M. Quigley, E. Berger, and A. Y. Ng, "STAIR: Hardware and software architecture," in *AAAI Robotics Workshop*, 2007.
- [2] H. Nguyen, C. Anderson, A. Trevor, A. Jain, Z. Xu, and C. Kemp, "El-e: An assistive robot that fetches objects from flat surfaces," in *Human Robot Interaction, The Robotics Helpers Workshop*, 2008.
- [3] "Personal robotics program," [www.willowgarage.com](http://www.willowgarage.com).
- [4] R. Diankov *et al.*, "The OpenRAVE project," hosted at <http://sourceforge.net/projects/openrave/>.
- [5] B. Gerkey *et al.*, "The Player/Stage project," hosted at <http://playerstage.sourceforge.net>.
- [6] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1999.
- [7] W. Burgard, A. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1998.
- [8] S. Thrun *et al.*, "MINERVA: A second generation mobile tour-guide robot," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1999.
- [9] I. Ulrich and J. Borenstein, "Vfh+: Reliable obstacle avoidance for fast mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1998.
- [10] D. Berenson, R. Diankov, K. Nishiwaki, S. Kagami, and J. Kuffner, "Grasp planning in complex scenes," in *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2007.