Interactive Object Modeling & Labeling for Service Robots

Alexander J. B. Trevor, John G. Rogers III, Akansel Cosgun, Henrik I. Christensen

Abstract— **We present an interactive object modeling and labeling system for service robots. The system enables a user to interactively create object models for a set of objects. Users also provide a label for each object, allowing it to be referenced later. Interaction with the robot occurs via a combination of a smartphone UI and pointing gestures.**

I. INTRODUCTION

Object recognition is an important capability for many service robotic tasks. Personal service robots may need to be able to recognize different objects in different environments. Object recognition with large numbers of objects can be computationally expensive, so choosing a relevant subset of objects to model is advantageous. We propose a system that enables a user to interactively build object models for selected objects, and annotate these models with a label. This enables the user to easily command the robot to build a small object database of the specific objects it will need to interact with. Additionally, because the user specifies the label, users can establish common ground with the robot for future references to modeled objects.

The video demonstrates our interactive object modeling and labeling system. First, the robot is teleoperated through the environment to build a map using our SLAM system. The resulting map contains the locations and extent of planar surfaces, which will be used by our system to detect tabletop objects. The robot can be commanded to model an object by using a smartphone UI to enter the desired label, and then pointing at an object on a table. The pointing gesture is then recognized. The user's elbow and hand define a ray, which is tested for intersections with nearby planar surfaces such as tables. If such an intersection is found, object segmentation is performed at the referenced location. The nearest detected object to the referenced point is modeled and annotated with the provided label, and added to the object database.

II. SYSTEM OVERVIEW

A. Mapping System

The presented video builds on our previous work on SLAM [3], semantic mapping [1], and interactive annotation of maps [2]. Our mapping system allows the robot to navigate and remain localized in it's environment. It additionally contains a list of planar surfaces, which is used in this work to assist in the object segmentation, as we assume objects are supported by one of these surfaces (such as a table or shelf). The planar surfaces can be used as landmarks for SLAM, but also provide useful semantic information about the environment, such as locations of tables, shelves, and walls.

B. Object Modeling System

The object modeling system consists of several components: object segmentation, feature extraction, and modeling. Once a user has given a referenced point at which to detect objects, the robot moves its pan-tilt unit to aim the sensors at this location. Object segmentation is performed using point cloud data from an RGBD sensor. First, the large planar surface corresponding to the table is detected. This is removed from the point cloud, and point clusters above this are detected. The cluster with a centroid nearest to the reference point is selected as the object to be modeled. The cluster's points are projected into our DSLR's image, and are used to generate a region of interest. SURF features are detected for the region of interest, and are stored as an object model along with the provided label.

C. Gesture Recognition & Smartphone UI

Users interact with the system using a combination of a smartphone user interface and pointing gestures. The phone UI communicates with the robot over wireless ethernet, which can forward commands to other software modules on our robot. The application was implemented using Apple XCode and iOS.

The gesture recognition system uses a Kinect-like Asus Xtion Pro sensor with the OpenNI skeleton tracking library. When the user enters a label into the smartphone UI and presses the "label object" button, the gesture recognition system attempts to detect a pointing gesture. To be recognized as a pointing gesture, the user's arm must not be vertical (not at their side) and must remain stationary for a short time period.

III. ACKNOWLEDGMENTS

This work was made possible through financial support from the Boeing corporation and ARL MAST CTA project 104953.

REFERENCES

[1] J.G. Rogers, A.J.B. Trevor, C. Nieto-Granda, and H.I. Christensen. Simultaneous localization and mapping with learned object recognition and semantic data association. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 1264–1270. IEEE, 2011.

[2] A. J. B. Trevor, A. Cosgun, J. Kumar, and H. I. Christensen. Interactive Map Labeling for Service Robots. In *IROS Workshop on Active Semantic Perception*, 2012.

[3] A. J. B. Trevor, J. G. Rogers III, and H. I. Christensen. Planar Surface SLAM with 3D and 2D Sensors. *International Conference on Robotics and Automation*, 2012.

A. Trevor, J. G. Rogers III, A. Cosgun, and H. I. Christensen are affiliated with the Georgia Institute of Technology's center for Robotics & Intelligent Machines. {atrevor, jgrogers, akanselcosgun}@gatech.edu, hic@cc.gatech.edu