Robotics and Autonomous Systems 56 (2008) 927–941

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Towards 3D Point cloud based object maps for household environments

Radu Bogdan Rusu *, Zoltan Csaba Marton, Nico Blodow, Mihai Dolha, Michael Beetz

Technische Universität München, Computer Science Department, Intelligent Autonomous Systems Group, Boltzmannstr. 3, 85748, Garching bei München, Germany

ARTICLE INFO

Article history:
Available online 20 August 2008

Keywords:
Environment object model
Point cloud data
Geometrical reasoning

ABSTRACT

This article investigates the problem of acquiring 3D object maps of indoor household environments, in particular kitchens. The objects modeled in these maps include cupboards, tables, drawers and shelves, which are of particular importance for a household robotic assistant. Our mapping approach is based on PCD (point cloud data) representations. Sophisticated interpretation methods operating on these representations eliminate noise and resample the data without deleting the important details, and interpret the improved point clouds in terms of rectangular planes and 3D geometric shapes. We detail the steps of our mapping approach and explain the key techniques that make it work. The novel techniques include statistical analysis, persistent histogram features estimation that allows for a consistent registration, resampling with additional robust fitting techniques, and segmentation of the environment into meaningful regions.

1. Introduction

In this article we investigate how an autonomous mobile robot (see Fig. 1) can acquire an environment object model of a kitchen, and more generally of human living environments. The acquired environment model is to serve as a resource for robotic assistants that are to perform household chores including setting the table, cleaning up, preparing meals, etc. For tasks like these, it is not sufficient to have models that serve localization and navigation purposes. Rather the robot has to know about cupboards, tables, drawers, shelves, etc, how to open and use them, what is stored in them, etc. Finally, the models have to represent ovens, dish washers and other appliances as specific kinds of cupboards.

Our scenario is as follows: a robot enters the room and repeatedly turns and makes a sweeping movement with its left arm that has a a laser sensor mounted on its end effector (see Fig. 1). Storing the individual laser readings during the sweep and combining them with the arm's joint angles results in a 3D laser scan. The scan is converted to a point cloud where each point is represented by a tuple containing: the 3D position in world coordinates $\langle x, y, z\rangle$, intensity and distance values $\langle i, d\rangle$. After processing, integrating, and interpreting the individual scans, the robot transforms the individual point clouds into an environment object model (simply called *object map*) as it is depicted in Fig. 2. This object map consists of a 3D mesh visually representing the parts of the environment that the robot considers as obstacles that it must not collide with, and cuboids and planes that represent the objects that are relevant for the robot's tasks — appliances, cupboards, drawers, and tables.

We represent the object map using OWL-DL[1] such that the environment models can be easily shared between different applications and robots, and that a robot may query the map in order to retrieve necessary information from the model. The objects in our map are hierarchically structured, such that kitchenettes consist of connected cupboards, a single cupboard is a box with a door and a handle. The objects are also instances of object classes where the object classes are organized in a specialization hierarchy. Thus, a cupboard is a specialization of a container and can therefore be used to store objects. Or, an oven is a specific kind of container that is used to prepare meals.

The main challenge of this scenario is the transformation of raw point cloud data into a meaningful, compact representation of the world, without losing fine grained details necessary for tasks like robot manipulation and activity recognition. The key contributions of this article include a novel multi-dimensional tuple representation for point clouds and robust, efficient and accurate techniques for computing these representations, which facilitate the creation of hierarchical object models for kitchen environments.

The tuples contain informative point feature histograms, persistent feature values, and region connectivity information, among other information. The techniques which operate on them include: a robust and fast registration using persistent feature histograms, robust polynomial resampling methods, and segmentation and extraction of objects in the map.

* Corresponding author. Tel.: +49 08928917780; fax: +49 08928917757.
E-mail addresses: rusu@cs.tum.edu (R.B. Rusu), marton@cs.tum.edu (Z.C. Marton), blodow@cs.tum.edu (N. Blodow), dolha@cs.tum.edu (M. Dolha), beetz@cs.tum.edu (M. Beetz).

[1] We use ResearchCyc as our basic encyclopedic knowledge base and add the object classes that are missing and needed for representing our maps. ResearchCyc knowledge bases can be exported into OWL-DL, which makes them usable as Semantic Web knowledge sources, and applicable to our map representation.

0921-8890/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2008.08.005

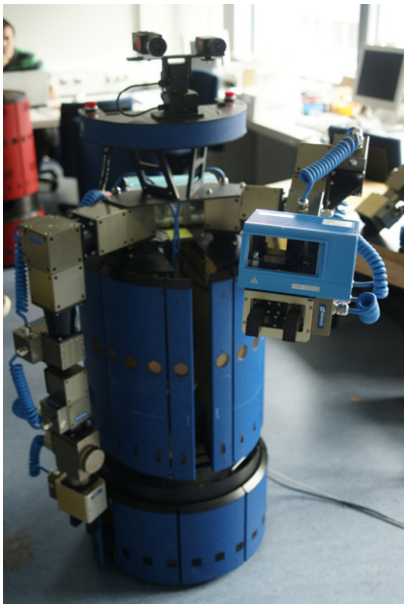

Fig. 1. Autonomous robot used for the experiments in this article. A laser scanner is mounted on the arm, which allows the robot to acquire 3D scans of the environment.

We show the effectiveness of the contributions, and their usage in a comprehensive application scenario. This includes constructing a complete point cloud model of the environment, the extraction of informative features for geometrical reasoning, segmentation into classes of objects, and supporting descriptive queries useful in high level action planning.

The remainder of the article is organized as follows. Section 2 presents related work, followed by a description of our system overview and architecture in Section 3. Sections 4 and 5 explain the technicalities of the two main modules of our system, namely Geometrical Mapping and Functional Mapping. We briefly reiterate our empirical results in Section 6, and conclude with Section 7.

2. Related work

Several efforts in the past have been made regarding the creation of environment object maps out of 3D range data. Since the creation of such maps is a highly complex process and involves the combination of several algorithms, we will try to address the most relevant publications for our work below. Related work on particular dimensions will be addressed in their respective technical sections.

An EM-based algorithm for learning 3D models of indoor environments is presented in [18]. The maps are created using mobile robots equipped with laser range finders, but they do not include any semantic information. The work in [13] uses a stereoscopic camera system and a knowledge base in the form of a semantic net to form 3D models of outdoor environments. Two parallel representations, one spatial and one semantic, are proposed in [8] for an indoor environment, but their approach needs further investigation. An object based approach for cognitive maps is used to recognize objects and classify rooms in different categories in [38]. The work presented in [21] provides a method for classifying different places in the environment into semantic classes like doorways, kitchens, corridors, rooms using simple geometric features extracted from laser data and information extracted from camera data. The semantic interpretation in [24] takes into account generic architectural elements (floor, ceiling, walls, doors) which are identified based on the relationships between the features (parallel, orthogonal, above, under, equal height). With few exceptions, in particular in the area of cognitive mapping [38,20], but also including [37], maps do not represent objects relevant for other robot tasks, besides navigation.

Since the input data leading to the creation of the object map has to be acquired in stages, an algorithm for automatically aligning (registering) the separate views into the same model is needed. One of the most popular registration methods to date is the Iterative Closest Point (ICP) algorithm [5,41]. The original version uses pairs of nearest 3D points in the source and model set as correspondences, and assumes that every point has a corresponding match. Obviously this is rarely the case with most applications, but a simple distance threshold can be used to disregard correspondences in order to deal with partially overlapping scenes. Additionally, to further improve the quality of correspondences, a lot of efforts have been made into the area of feature selection [10,35], as well as including extra information such as colors [17] or normals [2] that could improve the correspondence problem. Since ICP requires an exhaustive search through the correspondence space, several variants that address the problem of its computational complexity have been proposed [28,22]. Furthermore, methods for generating initial approximate alignments [10,35,19], as well as choosing alternate optimization methods [14,6], improved the results of the registration process. [23] presents a system for 3D laser scan registration using ICP based on semantic information (e.g. walls, floors, ceilings), which decreases the computation time by up to 30%. Our registration algorithm is based upon the previously mentioned work, and adds several extensions. In the following we will discuss our reasons and justify the need for these extensions.

While feature candidates such as surface curvature estimation [2] or integral volume descriptors [10] have already been used

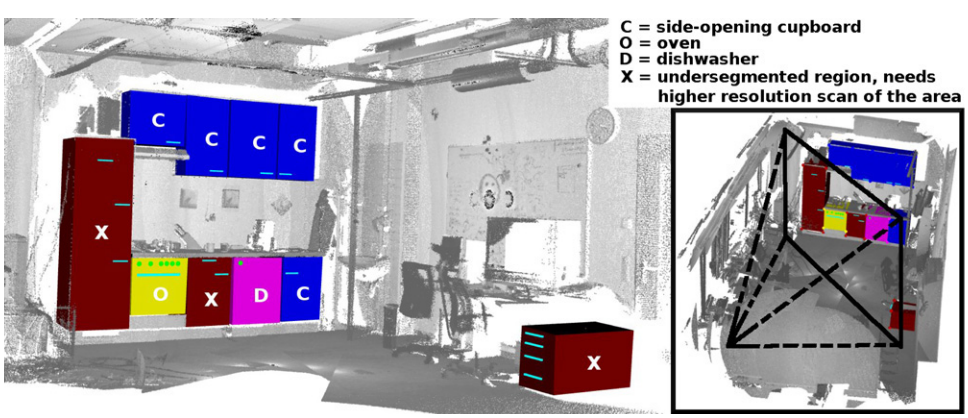

Fig. 2. An object map of the kitchen environment with superimposed object classes (remission values shown in grayscale).