Artificial Intelligence ••• (••••) •••–•••

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Model-based furniture recognition for building semantic object maps

Martin Günther [a,*], Thomas Wiemann [a], Sven Albrecht [a], Joachim Hertzberg [a,b]

[a] Institute of Computer Science, Osnabrück University, Albrechtstr. 28, 49076 Osnabrück, Germany
[b] DFKI Robotics Innovation Center, Osnabrück Branch, 49076 Osnabrück, Germany

ARTICLE INFO

Article history:
Received in revised form 16 December 2014
Accepted 20 December 2014
Available online xxxx

Keywords:
Semantic map
Incremental mapping
Closed-loop mapping
Model-based object recognition
3D point cloud
CAD model matching
OWL-DL ontology

ABSTRACT

This paper presents an approach to creating a semantic map of an indoor environment incrementally and in closed loop, based on a series of 3D point clouds captured by a mobile robot using an RGB-D camera. Based on a semantic model about furniture objects (represented in an OWL-DL ontology with rules attached), we generate hypotheses for locations and 6DoF poses of object instances and verify them by matching a geometric model of the object (given as a CAD model) into the point cloud. The result, in addition to the registered point cloud, is a consistent mesh representation of the environment, further enriched by object models corresponding to the detected pieces of furniture. We demonstrate the robustness of our approach against occlusion and aperture limitations of the RGB-D frames, and against differences between the CAD models and the real objects. We evaluate the complete system on two challenging datasets featuring partial visibility and totaling over 800 frames. The results show complementary strengths and weaknesses of processing each frame directly vs. processing the fully registered scene, which accord with intuitive expectations.

1. Introduction

1.1. Closed-loop incremental semantic mapping

Building 3D maps of indoor environments by mobile robots has received increasing interest since the launch of inexpensive 3D sensors such as the Microsoft Kinect. Several successful approaches exist that generate 3D point cloud maps (e.g., [1,2]) or mesh representations (e.g., [3]) based on RGB-D data. Yet, automatically providing additional *semantic* information to the maps, such as location and type of furniture present, is still not well understood. Information on a semantic level, however, is necessary for many advanced tasks of an autonomous robot, such as object search or place recognition. Also, it has advantages for the map building process itself: If the class and location of objects in the map are known, object models could be used to hypothesize missing sensor data, or loop-closing in mapping can be based on semantic as well as geometric information.

A semantic map is necessarily hybrid in the classical sense of Kuipers [4], including at least geometric information and semantic knowledge [5]. Intuitively, the process of generating a semantic map (semantic mapping, for short) should be

* Corresponding author.
E-mail addresses: martin.guenther@uni-osnabrueck.de (M. Günther), thomas.wiemann@uni-osnabrueck.de (T. Wiemann), sven.albrecht@uni-osnabrueck.de (S. Albrecht), joachim.hertzberg@uni-osnabrueck.de (J. Hertzberg).

http://dx.doi.org/10.1016/j.artint.2014.12.007
0004-3702/

closed-loop and incremental. *Closed-loop* means that object recognition and labeling in the sensor data ("bottom-up") should not strictly precede processing on the semantic level, but that knowledge and reasoning on the semantic level should be able to influence object classification, recognition and the mapping process as a whole ("top-down"). For example, the information that some room is an office room should lead reasoning on the semantic level to hypothesize that certain types of objects are likely or unlikely to be present, respectively; such hypotheses then bias or guide bottom-up sensor data processing. *Incremental* means that the semantic map building process does not have to wait for the sensor data of some scene or environment to be complete (no matter how such completeness would have to be defined and determined), but has to start right away, based on individual sensor takes, such as single RGB-D frames or 3D laser scans. This expectation fits with the closed-loop property, as the increase of environment knowledge in the semantic map over the mapping process is supported by both sensor data and prior knowledge, e.g., about object classes and their relations. Incrementality poses a challenge, though, as such individual sensor takes suffer greatly from occlusions and limitations due to sensor aperture or view pose constraints – in addition to the unavoidable regular sensor noise.

Closed-loop, incremental semantic mapping is currently not well understood. There is quite some body of literature about its ingredients, as will be discussed in the Related Work section; however, there are only few systems doing it in integration. This paper contributes a detailed case study of such a system. It presents an approach to semantic mapping that:

1. reconstructs the surfaces from noisy 3D data, captured from a Kinect camera, and creates a triangle mesh;
2. recognizes furniture objects in the point clouds based on structural descriptions from an OWL-DL ontology;
3. and finally adjusts their poses using ICP, and augments the created map with CAD models corresponding to the furniture objects.

We call *model-based object recognition* the ensemble of these three steps, used in integration with a knowledge base (given in OWL-DL, in this case) and a module for building geometric 3D point cloud maps. Using state of the art SLAM algorithms, these annotated point clouds can then be used to maintain a consistent semantic map of the complete environment, consisting of both the geometry and the semantic knowledge.

We would like to emphasize the role that using a formally well-understood knowledge representation and reasoning (KR&R) formalism plays in closed-loop, incremental semantic mapping, rather than using some ad-hoc set of object labels. Using arbitrary labels like "table", "tasse", or "q17", which may or may not have a meaning for humans, may suffice for purely bottom-up recognition, classification, or labeling of segments of sensor data. Whenever the intention is to reason with and about objects or events perceived by the robot, though, using some KR&R formalism with a well-defined semantics and, ideally, efficient reasoners available is the obvious choice. Such reasoning is needed for the top-down part of closed-loop semantic mapping in the first place; it may be employed in other robot tasks using the previously acquired semantic map, such as object search (e.g., [6,7]), human robot interaction on a high conceptual level [8], detecting norm violations [9].

Many researchers in semantic mapping have recently been using description logics (DL, [10]) as such a formalism, in particular the DL variants OWL-DL and OWL-lite, as available in the OWL W3C standard [11]; we are using OWL-DL in the work reported here, too. DL is an obvious choice for a KR&R formalism in semantic mapping, as it allows to represent and reason about object ontologies, providing a structured representation of object classes and instances, but of some relations between objects, too, which the declarative part of a semantic map is expected to contain. Existing DL reasoners provide reasoning services such as consistency checking and subsumption within an ontology for free and in a highly optimized way. They allow sound inferences to be made across all hierarchical levels of the ontology without further effort. For example, questions like "How many pieces of furniture does room R contain?", or "Which pieces of furniture on this floor are suitable storage places for a milk jug?" could be answered right away, based on perceptions of individual chairs, tables, shelves and so on. As many representation and reasoning problems in robotics naturally include uncertainty, such as by sensor noise and/or interpretation uncertainties, several researchers have recently embedded the ontological reasoning provided by DL reasoners into probabilistic frameworks like Markov Logic Networks [12,13] or Bayesian Logic Networks [14]. The bottom line here is:

1. Using a well-founded KR&R formalism for representing and reasoning in the semantic part of a semantic map is strongly advised, if not needed, in semantic mapping; previous AI work in KR&R has yielded a wide variety of such formalisms that are ready to be used.
2. Variants of DL have been used in much of the recent semantic mapping research, and we have done so in the work reported here.

"Pure" DL is certainly not the final word regarding a suitable KR&R formalism, as it cannot well handle uncertainty and n-ary relations, just to mention two points. Identifying or developing more fitting formalisms is an important issue for interdisciplinary research between AI and Robotics, which we recommend to put on the common agenda, but do not intend to detail in this paper.

Please cite this article in press as: M. Günther et al., Model-based furniture recognition for building semantic object maps, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2014.12.007