Robotics and Autonomous Systems 56 (2008) 915–926

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Towards semantic maps for mobile robots

Andreas Nüchter*, Joachim Hertzberg

University of Osnabrück, Institute of Computer Science, Knowledge-Based Systems Research Group, Albrechtstr. 28, D-49069 Osnabrück, Germany

ARTICLE INFO

Article history:
Available online 15 August 2008

Keywords:
3D mapping
6D SLAM
Scene interpretation
Object detection
Semantic mapping

ABSTRACT

Intelligent autonomous action in ordinary environments calls for maps. 3D geometry is generally required for avoiding collision with complex obstacles and to self-localize in six degrees of freedom (6 DoF) (x, y, z positions, roll, yaw, and pitch angles). Meaning, in addition to geometry, becomes inevitable if the robot is supposed to interact with its environment in a goal-directed way. A semantic stance enables the robot to reason about objects; it helps disambiguate or round off sensor data; and the robot knowledge becomes reviewable and communicable.

The paper describes an approach and an integrated robot system for semantic mapping. The prime sensor is a 3D laser scanner. Individual scans are registered into a coherent 3D geometry map by 6D SLAM. Coarse scene features (e.g., walls, floors in a building) are determined by semantic labeling. More delicate objects are then detected by a trained classifier and localized. In the end, the semantic maps can be visualized for human inspection. We sketch the overall architecture of the approach, explain the respective steps and their underlying algorithms, give examples based on a working robot implementation, and discuss the findings.

1. Introduction

1.1. What is a semantic map?

If it is agreed that semantic knowledge can help an autonomous robot act goal-directedly, then, consequently, part of this knowledge has to be about objects, functionalities, events, or relations in the robot's environment. The data structure holding the space-related information about this environment is the *map*. Typical state-of-the-art robot maps represent the environment geometry – often in 2D, sometimes in 3D, sometimes topologically – and, maybe, additional sensor-relevant information such as specific features, or texture [5]. This typical map content is in harmony with today's typical purpose of maps for mobile robots, namely, navigation. A *semantic map* augments that by information about entities, i.e., objects, functionalities, or events, that are located in space.

We assume that the main purpose, or family of purposes, for a semantic stance in map contents is some type of reasoning based on individual entities in the map and/or their classes; examples for such reasoning are planning, explanation, prediction, and sensor data interpretation. To enable this reasoning, some background knowledge about entities is required, an informal example being a rule like *A chair typically rests on the floor*. The knowledge may come in any suitable knowledge representation format [4], as needed for the type or types of reasoning to be associated with the entities in the map. Given that such knowledge is typically independent of space, it is not strictly part of the map; however, we require that it exists for entities represented in the semantic map. In brief, then:

> A *semantic map* for a mobile robot is a map that contains, in addition to spatial information about the environment, assignments of mapped features to entities of known classes. Further knowledge about these entities, independent of the map contents, is available for reasoning in some knowledge base with an associated reasoning engine.

In the technical part of this paper, we will use special instances of sensor configuration, map type, and reasoning. Note, however, that we understand semantic maps as being a more general concept than what we have experimented with and that we will describe below. We will get back to this in the discussion part of this paper.

1.2. System and paper architecture

Our approach uses 3D laser range and reflectance data for environment mapping and for perceiving 3D objects on an autonomous mobile robot. Starting from an empty map, multiple 3D scans, acquired by the robot in a stop-scan-go fashion, are registered consistently by 6D SLAM, i.e., by a version of *Simultaneous Localization and Mapping* that allows for using 6DoF robot poses (x, y, z positions; yaw, pitch and roll angles). Then, the coarse structure of the resulting 3D scene is interpreted using plane

* Corresponding author. Tel.: +49 541 969 2623.
E-mail address: andreas.nuechter@uos.de (A. Nüchter).

0921-8890/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2008.08.001

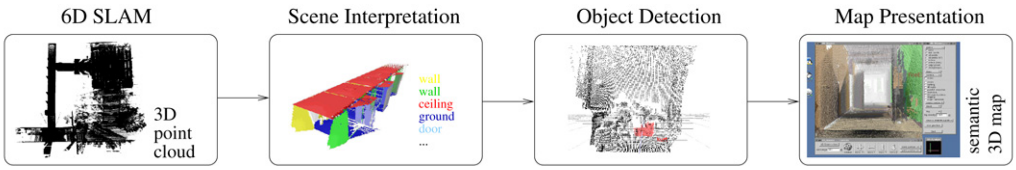

Fig. 1. System overview: From left to right: 6D SLAM acquires and registers a point cloud consisting of multiple 3D scans; scene interpretation labels the basic elements in the scene; object detection locates previously learned objects in 6 DoF; finally, the semantic map is presented to the user.

extraction and labeling, thereby exploiting background knowledge represented in a constraint network [34]. After that, the 3D range and reflectance data are transformed into 2D images by off-screen rendering, and they are used in this form for detecting and localizing objects by two alternative approaches [35,42]; the object localization is then transformed back into the 3D data. Finally, the semantic map is presented using tools from computer graphics. Fig. 1 gives a system overview. Our system contains the classical architecture to derive symbols from sensor data [22]. While these building blocks in isolation have been described in previous publications, we present in this paper for the first time how these components have been put together to build semantic maps according to our definition. Note that the simple cascade-style architecture just described is only an initial point for building semantic maps: In general, one would use feed-back, e.g., from object detection to scene interpretation. The matter will be discussed below.

The paper presents our work in terms of the building blocks in Fig. 1. We emphasize scene interpretation and object detection, and the feed-back control loop beyond the simple cascade in Fig. 1. So after finishing this introduction by remarks on related work, Section 2 summarizes some technical background concerning the 6D SLAM algorithm used. Section 3 describes the process of bottom-up interpretation of gross scene features. These are augmented by recognized objects, as presented in Section 4. Section 5 wraps up the process and provides example results. Section 6 discusses our own results and puts them into perspective with semantic mapping in general. We conclude the paper in Section 7.

1.3. Related work

Robotic mapping. In the considerable body of literature about robot maps and mapping, maps are metrical in most cases, and less frequently, topological. As in regular language, a map contains space-related information about the environment, i.e., not all that a robot may know or learn about its world need go into the map. Metric maps are supposed to represent the environment geometry quantitatively correctly, up to discretization errors. We will use the term *geometry map* henceforth to refer to maps that represent (metrically more or less truthfully) the environment geometry. [41, Ch. 5] gives a general introduction into the topics of maps and mapping; [46] covers probabilistic approaches in particular. Both textbooks also give introductions to SLAM, i.e., the process of building a map based on imprecise sensor data and on the imprecise robot motion model.

Most robot maps in the literature are given in 2D, usually upright projections of the scene. Since the early 2000s, some groups have been using pitching or rotating laser scanners for acquiring 3D data, e.g.,[43,48,50]. As these data are much richer in information than the 2D scans mostly used in 2D mapping, slightly different algorithms are used for 3D. Based on consistent 3D scans of the environment, scan matching variants are often applied for constructing a 3D map [13,18,26,30,36,40,44]. [33] summarizes the state of the art in 3D mapping.

Only few groups in robotics have been working on variants of semantic mapping. [21] presents a robot control architecture that fuses mapping and object detection, resulting in a labeled map. [20] presents a mapping system that reconstructs 3D models assuming 3 DoF, i.e., planar, robot motion in RoboCup Rescue. In the same context, [9] uses labeled maps for automatic behavior activation. [1] and [3] also present mapping approaches that include object detection. They repeatedly map environments and identify changing occupancy of grid cells using difference maps, focusing on representing uncertain object knowledge in such occupancy object maps. [25] describes a probabilistic approach for inserting in the map hierarchical environment structures and spatial relations, all based on 2D data. [12] is a study, also based on 2D data, to combine metric, topological, and semantic aspects in a map. It uses the semantic level for reasoning ("This room contains no sink, it cannot be the kitchen!").

Scene understanding. To understand *understanding* has been a topic in AI from its early days on. The problem could be described as [39, pg. 791]

> We are given a set of ambiguous inputs, and from them we have to work backwards to decide what state of the world could have created these inputs.

Prominent lines of research in AI include language/speech understanding, image understanding, and scene understanding — all in the sense just quoted. More recent AI research mostly avoids the term due to its generality, imprecision, and metaphorical overloadedness. Yet, it describes nicely what is needed for building semantic maps. Recent work in computer vision uses the term *Cognitive Vision*, cf. [7]. We will come back to the approach by Neumann and Möller [29]: They use a description logic domain theory and a representation of perceived environment objects and processes for aggregating bottom-up scene information from camera images and for hypothesizing top-down features to look for in the given image stream. That is clearly an important ingredient of building semantic maps. A point is typically lacking in scene understanding work that semantic mapping in closed-loop robot control should include: Physical robot action in sensor data acquisition, as by changing the pose or even physical interaction with the environment.

Symbol grounding. Object anchoring [8] is a line of robotics-related research that aims at building up and maintaining the links between symbolic representations of objects (as in a logic-based knowledge representation formalism) and their images in the sensor data stream. This is clearly related to semantic mapping; it is also more ambitious than the latter, as anchoring assumes projecting the development of anchored objects into the time-space future, which semantic mapping, as considered here, does not necessarily involve. On the same line, a semantic map is related to solving the symbol grounding problem [16]. Note, however, that semantic mapping deals only with a small fraction of symbol grounding in general.

2. Technical prolegomena: 6D SLAM

For building a semantic 3D map, we start with building some version of geometric 3D map of the environment first. In the cascade architecture of Fig. 1, this prior map even needs to have