HALMSTAD
UNIVERSITY

Semantic Mapping in Warehouses

Saeed Gholami Shahbandi

Supervisors: Björn Åstrand, PhD
Roland Philippsen, PhD
Antanas Verikas, Professor

LICENTIATE THESIS | Halmstad University Dissertations no. 23

Semantic Mapping in Warehouses

Halmstad University Dissertations no. 23
ISBN: 978-91-87045-49-3 (printed)
ISBN: 978-91-87045-48-6 (pdf)
Publisher: Halmstad University Press, 2016 | www.hh.se/hup
Printer: Media-Tryck, Lund

Abstract

This thesis and appended papers present the process of tacking the problem of environment modeling for autonomous agent. More specifically, the focus of the work has been semantic mapping of warehouses. A semantic map for such purpose is expected to be layout-like and support semantics of both open spaces and infrastructure of the environment. The representation of the semantic map is required to be understandable by all involved agents (humans, AGVs and WMS.) And the process of semantic mapping is desired to lean toward full-autonomy, with minimum input requirement from human user. To that end, we studied the problem of semantic annotation over two kinds of spatial map from different modalities. We identified properties, structure, and challenges of the problem. And we have developed representations and accompanied methods, while meeting the set criteria. The overall objective of the work is "to develop and construct a layer of abstraction (models and/or decomposition) for structuring and facilitate access to salient information in the sensory data. This layer of abstraction connects high level concepts to low-level sensory pattern." Relying on modeling and decomposition of sensory data, we present our work on abstract representation for two modalities (laser scanner and camera) in three appended papers. Feasibility and the performance of the proposed methods are evaluated over data from real warehouse. The thesis conclude with summarizing the presented technical details, and drawing the outline for future work.

To Zari and Ali.

List of Publications

The thesis summarizes the following papers:

A. Gholami Shahbandi, Saeed, and Björn Åstrand. *"Modeling of a Large Structured Environment: With a Repetitive Canonical Geometric-Semantic Model."* 15th Annual Conference, Towards Autonomous Robotic Systems (TAROS) 2014, Birmingham, United Kingdom, September 1-3, 2014. Springer, 2014.

B. Gholami Shahbandi, Saeed, Björn Åstrand, and Roland Philippsen. *"Sensor based adaptive metric-topological cell decomposition method for semantic annotation of structured environments."* Control Automation Robotics and Vision (ICARCV), 2014 13th International Conference on. IEEE, 2014.

C. Gholami Shahbandi, Saeed, Björn Åstrand, and Roland Philippsen. *"Semi-Supervised Semantic Labeling of Adaptive Cell Decomposition Maps in Well-Structured Environments."* European Conference on Mobile Robots (ECMR) 2015, Lincoln, United Kingdom, 2-4 September, 2015. 2015.

Contents

List of Figures

Chapter 1
Automation and Intelligent Warehouses

"I could not tell you in any detail how my computer works.
I use it with a layer of automation."
– Conrad Wolfram

Historically "automation" has been designed to delegate actions to machine. Robotics (and automation) as a multidisciplinary field benefits immensely from advances in research fields such as computer science, statistical learning, signal processing, and many other contributing fields. Consequently the state of the art in autonomous robotics has many core technologies to offer that are readily available to use. Such technologies empower innovative solutions that leverage insights from several specialist domains. As a result it becomes possible to conceptualize solutions that fulfill the objectives of automation, and furthermore, the resulting solution would behave more "intelligently".

An action often follows a decision, and to make a decision is not an easy task. This is where the concepts of "intelligence" comes into play. There are different approaches in defining and identifying the nature of intelligence. Intelligence and its related researches are not the concern of this thesis. The importance of intelligence in automation is due to an "intelligent behavior" expectation of the agents[1]. Naturally in this thesis we adhere to the "behavioristic" definition of intelligence. Unlike cognitive approaches, a behaviorist approach to intelligence does not concern with the internal state of the agent. But rather identify intelligent agents through their behavior. Such

[1] An agent is any entity in the environment that is capable of perception and action. An agent could be a human operator, a robot, an Auto-Guided Vehicle (AGV), or a centralized management system.

approach to intelligence has been helpful in formulating research questions and objectives in this chapter. Most behaviorist artificial intelligent methods rely on symbol manipulation (or similar approaches) for inference and decision making. But an inference engine is not the only requirement for an agent to make a decision. The other side of this challenge is the knowledge which the inference and consequently the decision are based on. An agent's knowledge is assessed relying on the concepts of "awareness" and "knowledge hierarchy". That is to say, what is the required level of knowledge for a system to be "aware" and make an informed decision. To this end, a general description for these two concepts is presented in chapter 2.

We characterize an intelligent *behavior* by a set of desired features[2] such as, but not limited to: *i*) *Adaptability*; *ii*) *Extensibility*; and *iii*) *Inference capability*. Adaptability requires a flexible system that can undertake adaptation responsibility when circumstances demand. If an autonomous system is not adaptive, it would loose its repeat-ability by over the course of environmental changes. That is to say the more complex the objective is, it is more likely for an ad hoc solutions to fail. Extensibility aims at delivering a generic solution, independent of the problem's setting. Extensibility guarantees that a solution is capable of handling a category of challenges, instead of only one problem. In other words, an extensible system is generic for a class of problems. So an implemented solution for an application could be readily used in similar applications, but with different settings. Inference capability enables the system to deduce reasonable and practical conclusions from available knowledge and observations, based on which the system can make a decision. The inference is performed in perception to improve an agent's understanding of the sensory data and enhance its knowledge base. The inference is more prominently practiced in action domain for decision making.

1.1 Warehouse Automation

There is a disparity between a researcher's assumptions and the public opinion. Researchers tend to take some assumptions for granted. In case of automation for instance, even the most basic advantages of automation such as safety and delegation of mundane tasks to robots might be under question in public opinion. This contrast I believe is rooted in researchers' over familiarity with automation concept. Researchers often deal with complex systems on a daily basis. As well, research practice brings the convenience of having automation skills at one's disposal, which in turn brings an insight into automation solutions in general. Therefore, before focusing on the works behind this thesis, we take a closer look at automation and its motivations. Motivations behind automation could be grouped into three fundamental categories.

[2] Please note that the mentioned properties or features of an intelligent agent are based on a "behavioristic" assessment instead of a *cognitive* approach to intelligence.

Fundamental in the sense that the motivations behind most automation processes are another variation or formulation of these categories, or ultimately are contributing to one of them as a subset. This is regardless of whether if the subject of discussion is an ad hoc solution of an specific application (e.g. conveyor belts or spam filters) or a general purpose tool (e.g. robotic manipulators.) These three categories are as follow:

- *Delegation of undesirable and mundane tasks.*
- *Safety*, which might have different causes, such as:
  - Human error. In comparison it's easier to make sure that robots won't make the same mistake twice (if the mistake is acknowledged and debugged).
  - Dangerous environments. For instance rescue missions, space missions, nuclear reactors, etc. Robots are employed in such tasks to save humans' life[3].
- *Increase throughput*, that could be realized with different objectives such as, but not limited to: *i*) Increasing capacity by speeding up the processes, *ii*) Improving quality of the product, *iii*) Robustness of the processes to failure, *iv*) High accuracy tasks.

Warehouses and their management processes are challenges with interesting features. Features that would fit right into above-mentioned motivations. And motivations are present from all categories. Repetitive tasks (*undesirable tasks*) are the building blocks of a warehouse management flow. A warehouse's work flow could be modulated and categorized into groups such as `navigate`, `locate`, `pick-up`, `drop`, `inventory maintenance` and so on. Warehouses are crammed with lift trucks in their corridors. Filled partially with Auto Guided Vehicles (AGVs) and manually driven trucks, there is a high chance of collision (*safety*). Throughput matters, since it is considered an economic drive for the society. Not to mention that the investors would like to have higher profit (*increase throughput.*) It's safe to assume that throughput is one of the main motivations of building warehouses in the first place.

Automatic Inventory and Mapping of Stock (AIMS)

The research works behind this thesis and appended papers have been conducted in the frame of AIMS project[4]. The AIMS project approaches the intelligent warehouse problem by deploying "computational awareness"[5] in a robotic systems. AIMS project intend to improve the behavior of AGVs and

[3] This might be one of the invoking reason for robots uprising in future.

[4] Link: `http://islab.hh.se/mediawiki/AIMS`

[5] Zhao [Zhao et al., 2012] defines awareness as "the ability [...] to be conscious of sensory patterns". See section 2.1 for more details.

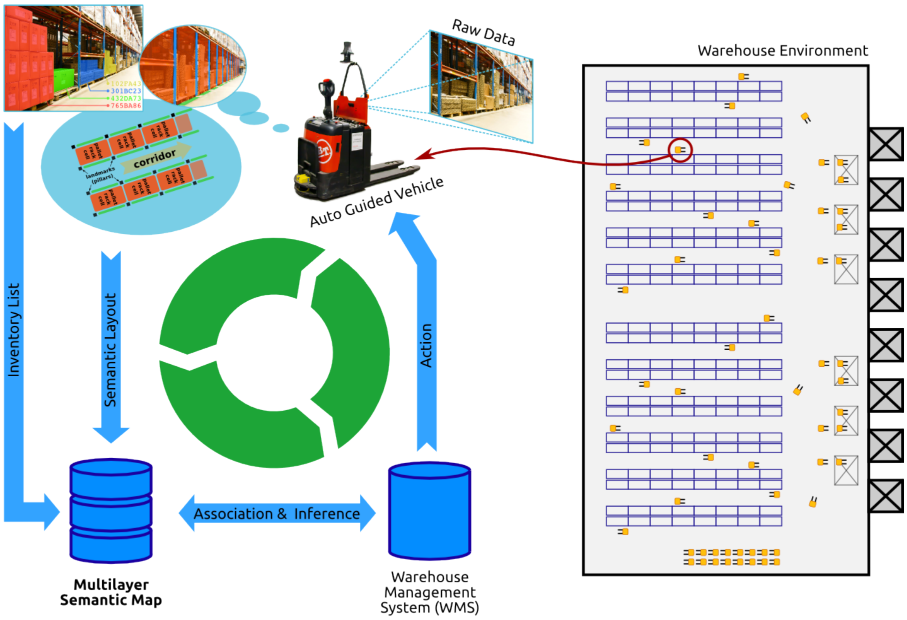


Figure 1.1: Automatic Inventory and Mapping of Stock (AIMS) project. The ultimate objective of the AIMS project is autonomous development of a multi-layer semantic map entailing a representation of the environment layout, inventory list, and the link to the Warehouse Management System (WMS).

autonomous systems through elevating their knowledge of their surrounding. Objectives of the AIMS project could be enumerated as: *i*) semantic mapping of the layout (and the focus of this thesis); *ii*) inventory management, that is the problem of object identification and volume estimation to maintain and track the inventory list over the time; and *iii*) obstacle avoidance, is the problem of detecting static and dynamic objects in the trucks' trajectory. Figure 1.1 illustrates the overall of the aims project. The ultimate objective of the AIMS project is autonomous development of a multi-layer semantic map, entailing a representation of the environment layout, inventory list, and trajectory of the moving objects. This semantic map would linked to the Warehouse Management System (WMS), for information and knowledge sharing between the two.

1.2 Semantic Mapping

Broadly speaking "semantics theory" is the study of the words' meaning and origin in natural languages. This means the semantic labels have a subtle and profound meaning based on some context. The terms (names, nouns,...) and their meanings are developed in natural languages throughout the history of human kind. That is why the terms and their meanings are contextual. They depend on human's sensory configurations, social context and so on. Therefore a spontaneous development of semantics in isolation from human knowledge seems not feasible. On the other hand, development of an artificial systems, able to perceive the surrounding on the human level is still intractable [6]. The alternative is to provide the system with semantics through the labels for categories of objects. That is to say, identifying categories for the system along with attributes for each category. So that the system is equipped with required knowledge to behave accordingly and intelligently. Such approach only concerns the knowledge base of an intelligent decision, not the inference ability.

In robotics a map is the robot's internal abstract and spatial representation of the world. The key criterion for a map to be qualified as "semantic", is to support semantic annotation of its content. The content refers to those objects [7] that are located and presented in the map. The instantiating of the map into objects requires background knowledge and depends on the focus of interest [8]. Consequently the implementations of semantic mapping are application driven. That is because there is not available a universal representation (e.g. feature space) capable of compromising all modalities, context, patterns and information in general. For instance different types of sensors or objects of interest, would lead to many different approaches for pattern recognition. If the scope of application is known in advance, the robot could be equipped with methods (often through training) that enable the robot to detect and instantiate the objects of interest. So the robot would understand its surrounding and is able to interact with its environment as intended, as it becomes "aware".

1.3 Objectives, Research Questions and Contribution

The objective is to develop a framework for autonomous handling of spatial information of the environment, with emphasize on minimizing the need for the human intervention. Toward an aware system for effective management of logistics and inventory in a warehouse environment, a rich geometric-semantic map is set as a goal. The technical objective is to create a multi-layer semantically annotated spatial map. Such a map that is enriched with information required for articulating intelligent behavior. We deduce part of

[6] Here, the "human level" refers to human's cognitive ability, awareness and consciousness.

[7] Here object refers to not only object such as bed and desk, but also places (e.g. room), the static physical world (e.g. wall.), and other instances.

[8] Whether if the layout of the surrounding is the matter of interest, what the environment accommodates, or else.

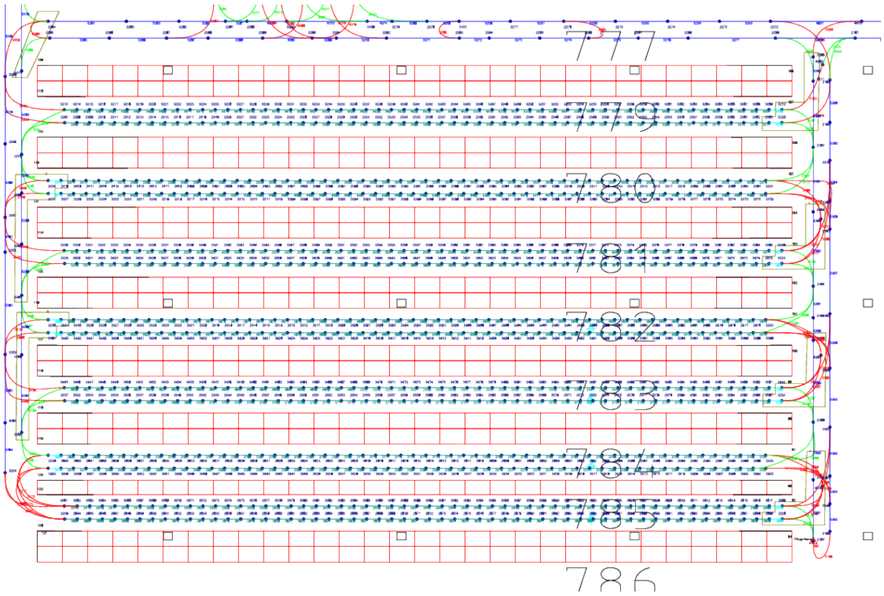


Figure 1.2: An example manually drawn CAD (Computer-Aided Design) sketch of a real-world warehouse. Red rectangles represent the infrastructure, that is to say the pallets locations in storage racks. Manually drawn "pick&drop" points connected to each other in blue, represent the pre-define path.)

this information set from an example CAD (Computer-Aided Design) drawing of a real-world warehouse, as depicted in figure 1.2.

In the works presented in the appended papers, we addresses this objective on the level of infrastructure and region modeling with four criteria.

- *Source and essence of the map;* a spatial map that is directly used for low level tasks, such as localization of the vehicles and inventory, that also builds the foundation of the final semantic map.
- *Structure of the representation;* a layout map (similar to the drawing in the figure 1.2) seems like a suitable example of how the desired map should be structured.
- *Context of the representation;* the map is expected to contain semantics of infrastructures for storage locations and of open-spaces for vehicles' trajectory.
- *Usability of the representation;* a representation must be understandable and usable by all involved agents (i.e. auto-guided trucks, human operators and the warehouse management system.)

Research Questions

Aforementioned objectives raises three fundamental research questions: *i*) what is the required knowledge for a system to be aware?; *ii*) how to acquire this knowledge?; and *iii*) what is a suitable representation for it? Awareness of a system and consequently the required knowledge for the system to be aware is objective. Nevertheless, enabling a more elaborate situation assessment depends on a high-level knowledge, suitable for inferences. To this end, chapter 2 presents two concepts of "knowledge hierarchy" and "computational awareness". Then we take a new alternative look at the knowledge hierarchy that helps the association between the levels of the two concepts. Autonomous acquisition of the high-level knowledge demands a system to handle *segmentation* of data and *associating* instances with categories from the conceptual level. Considering the "unconstrained and infinitely complex" [Freeman, 2001] real world, one of the most crucial steps in semantic analysis is the segmentation capability. It is the process of identifying meaningful instances (e.g. place, object, activity) in the data. In addition to segmentation, acquiring high-level knowledge requires an association between the sensory patterns and the categories of instances. Semantic annotation (and anchoring problem) is about bridging these gaps and extracting knowledge. Chapter 2 also covers the basics of semantic theories, with an emphasize on its implications in semantic mapping problem. However, an autonomous system is not capable of the association without being introduced to semantics. That is to say, at some point in system design some prior knowledge is incorporated in the method. Furthermore, the design of a system is often context-dependent. A brief summary of works related to this topic is presented in chapter 3. This summary evaluates related work with respect to the scope of the context (from local to global) and the level of semantic incorporation (from low to high.) Our intention is to develop high level of autonomy in our semantic analysis, by pushing the level of semantic incorporation higher. It means the desired system would not require training sets with manual labeling. The system is expected to acquire sufficient information and autonomously develop an interpretation of the data, that facilitates the bridging between sensory patterns and semantics. Consequently we can formulate the general objective of the work by considering the research questions as:

To develop and construct a layer of abstraction (models and/or decomposition) for structuring and facilitate access to salient information in the sensory data. This layer of abstraction connects high level concepts to low-level sensory pattern.

Decomposition and modeling are the process of dissecting the data into smaller parts (analytically or data-driven.) Decomposition and modeling could lead to instantiating the data, similar to the segmentation process. But unlike

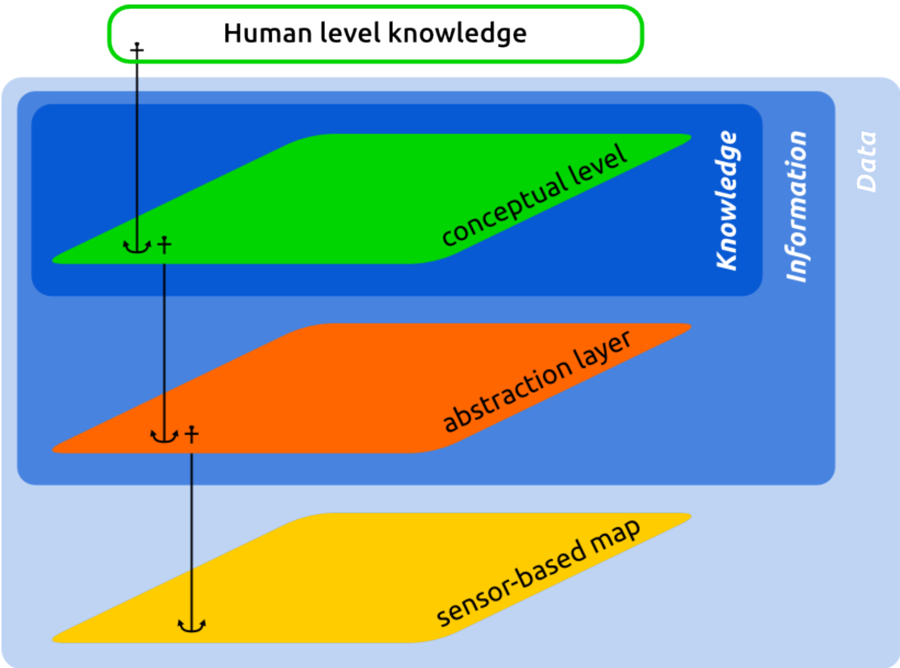


Figure 1.3: A conceptual visualization of the contribution and the core message of this thesis and appended papers.

segmentation, decomposition and modeling maintain the topological relation between instances (e.g neighborhood).

Contributions

The "thesis" of this document is the "layer of abstraction for structuring and facilitate access to salient information in the sensory data." Figure 1.3 visualizes a conceptual summary of this thesis.

General contributions of the presented work fall into three groups: *i*) studying the problem semantic mapping in well-structured environments and learning how to exploit the structure of the environment for abstraction; *ii*) as the framework of the thesis, a layer of abstraction for structuring and facilitate access to salient information in the sensory data; and *iii*) and technically, developing and implementing different representations and techniques for abstract modeling of a warehouse environment in different modalities.

Technical contributions of the appended papers (A, B, and C) could be listed as:

A)

- A canonical geometric-semantic model for the abstraction of structured environments. The model carries not only the structure, but also the semantic labels of environment's infrastructure.
- A method for generating and matching such models into the latent structure of the map. Different models are uniquely identified by the means of the model's parameters. Most of the steps and parameters of the model are independent, therefore it is adjustable according to different scenarios.
- A parametric formulation of Radon transformation integrated with a clustering technique, that detects alignment of the landmarks in the map and clusters them simultaneously.

B)

- The proposed method interprets occupancy maps to bring out underlying spatial characteristics in a format readily usable by machines and humans. The spatial arrangement of features is extracted via polygonal cell decomposition, and captured in a subdivision data structure to enable a higher-level topological analysis. After cell decomposition, two corresponding data structures capture topological characteristics and ease subsequent semantic labeling. Our novel representation characterizes the geometric structure of the environment in addition to providing an abstract metric representation of the map.
- A particular challenge that our method in Paper B addresses is the discrepancy between the importance of orientation information and the absence of corresponding clean long lines in the range data that can be obtained via laser scanners in warehouses. We employ radiography to fit independently oriented and spaced lines to the occupancy map.
- The key features of the proposed method are the flexible and robust direction detection and the adaptive spacing of lines. This focuses on an adaptive generic representation to support computing and storing of such information. Control of the spatial resolution influences the abstraction level of the resulting topological information.

C) The core of the last paper is the autonomous generation of a high-level representation that allows an efficient abstraction of the semantics based on human input. The method autonomously builds a high level spatial model of the world (based on the decomposition from paper B) and instantiates it, without prior knowledge of the environment. This allows an abstraction of human semantics, instead of providing this knowledge through a training set. Delivery of human semantics to the robot is done on a higher level, and directly supports semantic inference.

Structure of this thesis

Comparison of the approaches and methods in the field has been carried out by the help of concepts from "computational awareness" and "knowledge hierarchy". Brief descriptions of these concepts, along with an overview of *semantic mapping problem* is provided in chapter 2. Relying on the concepts that are portrayed in chapter 2, related works to this thesis are studied in chapter 3. The related works are organized in two categories of *emerging solutions* for intelligent warehouses and *semantic mapping* in robotics with more focus on the latter. Chapter 4 presents the methodology of presented works in the appended papers as a whole. Then each individual paper is described and reviewed with respect to the related works. In chapter 5 the overall of the presented work in this thesis has been discussed. It emphasizes the most outstanding features of the proposed methods, while discussing the limitations. The chapter concludes with the prospective of this work.

Chapter 2
Problem Statement

"Not to mean one thing alone is to mean nothing."
– Aristotle

The first half of this chapter intends to find a reference frame that enables analysis of semantic mapping approaches and their differences in terms of obtained knowledge. For this purpose, "Computational awareness" and "knowledge hierarchy" are introduced in this chapter. Awareness has been discussed in different fields of research with numerous definitions. The study of the awareness is often aimed at the question of "how to build aware systems?". Looking at it backward, awareness concept provides an implicit analysis frame for the assessment of an agent's awareness. With this frame it becomes possible to discuss the required level of "knowledge" for an agent to make an informed decision. Zhao summarizes the concept of awareness out of the application context under the title of "computational awareness" [Zhao et al., 2012, Zhao, 2013]. In section 2.1 we take a closer look at the computational awareness, and different levels of awareness are discussed in section 2.1.1. One type of awareness, namely "situation awareness", is presented section 2.1.2 as an example. As mentioned, most of the works in awareness researches contributed to building aware systems, and the assessment of the awareness levels is not well formulated. To that end, the concept of "knowledge hierarchy" proposed by Ackoff [Ackoff, 1989] is introduce in section 2.2. Knowledge hierarchy, often presented on a pyramid, has diverse applications and consequently very little consensus over the definitions in the respective literature. Section 2.2 discusses the limitations of the pyramid representation. Then the concept of hierarchical layers of knowledge are adapted to hierarchical overlapping properties. Based on the definitions that will follow in this chapter, levels of knowledge hierarchy are associated with different levels of awareness from section 2.1. This link between levels of knowledge and awareness make it possible to find a lower bound on the knowledge requirement of an aware system. However, the quantification of such link and the consequent requirements are objective dependent. The alternative look on the knowledge hierarchy, specifically the correlation between the levels of awareness and the knowledge hierarchy is a new perspective to these concepts. This correlation is addressed section 2.3.

The rest of this chapter aims at the problem of semantic mapping. Please note that the related works to semantic mapping in robotics are presented in chapter 3. The second part of this chapter is an attempt to characterize the *problem* of semantic mapping. For this purpose, a linguistic definition of semantics and semantic theories is followed by a very brief overview of semantic mapping. This chapter concludes by reviewing the role of prior knowledge in semantic mapping problem.

2.1 Computational awareness

Zhao [Zhao, 2013] studies the problem of awareness out of the application context under the title of "computational awareness". The concept of computational awareness presented here, is based on his works [Zhao et al., 2012, Zhao, 2013]. Zhao [Zhao et al., 2012] defines awareness as "the ability [...] to be conscious of sensory patterns". Based on the implied ability of the aware system to draw inferences from experiences, he considers awareness as "the bridge between perception and cognition", and thus a requirement for intelligence [Zhao, 2013]. That is to say awareness is the understanding of the context required for an intelligent behavior. Even though on its own, awareness does not ensure intelligence, it does help to understand the context. As a consequence, Zhao suggests that tackling the computational awareness problem is prior to developing artificial intelligence. In the context of automation, an aware system must be able to acquire data autonomously, process accordingly and develop a knowledge base that could be used for reasoning. On the other hand, Zhao considers the main purpose of computational awareness to provide context information. So that the recipient of the information (the agent) "can make decisions proactively" [Zhao, 2013]. This means the ability to communicate is a required feature of such systems. Communicability in addition to autonomous knowledge acquisition, demands the system to obtain semantics of the world's instances analogous to human knowledge.

2.1.1 levels of awareness

Awareness is the result of a multi-level process. It starts from perceiving the world through sensory input, and proceeds towards understanding the sensory patterns, so an informed decision could be made by an agent. According to Zhao [Zhao, 2013], different levels of awareness begin from low level perception

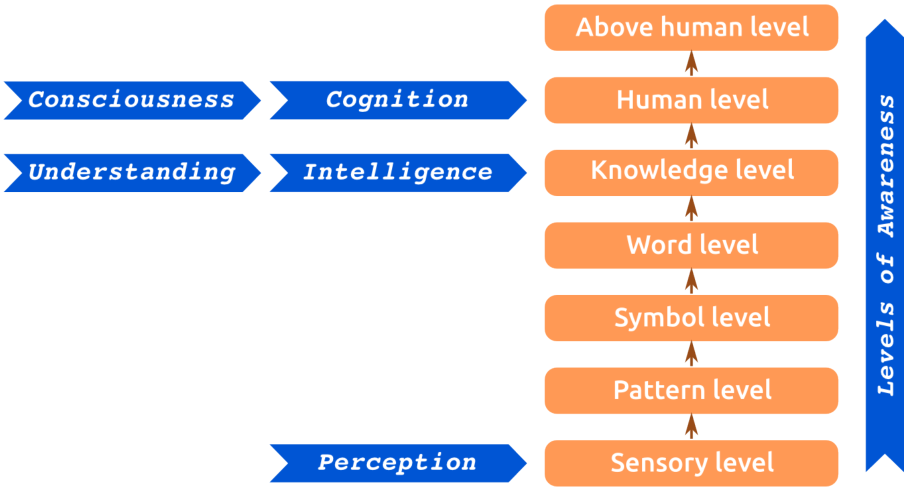


Figure 2.1: Levels of awareness proposed by Zhao [Zhao, 2013]. The figure presented here is augmented with extra detail to demonstrate the levels of awareness where “intelligence”, “understanding”, “cognition”, and “consciousness” could be achieved, according to Zhao [Zhao, 2013].

with no understanding, and advance to higher levels of awareness beyond individual humans (see figure 2.1). While a low level awareness may not directly understand the context, they do provide appropriate information for higher levels. Zhao anticipates a “step-by-step” creation of intelligence, on the higher levels of awareness [Zhao, 2013]. He asserts that if a systems is “just programmed” it might become aware to some level, but won’t be intelligent.

2.1.2 situation awareness

Situation aware system often refers to a system that is able to detect *events* in a timely manner. That is to say, its concern is to estimate (or predict) an event based on the history and current situation, in order to avoid a risk or gain from an upcoming opportunity. Such system could have many applications, from smart home and color print quality assessment [Lundström, 2014], to more general automation tasks [Endsley, 2011]. Situation awareness in particular is not the topic of interest for this thesis. The computational model of situation awareness proposed by Endsley [Endsley, 1995] is more general than just situation awareness. It also provides a framework for the prospective of our research (more in section5.1.) The model splits the process into three stages: *i*) *perception* of the environment; *ii*) *comprehension* of the situation; and *iii*) *projection* of future states. These three stages suggests that situation awareness as a framework is the counterpart of robotics paradigm

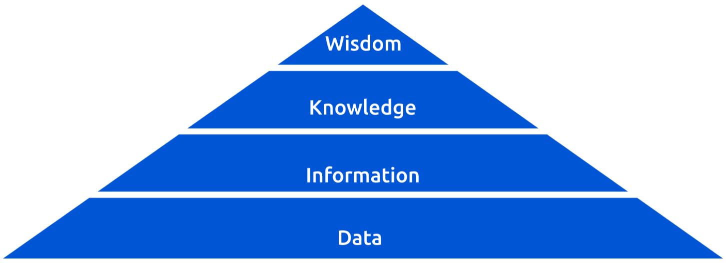


Figure 2.2: The knowledge hierarchy, with *four* levels [Rowley, 2007].

"Sense-Plan-Act". The slight difference between the two is in their objectives. Situation awareness focuses on identification and prediction of the events in the temporal field. On the other hand, the "Sense-Plan-Act" paradigm could be considered as the situation awareness with a focus on the spatial filed.

2.2 Knowledge hierarchy

The level of an agent's awareness is closely related to the level of knowledge it posses, and its understanding of that knowledge. The Knowledge hierarchy, originally articulated by Ackoff [Ackoff, 1989], provides a suitable representation for assessing these levels of understanding. A four-level pyramid (illustrated in figure. 2.2) has become the most common representation of the knowledge hierarchy.

Rowley [Rowley, 2007] provides a comprehensive reviews of the literature and the view points on the concept of the knowledge hierarchy. She writes:

"The hierarchy is used to contextualize data, information, knowledge, and sometimes wisdom, with respect to one another ... and to identify and describe the processes involved in the transformation of an entity at a lower level in the hierarchy to an entity at a higher level in the hierarchy. ... Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge"

"Transformation processes" are an important aspect of the hierarchy. Even though they are not illustrated explicitly in the pyramid, the transformation processes are central to the concept. Their role is to *elevate* an agent's understanding, by transforming data, information and knowledge to higher levels. Definition of the information and knowledge are often based on transformation processes, such as formatting and organizing.

2.2.1 definitions

Rowley emphasizes that there is not a consensus on the concept and definitions. Despite that, she manages to provide a helpful summary of definitions in [Rowley, 2007]. The data is recorded description of things, objective facts or observations which are: *i*) elementary and discrete; *ii*) unorganized and unprocessed; and *iii*) have no meaning or value. The definition of information is often based on data and how it is transformed. Information is data that is: *i*) formatted; *ii*) processed for a purpose; *iii*) a representation of reality; *iv*) has value to the recipient; or *v*) helps the understanding of a subject. Ackoff [Ackoff, 1989] asserts that "the difference between data and information is functional..." [Rowley, 2007]. Curtis and Cobham [Curtis and Cobham, 2008] and Bocij et al. [Bocij et al., 2008] also associate the definition of information to the transformation processes that make data functional and usable. When it comes to knowledge, the only consensus on its nature is that "it is based on perception that can provide a rational justification for it" [Rowley, 2007]. Knowledge builds on data and/or information that have been organized and processed. There is a notion of "added ingredient" in knowledge definitions which refers to expert opinion, skill, and experience. The objective of knowledge is to aid decision making by understanding experience, accumulated learning, and expertise. The concept of wisdom is very vague. Due to its unclear definition, it does not help modeling the levels of awareness, and therefore it is skipped here. To summarize, we expressed data, information and knowledge as:

Data is the result of *sensing* the reality through recording sensor(s).

Information is a recognizable pattern, *usable* by an agent to interpret the data or carry out an action.

Knowledge is what enables the agent to make an informed decision toward an *objective*.

Note that the key difference between information and data is the functionality aspect of information. And in turn, knowledge and information are differentiated in the decision making aspect. Objective is the a latent feature of the knowledge, which could be considered as a part of decision making process.

2.2.2 limitations of the pyramid

Although the hierarchy and its pyramid has been respected and used in many disciplines, it is mostly from a management, business and information systems perspective. The broad use and origin of the concept justify the abstract and generic representation of the pyramid. However, abstraction and generality come with a cost; limited expressiveness of the representation. It lacks the ability to represent technical details of an engineering system. These limitations could be identified with the following questions.

1. Is there a sharp division between the levels of the hierarchy?
2. What are the qualia for expressing the value of information?
3. How does the pyramid represent the transformation processes?
4. How does the pyramid incorporate the prior knowledge?

The mentioned limitations concern the pyramid representation, other than the knowledge hierarchy itself. These restrictions of the pyramid could motivate an alternative, and more expressive representation of the knowledge hierarchy. However, such proposal would be out of the scope of this thesis.

2.3 Knowledge hierarchy and the levels of awareness

In his paper [Ackoff, 1989], Ackoff originally articulated the concept of knowledge hierarchy, explaining that those higher levels of the hierarchy "include the categories that fall below them". He claims, for instance, there could be no knowledge without information. However there is not a clear explanation of what this "inclusion" means in terms of data, information and knowledge. The pyramid implies that each level is based on the lower levels, not including them. In order to resolve the conflict, this section presents an alternative interpretation of the hierarchy's levels. This new interpretation also facilitates establishing a correlation between levels of awareness and knowledge hierarchy. Assuming that the levels of the hierarchy (data, information, and knowledge) are different *properties*, and two properties could overlap, a set of records could have the one property or more at the same time (e.g. data, or both data and information). For instance, consider a set of range scanner readings from a robotics exploration mission. A map is the compilation of readings, associated with each other with the control inputs and/or time stamps. The set of readings, in its original form only holds the data property. Since the map is used for self-localization (i.e. it is usable and functional), it holds both the data and information property. Let's assume the robot explores the environment autonomously (e.g. frontier exploration.) This indicates the use of additional semantics of the map (i.e. frontier.) This "added ingredient" that could be used for decision making (i.e. where to explore next?) would add the extra property of knowledge to the map. The same map would be data and information if it is only used for localization. It would be data, information, *and also* knowledge, if the concept of "frontier" is used for exploration.

Figure 2.3 shows how these properties relate to the levels of awareness. The notion of property might suggest that any set of recordings that is information

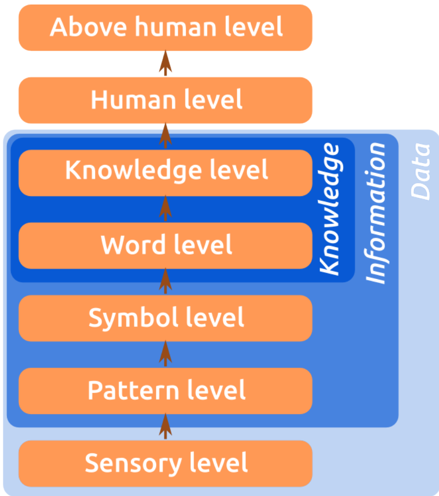


Figure 2.3: Overlapping properties of the knowledge hierarchy versus the levels of awareness.

must also be data. However, there are exceptions. For instance, prior knowledge in the abstract form presented to the aware systems only holds the knowledge property. Their transformation from data to knowledge has already happened outside the system, in people's mind. Introducing the aware systems to our prior knowledge only involves the association of available set of recording with those prior knowledge in an *abstract* form.

2.4 Semantics

Gärdenfors in [Gärdenfors, 2004] mentions four criteria for a theory of semantics for which to account. Later he added 2 additional criteria in [Gärdenfors, 2014].

1. ontological criterion;
2. semantic criterion;
3. epistemological criterion;
4. social criterion;
5. the relation between perceptual processes and meaning.
6. the relation between actions and meaning.

The ontological criterion concerns the *object of meaning* and its nature. It assures that the theory is philosophically and fundamentally sound.

On the next level toward pragmatism, the semantic criterion tries to address the relation between symbols and their meanings. The semantic criterion is the fundamental challenge in anchoring problem in robotics [Coradeschi and Saffiotti, 2003, Coradeschi and Loutfi, 2008, Daoutis, 2013]. Anchoring is the problem of associating perception and symbols, which by itself is a form of "symbol grounding problem". How does a child learn meaning and knows what to learn? The epistemological criterion accounts for that aspect of the semantic theory. Communication, borrowing from the Latin word *commūnis*, means "to share, or make common". Social criterion compels semantic theories to explain how the meaning in individuals and communal language are related, and it is the backbone of successful communications. Human Robot Interaction (HRI) is one of the applications that deals with the social criterion. Enable communication between agents is of particular importance for environment where robots and humans collaborate. Consequently the knowledge representation compatibility between human operators, management systems, as well as robots is desired. The two additional criteria account for the dependency between meaning and the agent's interface to the world, one way or another. That is to say, how perception shapes the meanings in an agent's mind, and in return how the meanings in the agent's mind would affect its perception. The other criterion is concerned with the same matter, but with action instead of perception.

2.4.1 semantic mapping

Fundamental problems in robotic mapping and localization are being tackled by pioneers (e.g. [Thrun et al., 2005]), and intelligent behavior of mobile robots is becoming more accessible. Semantic mapping in robotics shares its vision with similar fields, such as semantic web[1]. That is to say, adding a semantic layer to low level information to enable conceptualization of the real world data, and in turn be able to perform search and reasoning on knowledge level [Gärdenfors, 2014]. Semantic mapping has emerged long ago (e.g. [Kuipers, 2000]) in robotics, but it hasn't been long since it became fully functional on the knowledge level [2]. Semantic maps have became a reliable source in many applications, from navigation [Kostavelis and Gasteratos, 2013], and task scheduling [Kostavelis and Gasteratos, 2015] to exploration missions and object search [Aydemir et al., 2011].

[1]"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries." -quote from World Wide Web Consortium (W3C). December 11th, 2013. Retrieved April 1st, 2016.

[2] Note that semantics appear on different levels from *occupancy* in a grid map, to a *physical element* modeled by a line, to a concept such as *kitchen* represented by free space.

2.4.2 the role of human knowledge

Could an autonomous system acquire semantic knowledge without being introduced to semantics at any level? The answer to this question lies in how a semantics theory satisfies the epistemological criterion. One could take two approaches to this problem, namely machine learning and neurobiology. Neurobiology might be more promising among the two, since it attempts to understand human brain and mimic its functionality. However, it still has a long way go to achieve such a computational model[3]. On the other hand, categorization is the task of associating instances to the category to which they belong. Feature extraction, similarity measure and categorization algorithms in general might depend on each other, be context dependent, but they are not semantics dependent. The core question of semantic mapping in robotics is:

to find the appropriate association between "properties" in conceptual spaces (which could be measured and understood by the machine), and semantics that provides "meaning" to those properties analogous to what humans hold in their minds.

In other words, the problem of semantic mapping could be reduced to "semantic annotation" [4]. I share this opinion with Nüchter et. al. who consider a semantic map to contain "in addition to spatial information about the environment, assignments of mapped features to entities of known classes"[Nüchter et al., 2014]. The fact that one side of this association comes from human mind (semantics) suggests the need for human intervention at some point. This means that an autonomous system is not capable of the association without being introduced to semantics. More generally speaking, at some point in system design, our objectives must be introduced to the machine.

[3]This doesn't mean that if such a computational model is achieved, a machine can spontaneously develop semantics. Even human child has to learn semantic knowledge through training (recall the epistemological criterion of semantic theories.)

[4] Please note that I do not mean to undermine the sophisticated and elaborate methods, such as inference engines.

Chapter 3
Related Work

"If I have seen further, it is by standing on the shoulders of giants."
– Isaac Newton

In this chapter a review of the commercial solutions in intelligent warehouses is followed by the review of research works with an emphasis on semantic mapping. Commercial solutions gives an insight to the state of the art technology employed for warehouse automation. The contribution of this thesis is within the scope of semantic mapping toward awareness for auto guided vehicles in warehouses. Henceforth the main body of this chapter focuses on semantic mapping in robotics.

Commercial solutions provide methods of organizing the inventory and the resources to maximize the throughput. The challenge could be subdivided into *logistic management* and *resource automation*. Logistic process management is the first challenge that raises in the warehouse automation. Complexity of the logistics in a warehouse demands a knowledge management system, from data management to decision making support systems. It involves inventory list maintenance, and managing the storage of incoming and delivery of orders. The second part of warehouse automation addresses the logistic process itself. It is an attempt to automate the resources and the infrastructure. That is to say, to use auto guided vehicles (AGVs), trajectory planning for lift trucks, task scheduling, managing human collaboration with automated systems, and so on.

IBM's "Forward View" e-magazine mentions[1] the tendency of wholesalers in moving toward intelligent warehouses. *DEXION*[2] provides a variety

[1] Link: `http://www.ibm.com/midmarket/us/en/forwardview/`
[2] Link : `http://www.dexion.com/`

of warehouse automation solutions, from pallet racking to warehouse management systems. *Panasonic*[3] also provides a solution for surveying and managing warehouses. *Vargo*[4] offers a broad range of process automation such as conveyors, sorting, and warehouse management software. Kollmorgen's[5] Pick-n-Go® Automated Guided Vehicle, is an example of automating resources such as fork lift trucks in industrial environments. "OTTO" from Clearpath[6] is another example of automated resource. *Advantech*[7] provides wireless communication infrastructure for industrial automation. *Eosift*[8] offers a variety of products from palletizing systems, to pallet trucks and automated storage facilities. Of course many solutions attempt to deliver on both aspects of logistic management and resource automation. *Amazon Robotics* [9] aims at warehouse automation by automating the warehouse's infrastructure. In such warehouse, storage racks are mobile and moved around by autonomous robotic platforms. There are also examples of robotic solutions finding their ways to retail and convenient stores. "OSHbot" from *Lowe's innovation labs*[10] and "Tally" from *Simbe Robotics*[11] are two examples of such cases.

Roboticist have contributed immensely to automation tools such as auto guided vehicles, and other valuable technologies. Every advances in robotics have eventually (more often immediately) appeared in commercial solutions. Researchers have studied the challenges in intelligent warehouses. Few examples are, "Pan Robots" [12], or in specific topics such as mapping and localization [Vasiljevic et al., 2014, Beinschob and Reinke, 2014, Valencia et al., 2014, Nur et al., 2015, Stoyanov et al., 2013] article manipulation [Krug et al., 2016, Stoyanov et al., 2016], safety [Mosberger et al., 2014, Cardarelli et al., 2014], and navigation [Cardarelli et al., 2015]. Providing a comprehensive review of the related work in semantic mapping is not feasible here. Kostavelis and Gasteratos [Kostavelis and Gasteratos, 2015] present a thorough review of the semantic mapping methods. The aims of this chapter (3) is to deliver a categorical review of the related work, from the knowledge hierarchy and awareness's perspective. Please also note that the scope of the related work studied here has been narrowed in quantity according to two criteria.

[3] Link : `http://business.panasonic.co.uk/`
[4] Link : `http://www.vargosolutions.com/`
[5] Link : `http://www.kollmorgen.com/`
[6] Link : `http://www.clearpathrobotics.com/`
[7] Link : `http://www.advantech.eu/`
[8] Link : `http://www.eoslift.com/`
[9] Link : `http://www.amazonrobotics.com/`
[10] Link : `http://www.lowesinnovationlabs.com/`
[11] Link : `http://www.simberobotics.com/`
[12] A European project aiming at "automated logistics system supporting future factories": `http://www.pan-robots.eu/`

1. The method must aim to deliver a semantic map, where the map is a spatial representation of the environment. Many object recognition methods contribute to the "scene understanding" topic and could be readily integrated in a semantic mapping process. However they are not included in the discussion of this chapter, since the semantic mapping itself is the focus here.

2. Focusing on semantic maps for mobile robot, the other applications of semantic scene understanding such as object manipulation has been discarded.

3.1 Semantic perception

Nüchter et. al. [Nüchter et al., 2014] define *semantic perception* as "interpreting and organizing sensor information in symbolic form." Consequently they define *semantic mapping* as the process of "combining semantics with maps." They introduce three inputs for semantic perception (see figure 3.1); *i*) sensor data; *ii*) context; and *iii*) background knowledge in symbolic form. Even though both background knowledge and context count as prior knowledge, however they could be distinguished based on the type of knowledge they bring. Nüchter et. al. [Nüchter et al., 2014] do specify the difference by including the "symbolic form". But they do not provide definitions for the three above-mentioned terms. Here we provide definitions of the terms, within the scope of this thesis.

- *sensor data* is the recorded signals from sensors. Semantic mapping methods could be categorized with respect to the choice of sensors. However that would not be informative on the underlying method of the semantic mapping, and therefore irrelevant to this thesis.

- *context* is the knowledge that helps the understanding and the interpretation of the map. Context is what makes the interpretation of the signal (i.e. sensor data) possible.

- *background knowledge in symbolic form* is a set of abstract semantic labels and properties of object classes to be associated with instances in the map. This background knowledge enables the robot to infer the semantic labels of the regions in the map. A semantic mapping method might rely only on the context for inference and semantic annotation.

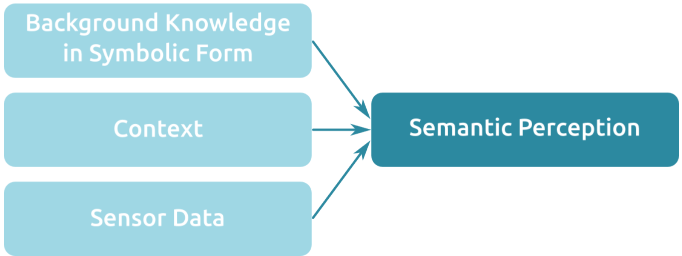


Figure 3.1: Sources of information in Semantic Perception as presented by Nüchter et. al. in [Nüchter et al., 2014].

3.2 Semantic mapping

Recalling the fifth and sixth criteria of the semantic theories from section 2.4, proposed by Gärdenfors [Gärdenfors, 2014], semantic maps could be categorized into *action-based*[13] or *perception-based*. An example of action-based semantic map is the Spatial Semantic Hierarchy (SSH) [Kuipers, 2000]. Kuiper in his seminal paper [Kuipers, 2000] includes a control-level into the semantic hierarchy. The SSH explains the relation between control-layer and the semantics of the spatial map (satisfying the sixth criterion of semantic theories according to Gärdenfors).

Since most of the works and researches on semantic mapping are perception-based, the scope of this chapter (3) is limited to that category. The rest of this section is organized by grouping the related work into different categories. Categorization of the related work relies on two measures of: *i*) levels of semantics incorporation (from low to high); and *ii*) scope of the context (from local to global). Levels of semantics incorporation is concerned with the association between semantics and the mapping processes. The semantic labels of instances could be either incorporated into the map through the mapping process, or augmented over a spatial map inferred from the stored information and the prior knowledge. The two extremes on this spectrum are "low level" (e.g. sensory) and "high level" (e.g. templates). Scope of the context is concerned with the prior knowledge, and the two extremes on this spectrum are "local" (e.g. visual cues) and "global" (e.g. region shape). Both measures are contextual and are laid over a spectrum, without sharp boundaries on either of them. The related work is analyzed by reviewing the two extremes of each spectrum. It should be noted that such categorization does not imply that a method falls only in one group. On the contrary most of the methods

[13] An action based approach to semantic mapping still requires the perception in the loop in order to build an internal representation. The difference is that in action based semantic maps the meaning is not based on appearance and consequently categorical, but the meaning is rather derived from responses to actions.

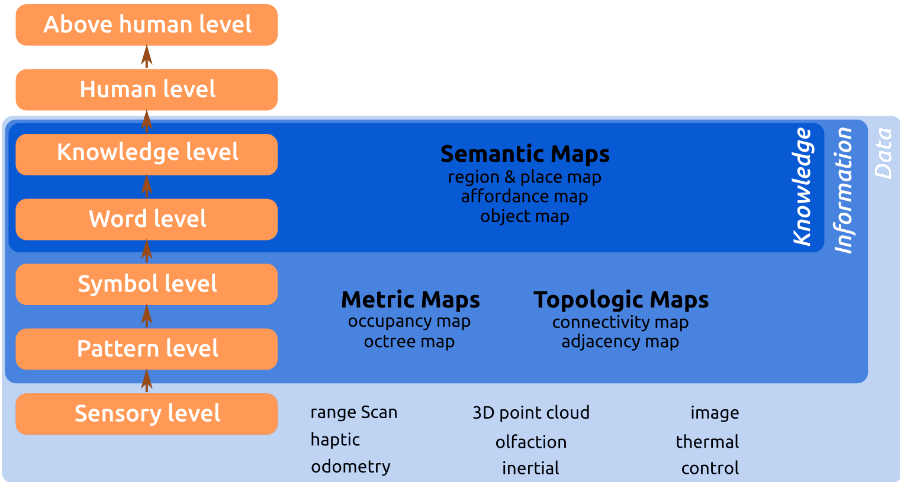
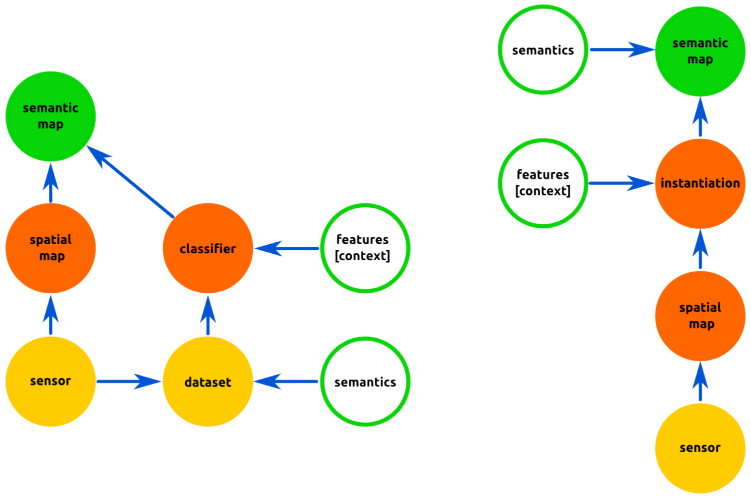


Figure 3.2: Mapping and its associated information, compared to levels of awareness and the knowledge hierarchy.

fall into more than one group. In some cases they might even use both local and global context, or low-level and high-level background knowledge delivery simultaneously. In figure 3.2 we show a separation between different types of maps in robotics and their associated information. As explained in chapter 2, even an occupancy grid map has implicit semantics that might be exploited. However in figure 3.2 we only considered the explicit information and semantics.

3.2.1 levels of semantics incorporation, from low to high

Looking at the core of semantic annotation problem in the perception-based category, it could be boiled down to the pattern recognition problem. That is to say, identifying patterns in sensory data (or the resulting map) and associating them with prior knowledge. These pattern are represented by symbols *anchored* to meanings. In this section (3.2.1) the semantic mapping methods are analyzed with respect to the pattern recognition process. More precisely the methods are laid against an abstraction spectrum, where the semantics could be incorporated on lower levels (sensory) or higher levels (template matching and beyond). Figure 3.3 illustrates samples on the two ends of this spectrum. Sub-figure 3.3a exemplifies a mapping process where the semantics are introduced on sensory level (low-level). The training data set is constructed for the later stages of pattern recognition. Practically, any method relying on a supervised pattern recognition falls in this category, let it be an object recognition method (e.g. object map), texture identification for terrain



(a) low level delivery (b) high level delivery

Figure 3.3: Different levels of incorporating the semantics into the mapping procedure.

classification (e.g. property map), or geometry based place classification. On the other hand, sub-figure 3.3b demonstrates the semantic mapping process, in which the system is capable of autonomously differentiating between instances in the sensory data. Thereafter, the semantics are represented to the system, as labels for the already discriminated instances on a higher level and in a symbolic and abstract form (such as [Gunther et al., 2013, Walter et al., 2014].)

3.2.2 scope of the context, from local to global

In the previous section, the semantic mapping methods have been analyzed with respect to the pattern recognition process. This section shifts the focus to the type and nature of the patterns and the source of those information. Regardless of the scope of the context, still the pattern recognition process is an essential part of the methodology. The two extremes of scope's spectrum are *local* and *global*. Methods relying on visual cues for pattern recognition lean more toward the local context. That is due to the dependency of visual cues on the features and the patterns of the point and its local neighborhood. Visual cues often come from cameras, but also could be based on laser scanners or 3D sensors. Object maps are one example group of such semantic maps which often rely on the *forward* and *backward chaining induction*. That is to say, the semantic label of a region is derived from the objects found in each region. The aforementioned “kitchen-fridge” example falls into this category. If a robot sees a “fridge”, relying on the fact that “fridge is located in kitchen” or “kitchen contains fridge”, infers the “kitchen” label for that area. A few examples of semantic mapping methods with this approach are [Aydemir et al., 2011, Galindo et al., 2005, Kostavelis and Gasteratos, 2013, Nüchter and Hertzberg, 2008, Zivkovic et al., 2007, Ranganathan and Dellaert, 2007, Rusu et al., 2008, Rusu, 2010, Vasudevan et al., 2007, Rottmann et al., 2005, Zender et al., 2008]. Property based semantic maps [Wolf and Sukhatme, 2008, Wurm et al., 2014, Milella et al., 2014] are a special case of object maps. While mentioned researchers aimed to derive semantic labels from the objects into the map, some others introduces properties of the regions as semantic label. Examples of property maps are “drivable” versus “grass” in the context of terrain semantics for outdoor autonomous mobile robots. The property could be the texture of the surface [Persson et al., 2007, Vatsavai et al., 2010], or even orientation of the surface [Nüchter et al., 2005]. When a method considers the shape of the region beyond the reach of a single sensor reading to infer the semantic label of the place, it is leaning toward the global context. Focusing on the configuration of the environment, the desired pattern could be described in terms of the *geometry* of region [Pronobis and Jensfelt, 2012, Kuipers, 2000, Rottmann et al., 2005, Zender et al., 2008, Joo et al., 2010, Liu and von Wichert, 2013, Mozos et al., 2005] or the *topology* of region [Zivkovic et al., 2007, Friedman et al., 2007, Fabrizi and Saffiotti, 2000, Aydemir et al., 2011, Pronobis and Jensfelt, 2012, Kuipers, 2000]. Although, it should be mentioned that the topology of the map is more often used for segmentation of places, other than semantic annotation.

3.2.3 structure of the environment

The structure refers to the arrangement (or shape) of the environment’s architecture. If ignored, the repetitive and closed-loop paths of an environment could challenge a SLAM (Simultaneous Localization And Mapping) algorithm. Such structure of the environment is another type of context. Even though the semantic of the structure is not of importance, it has been exploited for improving the performance of the mapping. The result of considering the context of environment structure in the mapping process is a more consistent global map. [de la Puente and Rodriguez-Losada, 2014] is an example of using the structure of the environment in order to improve mapping. The developed method for relies on multiple procedures for detection, evaluation, incorporation, and removal of structure constraint from latent structure of the environment into a graph-based SLAM. However, the structure doesn’t bear any semantic.

Chapter 4
Appended Papers

"Let no man enter here who is ignorant of geometry."
– Plato

Toward an aware system for effective management of logistics and inventory in a warehouse environment, a rich geometric-semantic map is set as a goal. In the works presented in the appended papers, we addresses this objective on the level of infrastructure and region modeling with four criteria.

- *Source and essence of the map;* a spatial map that is directly used for low level tasks, such as localization of the vehicles and inventory, that also builds the foundation of the final semantic map. It acts as the source of information, providing details to other layers of the map. In addition, the geometric and topological relations between semantically annotated instances of the map must be maintained.
- *Structure of the representation;* a layout map ("blueprint" such as a Computer Aided Design (CAD) drawing) is a suitable example of how the desired map should be structured. However, the expected semantic map shall support more features than a simple layout. It must improve the integrity of the map, while satisfying compatibility of the representation between agents.
- *Context of the representation;* the map is intended to be used for localization of both the vehicles and the articles in the environment. Therefore the map is expected to contain semantics of infrastructures for storage locations and of open-spaces for vehicles' trajectory.
- *Usability of the representation;* a representation must be understandable and usable (functional) by all involved agents.

Figure 4.1 demonstrates how the methodologies of the appended papers are situated with respect to the related work. The two attributes on the axes

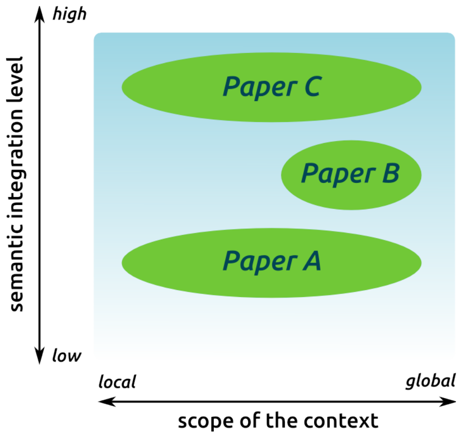


Figure 4.1: This figure shows the design space for a semantic mapping method. The two attributes on the axes have been discussed in the previous chapter (3).

have been discussed in the previous chapter (3).

The rest of this chapter presents summaries of the appended papers. Please note that the technical details of the developed methods are left in the appended papers. In this chapter, the aim is to provide a bigger picture overview than what each individual paper has to offer. To that end, the details of each appended paper is analyzed, following the concepts presented in chapters 2 and 3. Each summary consists of: *i*) an *abstract* overview of the paper; *ii*) *setup and method* covering sensors and prior knowledge, and describes the proposed algorithms in terms of transformation processes; *iii*) *result and review* covers the resulting semantic map, and addresses the levels of semantics incorporation and the scope of the context.

4.1 Paper A

In Paper A we present a canonical geometric-semantic model (figure 4.3d), along with a method for generating and matching these models into the latent structure of the map. Semantics are encoded into the model through the choice of landmarks (see figures 4.3b.) Figure 4.3c shows that the model geometrically represents the boundaries of corridors and pallet rack cells. Such a model enables the further processing of geometric-topological modeling of the environment, for the purpose of semantic annotation of infrastructures and geometric layout extraction. Different models are uniquely identified by the means of the model's parameters. Most of the steps and parameters of the model are independent, therefore it is adjustable according to different

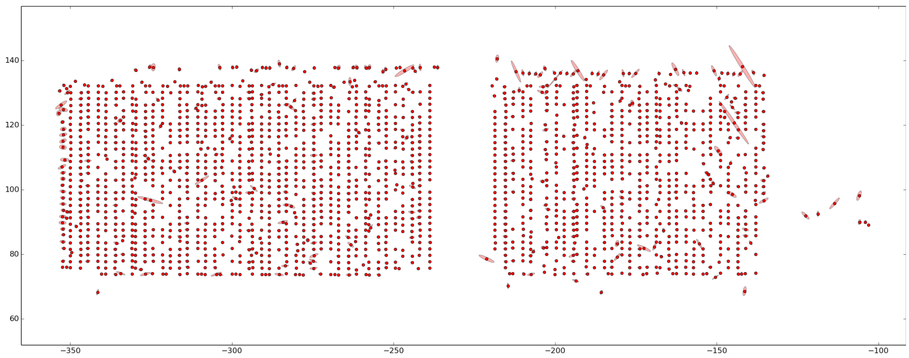


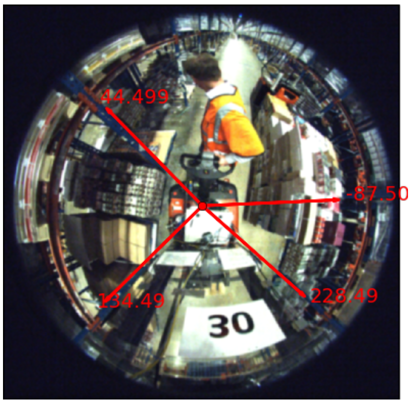
Figure 4.2: In this figure a landmark (pillars) map of a typical warehouse is presented. The map is the result of a bearing-only SLAM from a downward fisheye camera as in figure 4.3a.

scenarios. One example of such adjustment is where each individual model's parameters adopt the global characteristic of the map for a better consistency.

4.1.1 setup and method

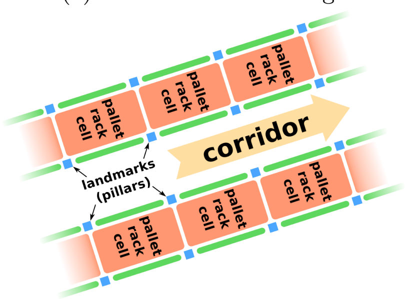
For this paper we used images from a downward looking fisheye camera as in figure 4.3a. Mapping is performed with an adapted version of a bearing-only SLAM proposed by [Bailey, 2003]. A set of prior knowledge contribute to the developed model and method. Most significant ones are: *i*) the appearance of infrastructures in a warehouse, that is used for landmark detection and mapping; *ii*) the knowledge of the environment's layout has contributed to the choice of landmark and models; and *iii*) a well-structured environment with repetitive pattern enabled a globally consistent modeling.

Each model is identified uniquely by five parameters (see figure 4.3d). θ is the angle of the model, and (x, y) is the starting point of the cluster. The last parameters n and d denote the number of and the distance between points (landmarks) in each model respectively. Paper A describes the methodology of finding these parameter for each model in detail. Briefly, modeling consists of 3 stages, *i*) a *closed-form Hough transform* developed for this purpose, which not only clusters landmarks in groups, but also gives an estimate of (θ, x, y); *ii*) frequency (n) and the length of the model (d) are estimated via Fourier transform analysis; and *iii*) in an optimization stage, estimated model is compared with the data it has been generated for fine-tuning. After segmenting the map and fitting with models, a global consistency check is carried out. The global consistency is expected, relying

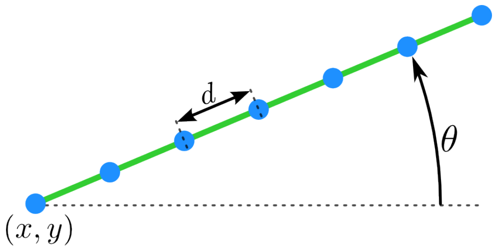


(a) landmarks in an image

(b) a scene in a warehouse corridor

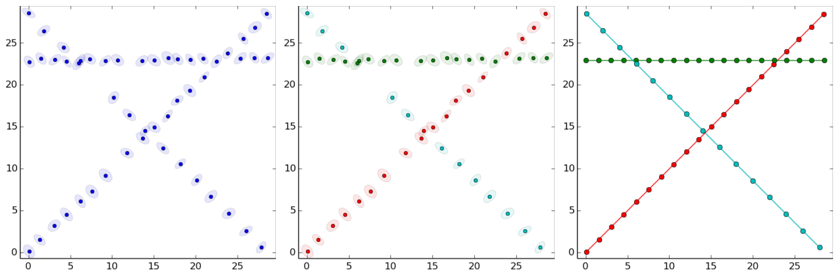


(c) target semantics of the model



(d) canonical geometric-semantic model

Figure 4.3: This figure illustrates how the latent structure of a typical warehouse scene could be semantically annotated. The proposed canonical model, captures the geometry of the infrastructures and suggests semantic labels for related instances. Please note that the proposed model targets to model a landmark (pillars) map such as the one in figure 4.2.



(a) synthetic landmarks (b) clustering (c) fitting models

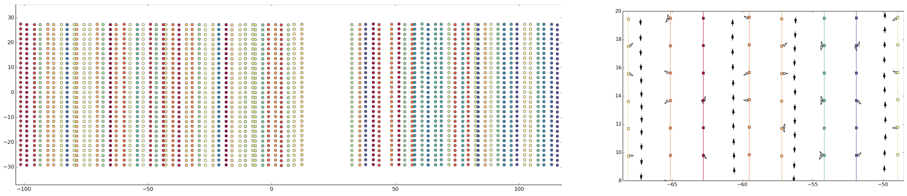
Figure 4.4: Closed-form Hough transform maps the points from Cartesian space into sinusoid in Hough space. A clustering is performed over intersections of the sinusoid in Hough space to cluster the aligned points in Cartesian space (details in paper A). After clustering the landmarks, each segment is fit with one model.

on the above-mentioned prior knowledge (that the environment is highly structured). To this end, the most dominant values (statistical mode) of different parameters is calculated, and the models are adjusted accordingly.

4.1.2 results and reviews

In this paper, we rely on the implication of the proposed method for the semantic annotation. That is to say, each point-line model is a corridor wall, and the space between two consecutive points in the model is the opening to a cell (as illustrated in figure 4.3.) The resulting semantic map consists of parametric models overlaid on top of the sensor-based map (the landmark map), serving as an abstraction layer. The results of modeling a real-worlds map (from figure 4.2) is illustrated in figure 4.5a.

Figure 4.6 summarizes the proposed method, in terms of the data flow in the system. The data flow is color codes, to distinguish between different levels of information, as presented in chapter 2. Yellow circles denote data, orange circles are information, and knowledge is represented by green circles. Note that the hollow green circles represent prior knowledge, which are not generated by method, but comes from human mind. The level of semantics incorporation in this work is low, as the semantics are hard coded into the models. The scope of the context used for the mapping is local in landmark detection, and global in modeling and semantic mapping.



(a) modeling the map from figure 4.2

(b) models on ground truth

Figure 4.5: Figure 4.5a shows the modeling of a warehouse (figure 4.2) with geometric-semantic models (all axes' unit in meter.) Figure 4.5b demonstrates a crop of the models, augmented with the truck's pose (black arrow) and reflectors (white triangles).

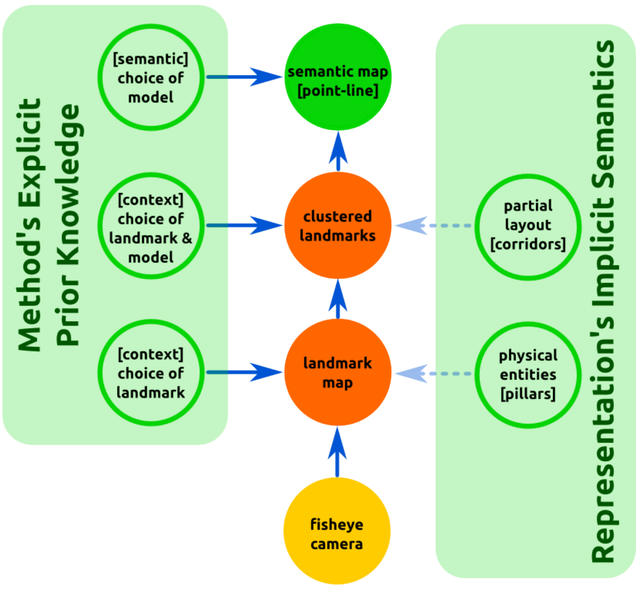


Figure 4.6: A review of data flow of the proposed method in paper A. Yellow circles denote data, orange circles are information, and knowledge is represented by green circles. For instance, based on the context (choice of the landmark) and input data, the map is generated on information level. Note that the hollow green circles represent prior knowledge, which are not generated by method, but comes from human mind.

4.2 Paper B

The proposed method in paper B interprets occupancy maps to bring out underlying spatial characteristics in a format readily usable by machines and humans. A particular challenge that our method in Paper B addresses is the discrepancy between the importance of orientation information and the absence of corresponding clean long lines in the range data that can be obtained via laser scanners in warehouses. The spatial arrangement of features is extracted via polygonal cell decomposition, and captured in a subdivision data structure to enable a higher-level topological analysis. We employ radiography to fit independently oriented and spaced lines to the occupancy map. After cell decomposition over those fitted lines, two corresponding data structures capture topological characteristics and ease subsequent semantic labeling. Our novel representation characterizes the geometric structure of the environment in addition to providing an abstract metric representation of the map. This focuses on an adaptive generic representation to support computing and storing of such information. The key features of the proposed method are the flexible and robust direction detection and the adaptive spacing of lines. Control of the spatial resolution at which grid lines should be inferred (scale of the Ricker wavelet) influences the abstraction level of the resulting topological information.

4.2.1 setup and method

A set of prior knowledge contributes to the developed model and method. Most significant ones are: *i*) occupancy, to detect open spaces and long lines (supposedly walls of the corridor); *ii*) well-structured environment, so the map could be abstracted via cell decomposition; and *iii*) shape of places, that are used as template graphs for pattern matching.

In paper B, occupancy grid maps of indoor environments (particularly warehouses) have been used for semantic annotation. A particular challenge that our method in Paper B addresses is to detect the salient structure of the environment in the absence of clean long lines in the range data. Figure 4.7 demonstrates the initial stages of the paper B. The initial stages are the core of the proposed method in paper B, which is decomposition and alternative representation. It starts by finding the salient structure of the environment and the dominant orientations of the environment (second from left in figure 4.7.) Then the principle lines (supposedly walls) of the environment are detected through a radiography of the map in the direction of the dominant orientations (third from left in the same figure.) The occupancy map is decomposed according to those dominant directions and lines (last from left in the same figure.) The alternative representation is composed of two

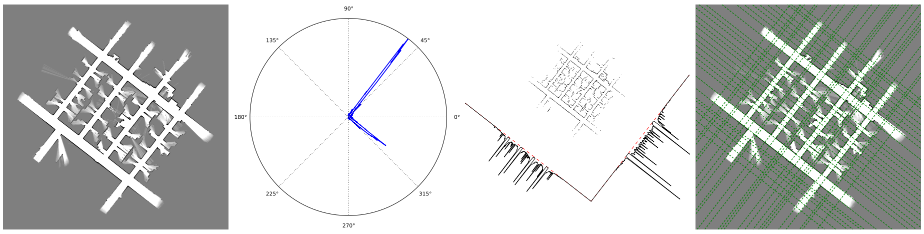


Figure 4.7: The core of the proposed method in paper B is decomposition. First the dominant orientations are extracted from the salient structure of the environment. The principle lines of the environment are detected through a radiography of the map in the direction of the dominant orientations. The occupancy map is decomposed according to those dominant directions and lines. The method is demonstrated on a map of the *Intel Jones Farms Campus*. This data set was obtained from the Robotics Data Set Repository (Radish[Howard and Roy, 2003]). We thank Maxim Batalin for providing this data.

topologically connected data structures, denoting adjacency and connectivity of the decomposed cells (see figure 4.8.) Figure 4.8 illustrated the dual structure overlaid on top of each other. In this paper the attribute to each decomposed cell comes from the occupancy value of all underlying occupancy grid cells (pixels) from the original map. Two cells are connected only if both represent open space, and the connection between them is not obstructed by a physical object. Semantic annotation is applied to open-spaces, through pattern matching between predefined template graphs denoting particular labels and the connectivity graph of decomposed cells (see figure 4.9.)

4.2.2 results and review

The result of the work is abstraction through decomposition of the occupancy map. The decomposition is represented by two graphs, who not only represent environment in an abstract form but also maintain the topological relations between the cells. The dual data structures lay on top of the occupancy grid map as an abstraction layer, and connect the low-level sensory data to higher conceptual layers. This aspect is more investigated in paper C. In this paper the attributes of each cell comes from the voting of underlying pixels from the occupancy map. Semantic labels in this work are those of open space, namely corridor, crossing and junction. Figure 4.9 illustrates the template graphs that are used for semantic annotation, and the result on the occupancy map. Please note that the target application are warehouses and a real world example of a warehouse is presented with details in paper B. However, the

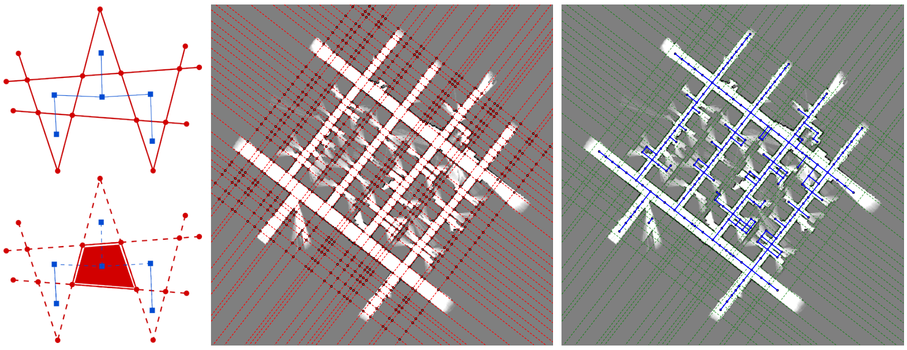


Figure 4.8: The alternative representation is composed of two topologically connected data structures (graphs), denoting adjacency (in red) and connectivity (in blue) of decomposed cells. The attribute to each decomposed cell comes from the occupancy value of all underlying occupancy grid cells from the original map. Two cells are connected only if both represent open space, they are connected and there is not physical obstruction in between. On the right side, you can see the two graphs being separately imposed on a real map (from figure 4.7)

example provided in this chapter is kept simple and from an indoor office environment (i.e. from figure 4.7 to figure 4.10.)

Figure 4.10 summarizes the proposed method, in terms of the data flow in the system. The level of semantics incorporation in this work is high, as the semantics are represented as templates. However, as it is signified in figure 4.10, there are implicit semantics in the representation forms that our method relies on. For instance, the *occupancy* is intrinsically implied by the representation of the occupancy grid map. Such semantics and contexts are the underlying prior knowledge in our system proposal. The scope of the context used for the detection of dominant orientation is global, which is based on the histogram of oriented gradient over the map. The context of the decomposition (e.g. long line detection, and dual graphs) and semantic labeling (template graphs) are global.

4.3 Paper C

The focus of paper C is the construction of a semantic map without providing semantic training data as an input to the system. The robot autonomously builds a high level spatial model of the world (based on the decomposition from paper B) and instantiates it, without prior knowledge of the environment. This allows an abstraction of human semantics, instead of providing this

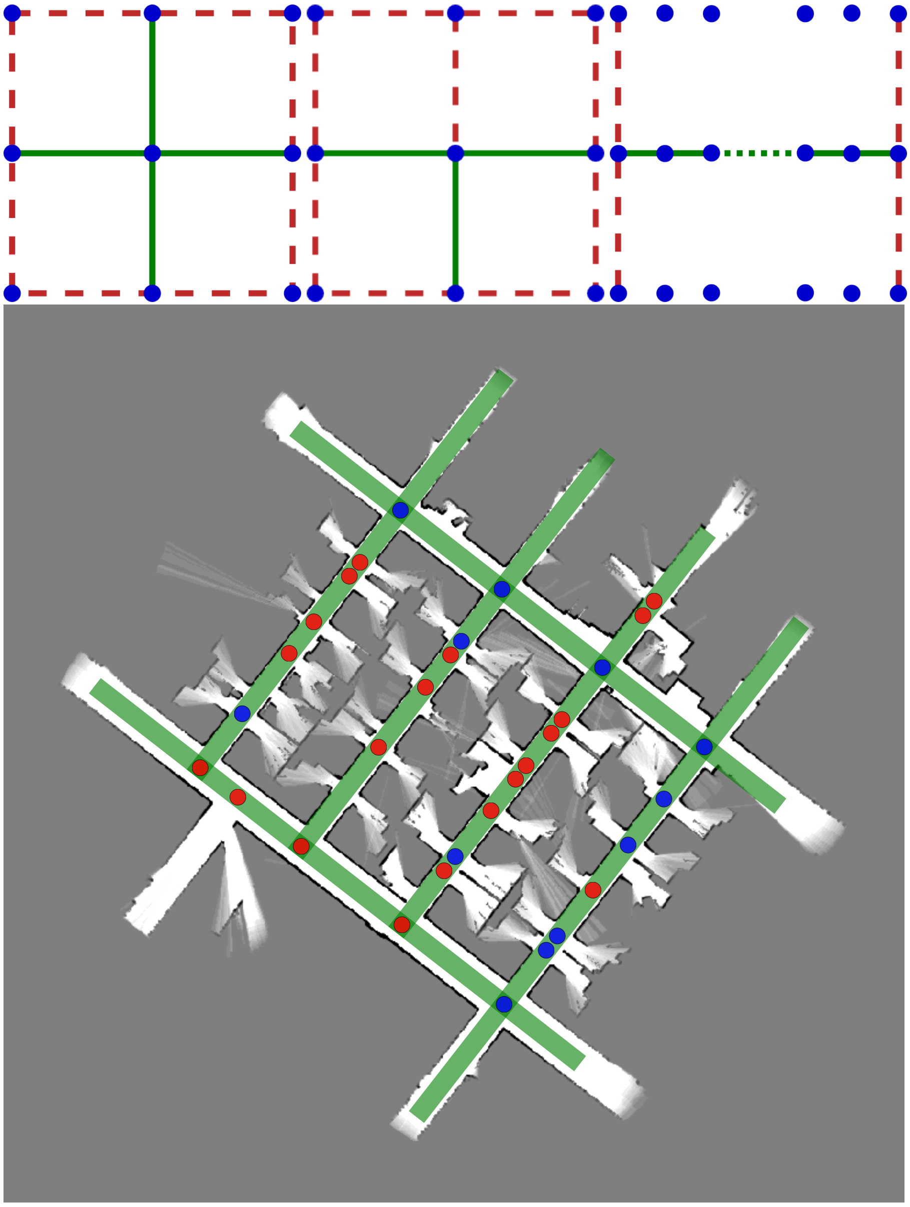


Figure 4.9: Result of paper B. The template graphs that are used for semantic annotation (top row) and the resulting semantic annotation of the map (bottom). The left template represents crossings, annotated with blue circles in the map. The middle template represents junctions, annotated with red circles in the map. The right template represents corridors, annotated with green stripes in the map.

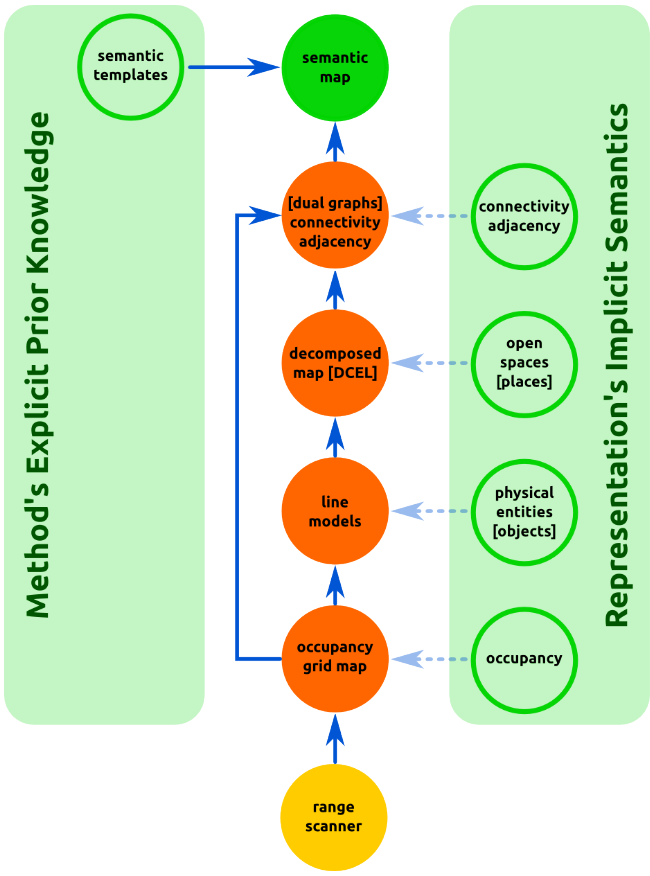


Figure 4.10: A review of data flow of the proposed method in paper B, and it follows the same color coding from figure 4.6.

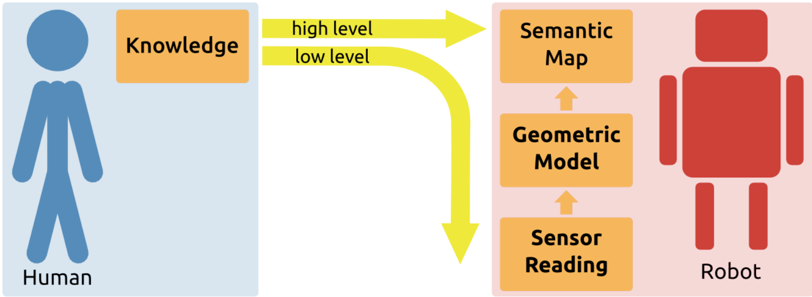


Figure 4.11: In the context of semantic mapping, semantics must be provided to the systems. The aim of our proposed method in paper C is to lift the input required from humans to a higher level of abstraction.

knowledge through a training set. Delivery of human semantics to the robot is done on a higher level, and directly supports semantic inference. These semantics are provided as labels accompanied with their functionality and possible relations between them. This allows the robot to perform an inference over its model and provided semantics, resulting in the desired semantic map suitable for operating autonomous robots. The contribution of this work is the autonomous generation of a high level representation that allows an efficient abstraction of the semantics based on human input.

4.3.1 setup and method

The set of prior knowledge contributing to the paper C is very similar to paper B. That means the most significant ones are: *i*) occupancy, to detect open spaces and long lines (supposedly walls of the corridor); *ii*) well-structured environment, so the map could be abstracted via cell decomposition; and *iii*) shape of places, affects the choice of feature extraction and clustering techniques.

Similar to paper B, occupancy grid map of indoor environments (particularly warehouses) are the target for semantic annotation. For cell decomposition and automatic spatial abstraction, we directly use the method proposed in paper B. In paper B the occupancy values has been transferred to the abstract layer (cell decomposition) for inference and semantic annotation. In paper C, each pixel of the occupancy grid map denoting open space is described via a feature vector, representing its surrounding appearance. The feature vector is proposed by [Mozos et al., 2005] and is based on the shape of the "raycast" from each point in the map (see figure 4.12). In [Mozos et al., 2005] the feature vector has been used for training classifiers

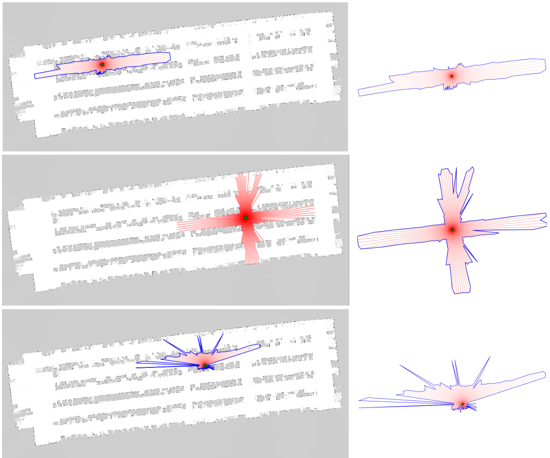


Figure 4.12: This figure shows examples of raycasting from different cells in an occupancy grid map of a warehouse. Examples are chosen from different type of places to emphasize the effect of the appearance in the resulting raycast.

with a training dataset. The classifier is then applied to test subjects for classification and semantic annotation. In paper C, we use a similar feature vector for unsupervised place categorization, instead of training classifiers. The method clusters the pixels of the occupancy grid map into different categories. Such categories are then passed into higher level of abstraction (cell decomposition) as attributes of the cell. Finally, the semantic labels are introduced to the system by the user on the highest level of abstraction. Figure 4.13 illustrates the parallel processes of the *spatial abstraction* and *unsupervised place categorization*. Merging of the two processes will result in a abstract representation of the environment, in which the different instances of the environment are discriminated.

4.3.2 results and review

Proposed method in paper C relies on the abstract representation of the cell decomposition, and the direct access that it provides to the sensory data (as in paper B). However in paper C, instead of direct occupancy values, a feature vector representing the appearance is of the region is made available to higher levels for inference and semantic annotation. Figure 4.14 illustrates how the semantic labels and their attributes are assigned to the categories resulted from the unsupervised place categorization. Resulting semantic map is illustrated in figure 4.13 (the top map). For more examples please see paper C.

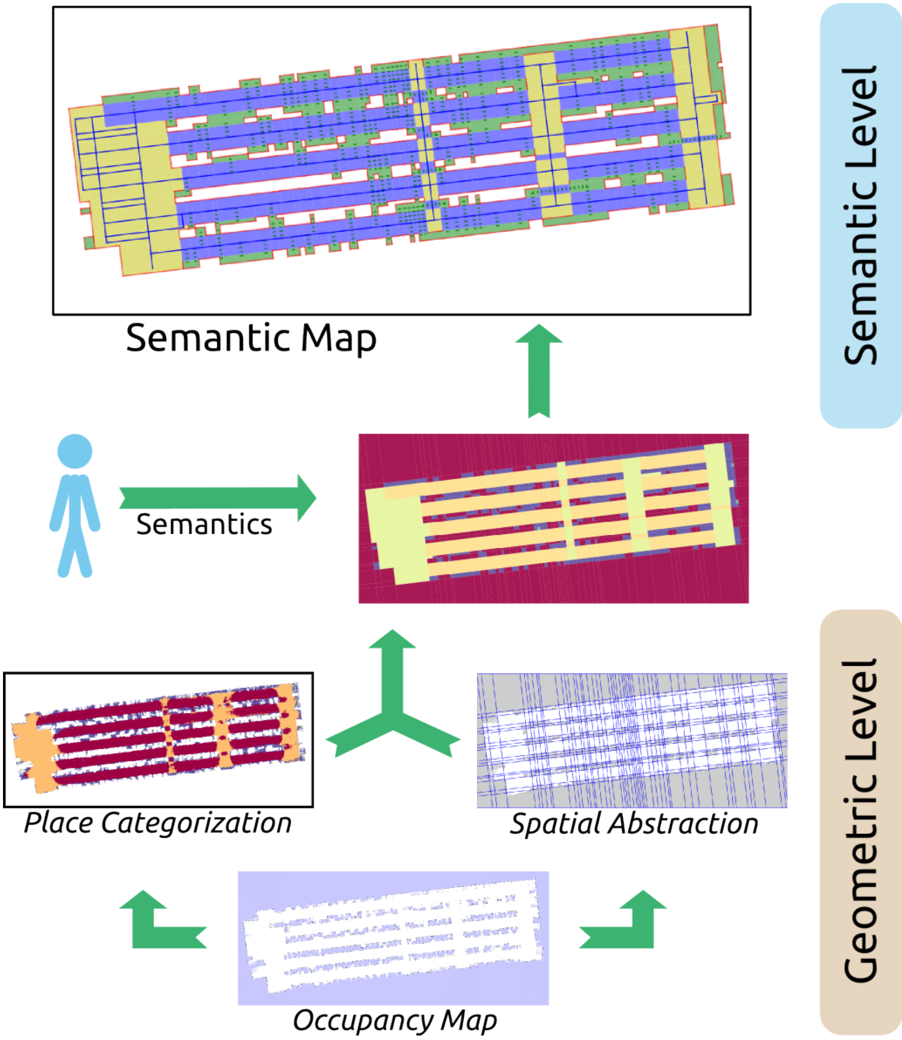
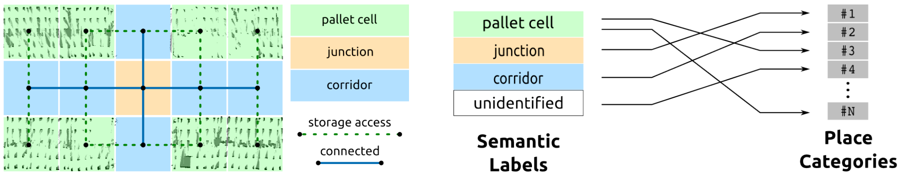


Figure 4.13: Overall flow of the paper C. The two parallel processes of the *spatial abstraction* and *unsupervised place categorization* result in an abstract representation of the environment, which differentiates between types of places. This allows the user to introduce semantics to the system on the highest level of abstraction (see figure 4.14.)



(a) semantic labels (b) assignment of labels to categories

Figure 4.14: Semantic labels of places, their connection, and assignment. For instance we can see corridors are traversable (connected) to other corridor and junctions. But corridor only have storage access (non-traversable) to the pallet cells. In this work the assignment of the predefined semantics to unknown place categories is performed manually in the last stage of the method.

Figure 4.15 summarizes the proposed method, in terms of the data flow in the system. The figure shows that paper C has an extra source of explicit knowledge in comparison to the proposed method in paper B (figure 4.10). The extra source is the choice of features. But this extra source of knowledge, simplifies the form of final semantic labels (as shown in figure 4.14). The level of semantics incorporation in this work is high, as the semantics are delivered as abstract labels. The scope of the context used for the detection of dominant orientation is local, which is based on the histogram of oriented gradient over the map. As well, feature extraction for each point from ray casting is somehow local. Yet, the comparison of all feature vector from all over the map with each other through the process of unsupervised place categorization (clustering) takes the global context into account.

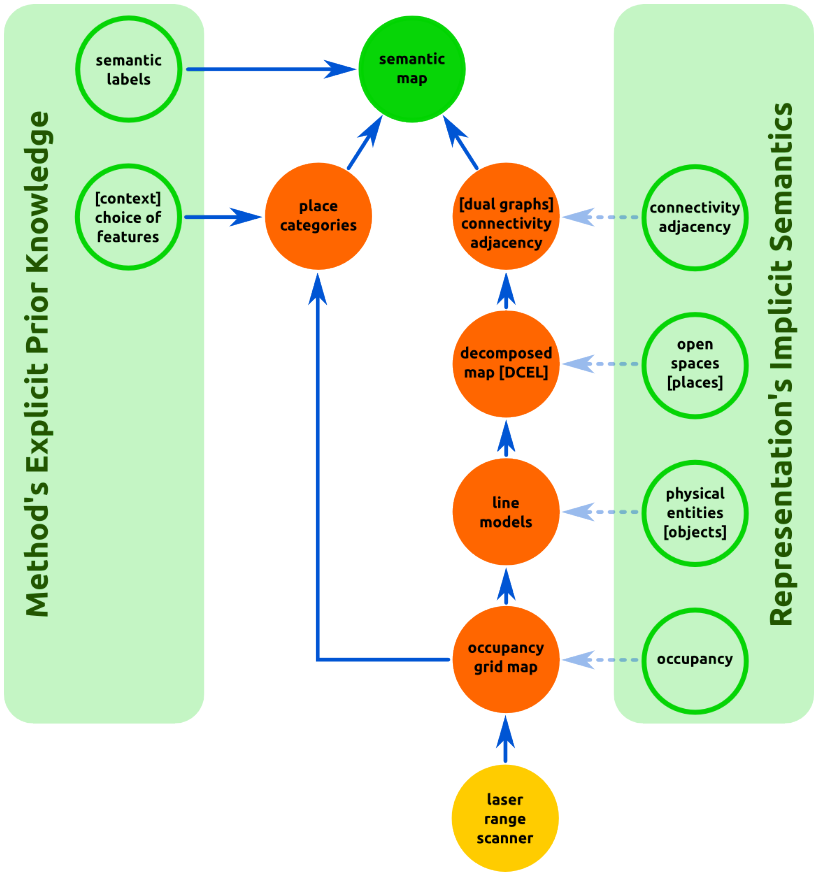


Figure 4.15: A review of data flow of the proposed method in paper C, and it follows the same color code from figures 4.6.

Chapter 5
Conclusion

"How will Socrates and Meno know when they have found the correct answer?"
–Thinking about android epistemology

In this thesis and appended papers we presented the process of studying the problem of environment modeling for autonomous agent. More specifically, the focus and objective of the works has been semantic mapping of warehouses. A semantic map for such purpose is expected to be layout-like and support semantics of both open spaces and infrastructure of the environment. The representation of the semantic map is required to be understandable by all involved agents (humans, AGVs and WMS.) And the process of semantic mapping is desired to lean toward full-autonomy, with minimum input requirement from human user. To that end, we studied the problem of semantic annotation over two kinds of spatial map from different modalities. We identified properties, structure, and challenges of the problem. And we have developed representations and accompanied methods, while meeting the set criteria. Contributions of the presented works fall into three groups: *i*) studying the problem and learning how to exploit the structure of the environment for abstraction; *ii*) as the framework of the thesis, a layer of abstraction for structuring and facilitate access to salient information in the sensory data; and *iii*) and technically, developing and implementing different representations and techniques for abstract modeling of a warehouse environment in different modalities.

Summary: The core of this thesis could be summarized as:

To develop and construct a layer of abstraction (models and/or decomposition) for structuring and facilitate access to salient information in the sensory data. This layer of abstraction connects high level concepts to low-level sensory pattern.

Figure 5.1 illustrates a conceptual summary of this thesis and appended papers.

Usability and Limitations: The proposed methods and their implementations have been successfully verified over data from real-world. While the proposal of paper A is application-driven and context-dependent, paper B and C propose methods that are not specific to warehouse environment. The performance of the methods in paper B and C have been successfully verified in environments other than warehouses (e.g. office.) However, the specific methods and representations are developed for well structured environments, with the assumption that the map could be effectively abstracted via straight lines. While the implementation might be modality-dependent, however the link to conceptual layer does not have to be dependent on the sensor type.

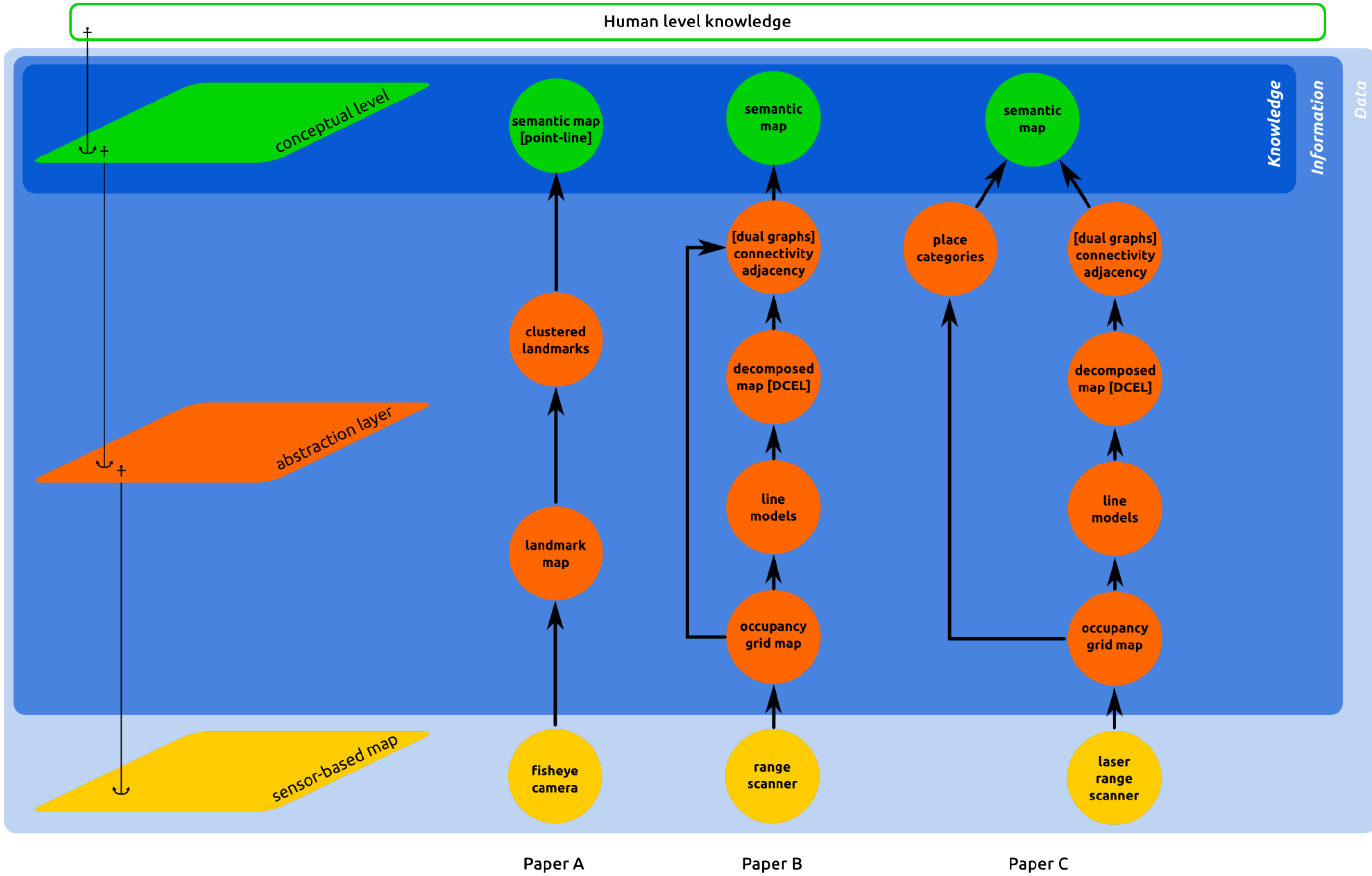


Figure 5.1: A summary of the conceptual contribution and the core message of this thesis and appended papers.

5.1 Future Work

Following this thesis, we plan to extend the scope of our application by generalization of the proposed methods into a framework. Figure 5.2 shows an implementation proposal for the system. Aiming to reduce modality and context dependency, a more generic knowledge representation and the use of description languages[1] for annotation will be considered. In order to evaluate system's ability to handle different modalities, adding other sensors such as 3D cameras is in order (an extra inputs in figure 5.2.) While supporting multi-modal perception, we plan to develop a semantic-based map merging and information fusion method (the association stage in figure 5.2.) To that end, a minimalistic representation of concepts that supports multi-resolution (the abstraction layer) and provides a hierarchy of the information is desired. Temporal identification of semantic labels is of great importance in dealing with dynamic environments. Including temporal behavior of each category in its semantic description, enables behavior-based inference in addition to identity-based approaches. The challenge in temporal modeling of semantics consists of invariant motion model representation and inferences methodologies. The temporal characteristic of an object is particularly important for inventory list maintenance and obstacle avoidance.

[1] Based on the concept of ontology, similar to "Unified Robot Description Format" (URDF) or "Web Ontology Language" (OWL)

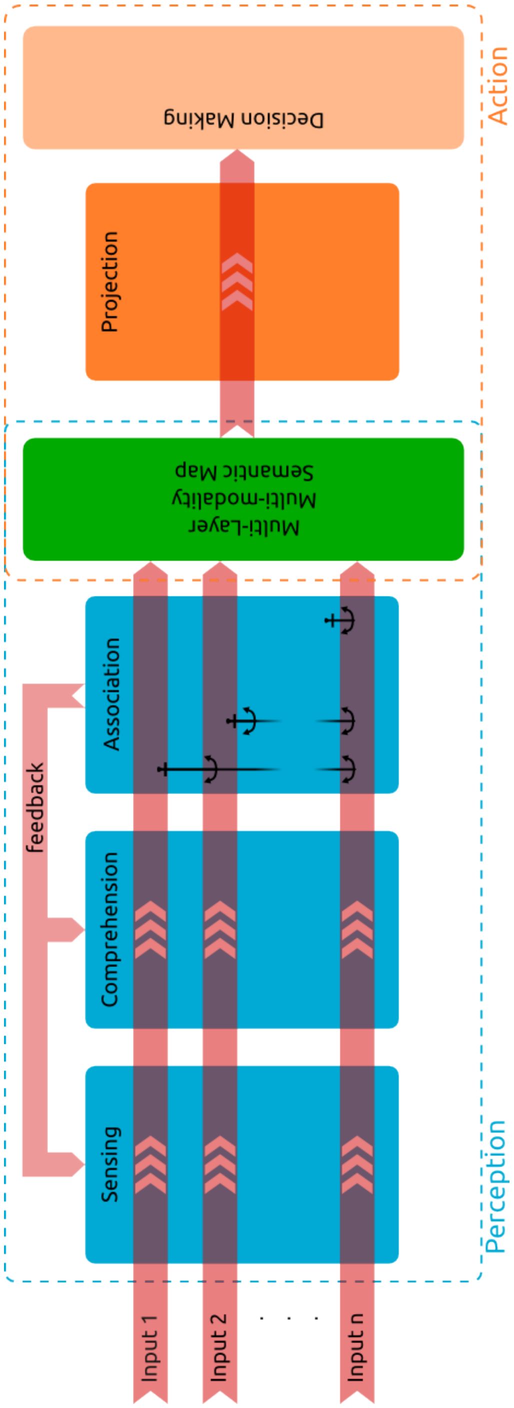


Figure 5.2: Generalization of proposed methods of this thesis, a future plan.

References

[Ackoff, 1989] Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1):3–9. (Cited on pages 11, 14, 15, and 16.)

[Aydemir et al., 2011] Aydemir, A., Göbelbecker, M., Pronobis, A., Sjöö, K., and Jensfelt, P. (2011). Plan-based object search and exploration using semantic spatial knowledge in the real world. In *ECMR*, pages 13–18. (Cited on pages 18 and 27.)

[Bailey, 2003] Bailey, T. (2003). Constrained initialisation for bearing-only slam. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 2, pages 1966–1971. IEEE. (Cited on page 31.)

[Beinschob and Reinke, 2014] Beinschob, P. and Reinke, C. (2014). Advances in 3d data acquisition, mapping and localization in modern large-scale warehouses. In *Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on*, pages 265–271. IEEE. (Cited on page 22.)

[Bocij et al., 2008] Bocij, P., Greasley, A., and Hickie, S. (2008). *Business information systems: Technology, development and management*. Pearson education. (Cited on page 15.)

[Cardarelli et al., 2014] Cardarelli, E., Sabattini, L., Secchi, C., and Fantuzzi, C. (2014). Multisensor data fusion for obstacle detection in automated factory logistics. In *Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on*, pages 221–226. IEEE. (Cited on page 22.)

[Cardarelli et al., 2015] Cardarelli, E., Sabattini, L., Secchi, C., and Fantuzzi, C. (2015). Cloud robotics paradigm for enhanced navigation of autonomous vehicles in real world industrial applications. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4518–4523. IEEE. (Cited on page 22.)

[Coradeschi and Loutfi, 2008] Coradeschi, S. and Loutfi, A. (2008). *A review of past and future trends in perceptual anchoring*. INTECH Open Access Publisher. (Cited on page 18.)

[Coradeschi and Saffiotti, 2003] Coradeschi, S. and Saffiotti, A. (2003). An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2):85–96. (Cited on page 18.)

[Curtis and Cobham, 2008] Curtis, G. and Cobham, D. (2008). *Business information systems: Analysis, design and practice*. Pearson Education. (Cited on page 15.)

[Daoutis, 2013] Daoutis, M. (2013). Knowledge based perceptual anchoring. *KI-Künstliche Intelligenz*, 27(2):179–182. (Cited on page 18.)

[de la Puente and Rodriguez-Losada, 2014] de la Puente, P. and Rodriguez-Losada, D. (2014). Feature based graph-slam in structured environments. *Autonomous Robots*, pages 1–18. (Cited on page 27.)

[Endsley, 1995] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64. (Cited on page 13.)

[Endsley, 2011] Endsley, M. R. (2011). *Designing for situation awareness: An approach to user-centered design*. CRC press. (Cited on page 13.)

[Fabrizi and Saffiotti, 2000] Fabrizi, E. and Saffiotti, A. (2000). Extracting topology-based maps from gridmaps. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 3, pages 2972–2978. IEEE. (Cited on page 27.)

[Freeman, 2001] Freeman, W. J. (2001). The neurobiology of semantics: how can machines be designed to have meanings? *What Should be Computed to Understand and Model Brain Function?: From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy*, 3:207. (Cited on page 7.)

[Friedman et al., 2007] Friedman, S., Pasula, H., and Fox, D. (2007). Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, volume 35. (Cited on page 27.)

[Galindo et al., 2005] Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernandez-Madrigal, J.-A., and González, J. (2005). Multi-hierarchical semantic maps for mobile robotics. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 2278–2283. IEEE. (Cited on page 27.)

[Gärdenfors, 2004] Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press. (Cited on page 17.)

[Gärdenfors, 2014] Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press. (Cited on pages 17, 18, and 24.)

[Gunther et al., 2013] Gunther, M., Wiemann, T., Albrecht, S., and Hertzberg, J. (2013). Building semantic object maps from sparse and noisy 3d data. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2228–2233. IEEE. (Cited on page 26.)

[Howard and Roy, 2003] Howard, A. and Roy, N. (2003). The Robotics Data Set Repository (Radish). (Cited on page 36.)

[Joo et al., 2010] Joo, K., Lee, T.-K., Baek, S., and Oh, S.-Y. (2010). Generating topological map from occupancy grid-map using virtual door detection. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–6. IEEE. (Cited on page 27.)

[Kostavelis and Gasteratos, 2013] Kostavelis, I. and Gasteratos, A. (2013). Learning spatially semantic representations for cognitive robot navigation. *Robotics and Autonomous Systems*, 61(12):1460–1475. (Cited on pages 18 and 27.)

[Kostavelis and Gasteratos, 2015] Kostavelis, I. and Gasteratos, A. (2015). Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66:86–103. (Cited on pages 18 and 22.)

[Krug et al., 2016] Krug, R., Stoyanov, T., Tincani, V., Andreasson, H., Mosberger, R., Fantoni, G., and Lilienthal, A. J. (2016). The next step in robot commissioning: Autonomous picking & palletizing. *IEEE Robotics and Automation Letters, under review–manuscript available at: http://www. aass. oru. se/Research/Learning/publications/APPLE. pdf*. (Cited on page 22.)

[Kuipers, 2000] Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial intelligence*, 119(1):191–233. (Cited on pages 18, 24, and 27.)

[Liu and von Wichert, 2013] Liu, Z. and von Wichert, G. (2013). Extracting semantic indoor maps from occupancy grids. *Robotics and Autonomous Systems*. (Cited on page 27.)

[Lundström, 2014] Lundström, J. (2014). Situation awareness in colour printing and beyond. (Cited on page 13.)

[Milella et al., 2014] Milella, A., Reina, G., Underwood, J., and Douillard, B. (2014). Visual ground segmentation by radar supervision. *Robotics and Autonomous Systems*, 62(5):696–706. (Cited on page 27.)

[Mosberger et al., 2014] Mosberger, R., Andreasson, H., and Lilienthal, A. J. (2014). A customized vision system for tracking humans wearing reflective safety clothing from industrial vehicles and machinery. *Sensors*, 14(10):17952–17980. (Cited on page 22.)

[Mozos et al., 2005] Mozos, O. M., Stachniss, C., and Burgard, W. (2005). Supervised learning of places from range data using adaboost. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1730–1735. IEEE. (Cited on pages 27 and 40.)

[Nüchter and Hertzberg, 2008] Nüchter, A. and Hertzberg, J. (2008). Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 56(11):915–926. (Cited on page 27.)

[Nüchter et al., 2014] Nüchter, A., Rusu, R. B., Holz, D., and Munoz, D. (2014). Semantic perception, mapping and exploration. *Robotics and Autonomous Systems*, 62(5):617 − 618. Special Issue Semantic Perception, Mapping and Exploration. (Cited on pages 19, 23, and 24.)

[Nüchter et al., 2005] Nüchter, A., Wulf, O., Lingemann, K., Hertzberg, J., Wagner, B., and Surmann, H. (2005). 3d mapping with semantic knowledge. In *RoboCup 2005: Robot Soccer World Cup IX*, pages 335–346. Springer. (Cited on page 27.)

[Nur et al., 2015] Nur, K., Morenza-Cinos, M., Carreras, A., and Pous, R. (2015). Projection of rfid-obtained product information on a retail stores indoor panoramas. *Intelligent Systems, IEEE*, 30(6):30–37. (Cited on page 22.)

[Persson et al., 2007] Persson, M., Duckett, T., Valgren, C., and Lilienthal, A. (2007). Probabilistic semantic mapping with a virtual sensor for building/nature detection. In *Computational Intelligence in Robotics and Automation, 2007. CIRA 2007. International Symposium on*, pages 236–242. IEEE. (Cited on page 27.)

[Pronobis and Jensfelt, 2012] Pronobis, A. and Jensfelt, P. (2012). Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3515–3522. IEEE. (Cited on page 27.)

[Ranganathan and Dellaert, 2007] Ranganathan, A. and Dellaert, F. (2007). Semantic modeling of places using objects. In *Proceedings of the 2007 Robotics: Science and Systems Conference*, volume 3, pages 27–30. (Cited on page 27.)

[Rottmann et al., 2005] Rottmann, A., Mozos, Ó. M., Stachniss, C., and Burgard, W. (2005). Semantic place classification of indoor environments with mobile robots using boosting. In *AAAI*, volume 5, pages 1306–1311. (Cited on page 27.)

[Rowley, 2007] Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*. (Cited on pages 14 and 15.)

[Rusu, 2010] Rusu, R. B. (2010). Semantic 3d object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, 24(4):345–348. (Cited on page 27.)

[Rusu et al., 2008] Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M., and Beetz, M. (2008). Towards 3d point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11):927–941. (Cited on page 27.)

[Stoyanov et al., 2013] Stoyanov, T., Saarinen, J., Andreasson, H., and Lilienthal, A. J. (2013). Normal distributions transform occupancy map fusion: Simultaneous mapping and tracking in large scale dynamic environments. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 4702–4708. IEEE. (Cited on page 22.)

[Stoyanov et al., 2016] Stoyanov, T., Vaskevicius, N., Müller, C., Fromm, T., Krug, R., Tincani, V., Mojtahedzadeh, R., Kunaschk, S., Ernits, R. M., Canelhas, D. R., et al. (2016). No more heavy lifting: Robotic solutions to the container unloading problem. *IEEE Robotics and Automation Magazine, under review–manuscript available at: http://www. aass. oru. se/Research/Learning/publications/RobLog. pdf*. (Cited on page 22.)

[Thrun et al., 2005] Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic robotics*. MIT press. (Cited on page 18.)

[Valencia et al., 2014] Valencia, R., Saarinen, J., Andreasson, H., Vallvé, J., Andrade-Cetto, J., and Lilienthal, A. J. (2014). Localization in highly dynamic environments using dual-timescale ndt-mcl. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 3956–3962. IEEE. (Cited on page 22.)

[Vasiljevic et al., 2014] Vasiljevic, G., Petric, F., and Kovacic, Z. (2014). Multi-layer mapping-based autonomous forklift localization in an industrial environment. In *Control and Automation (MED), 2014 22nd Mediterranean Conference of*, pages 1134–1139. IEEE. (Cited on page 22.)

[Vasudevan et al., 2007] Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. (2007). Cognitive maps for mobile robotsâĂŤan object based approach. *Robotics and Autonomous Systems*, 55(5):359–371. (Cited on page 27.)

[Vatsavai et al., 2010] Vatsavai, R. R., Cheriyadat, A., and Gleason, S. (2010). Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 273–280. IEEE. (Cited on page 27.)

[Walter et al., 2014] Walter, M. R., Hemachandra, S., Homberg, B., Tellex, S., and Teller, S. (2014). A framework for learning semantic maps from grounded natural language descriptions. *The International Journal of Robotics Research*, 33(9):1167–1190. (Cited on page 26.)

[Wolf and Sukhatme, 2008] Wolf, D. F. and Sukhatme, G. (2008). Semantic mapping using mobile robots. *Robotics, IEEE Transactions on*, 24(2):245–258. (Cited on page 27.)

[Wurm et al., 2014] Wurm, K. M., Kretzschmar, H., Kümmerle, R., Stachniss, C., and Burgard, W. (2014). Identifying vegetation from laser data in structured outdoor environments. *Robotics and Autonomous Systems*, 62(5):675–684. (Cited on page 27.)

[Zender et al., 2008] Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G.-J., and Burgard, W. (2008). Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502. (Cited on page 27.)

[Zhao, 2013] Zhao, Q. (2013). Computational awareness: another way towards intelligence. In *Computational Intelligence*, pages 3–14. Springer. (Cited on pages 11, 12, and 13.)

[Zhao et al., 2012] Zhao, Q., Hsieh, C.-H., Naruse, K., and She, Z. (2012). Awareness science and engineering. *Applied Computational Intelligence and Soft Computing*. (Cited on pages 3, 11, and 12.)

[Zivkovic et al., 2007] Zivkovic, Z., Booij, O., and Kröse, B. (2007). From images to rooms. *Robotics and Autonomous Systems*, 55(5):411–418. (Cited on page 27.)

Paper A - Modeling of a Large Structured Environment With a Repetitive Canonical Geometric-Semantic Model.

Modeling of a Large Structured Environment
With a Repetitive Canonical Geometric-Semantic Model *

Saeed Gholami Shahbandi and Björn Åstrand

Center for Applied Intelligent Systems Research (CAISR),
Intelligent Systems Lab, Halmstad University, Sweden

Abstract. AIMS project attempts to link the logistic requirements of an intelligent warehouse and state of the art core technologies of automation, by providing an awareness of the environment to the autonomous systems and vice versa. In this work we investigate a solution for modeling the infrastructure of a structured environment such as warehouses, by the means of a vision sensor. The model is based on the expected pattern of the infrastructure, generated from and matched to the map. Generation of the model is based on a set of tools such as closed-form Hough transform, DBSCAN clustering algorithm, Fourier transform and optimization techniques. The performance evaluation of the proposed method is accompanied with a real world experiment.

1 Introduction

Following the advances of the state of the art in autonomous vehicles, and growing research on the design and development of innovative solutions, *intelligent warehouses* emerge leveraging insights from several specialist domains. The Automatic Inventory and Mapping of Stock (AIMS) project targets the traditional warehouses where not necessarily infrastructures are designed or installed for automation. AIMS project intends to develop a process, through which an awareness of the surrounding environment is embedded in a "live" semantic map, for effective management of logistics and inventory (see fig. 1). Achieving this objective requires different sensors contributing to this semantic map in multiple layers. Forklift trucks enriched with such an awareness enable safe and efficient operations while sharing the workspace with humans. In such a shared workspace, compatibility between vehicles' knowledge, humans' cognition and Warehouse Management Systems (WMS) is important.

This paper focuses on mapping and modeling the infrastructure of the warehouse, as a foundation for addressing the location of both vehicles and storage of the warehouse. We present a method to extract structural pattern of the map that serves as the foundation of a multilayer geometric-semantic map to be used for logistic planning, Auto Guided Vehicle (AGV) path planning and WMS interaction. Semantic labels are subject to the context, in order to be functional

* This work as a part of AIMS project, is supported by the Swedish Knowledge Foundation and industry partners Kollmorgen, Optronic, and Toyota Material Handling Europe.

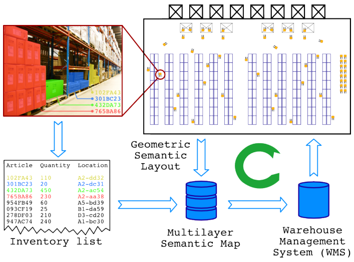


Fig. 1: The objective in AIMS project is to provide an awareness framework. The framework interacts with Warehouse Management System (WMS).

they shall be defined based on the environment and application. Accordingly we integrate multiple layers in this map, from geometric layout design of the environment to semantic annotation of the regions and infrastructures as a model. For this purpose a map of pillars of the pallet rack cells (see fig. 1) is generated, from a fish-eye camera based on a bearing-only mapping method. This map is used for inference and generation of a layout map of the warehouse, which integrates the desired conceptual meaning. Contribution of this work is a semantic-geometric model, based on an object map entailing infrastructure elements as landmarks. This is followed by an easement in inferences through a method for extracting and matching the model from and to the aforementioned map. Reliability and performance of the proposed model and accompanying method is demonstrated on a map acquired from a real world warehouse.

1.1 Related Works

In a time where robots are embedded in our daily life, robot's awareness of their surrounding is a crucial competence and semantic mapping is a particularly important aspect of it. That is because, while a geometrically consistent map is sufficient for navigation, it is not enough for task planning and reasoning. Many researchers have been contributing to this particular aspect, from different points of view.

Some tried to model the environment through the topology of open space in geometric map, like [6], where they employed a series of kernels for semantic labeling of regions. Some others like [5], [14] and [9] proposed spatial maps, enhanced conceptually by object recognition in regions of the map. [5] proposed a composition of two hierarchical maps, semantic and geometric anchored together. In [14] a framework of a multilayer conceptual map is developed, representing the spatial and functional properties of the environment. And [9] introduced a comprehensive framework of spatial knowledge representation for large scale semantic mapping.

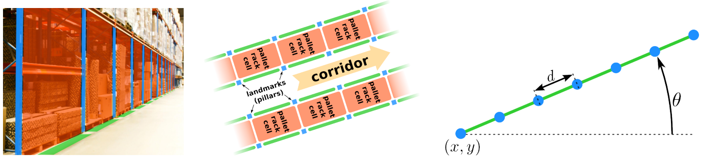
While mentioned researchers aimed to derive semantic concept from the functionality of the objects into the map, some others such as [8], [12] and [7], introduced properties of the regions as semantic label. [8] annotates an occupancy map with the properties of the regions, either "building" or "nature", through data from range scanner and vision. In [12] properties of the environment such as terrain map and activity map, are embedded into a metric occupancy map. Concerning the global localization, [7] employed hybrid geometric-object map.

Mentioned works are proposed for cases where the global structure of the environment is not a concern, and semantic information is extracted locally. The conceptual meaning is the property or functionality of the content of those regions, and does not link to the structural model of the environment. Researchers have also taken into account the environment's structure. [13] attempted mastering the SLAM problem by a geometrical object-oriented map representation, through modeling the boundaries of obstacles with polygons, employing a Discrete Curve Evolution (DCE) technique. An interesting recent work [10], developed a method for detection, evaluation, incorporation and removal of structure constraint from latent structure of the environment into a graph-based SLAM. Two last examples take into account the structure of the environment, for improving the solution to SLAM problem and providing a more consistent map. However there is no conceptual meaning accompanying the extracted structure.

Spatial semantic of open space from occupancy map is an interesting aspect and we investigate it in another work. It does provide useful knowledge of the open space in a warehouse, such as connectivity, corridors, or crossing of the corridors. But it does not provide semantic labels for infrastructure, such as the entrance of a pallet rack cell (see fig. 2). Such an information is very useful when the articles are localized in the layout, for logistic and AGVs' task planning. The other approaches to semantic mapping, where the semantic labels are derived from objects in the region is not very beneficial to our work either. That is because the smallest entities of regions are pallet rack cells with same semantics. Stored articles in those pallet rack cells and their identities do not carry any conceptual meaning for their region. The objects that we are interested in are the infrastructure of the warehouse, such as pillars of the pallet rack cells which represent the structure of the environment (see fig. 2). Therefore a more suitable approach for us would be to create a map of the environment, using the infrastructure as landmarks, and then extracting those patterns that are meaningful for us.

1.2 Our Approach

This paper presents a canonical geometric model and describes how to match it into the latent structure of a map. Such a model enables the further processing of geometric-topological modeling of the environment, for the purpose of semantic annotation of structures and geometric layout extraction. Semantic concept is encoded into the model through the choice of landmarks in the map as shown in fig. 2. It is assumed that the environment is highly structured and a pattern is frequent enough, so that it is possible to effectively represent the whole structure by a set of this canonical model with different parameters. By choosing pillars as landmarks, opening of a pallet rack cell is implied by two neighboring landmarks, while the sequence of landmarks creates a layout map by the "boundaries" of the corridors (a similar concept of representation as in



(a) warehouse scene (b) top view of warehouse (c) geometric semantic model

Fig. 2: Design of the model derived from the map's context, encoding semantic-geometric relation of landmarks.

[13]). This explains how the model in fig. 2 embeds a sufficient level of semantic and geometric knowledge while it is expressed only by a set of points uniformly distributed on a straight line. Different models are uniquely identified by the means of 5 parameters n, d, x, y, θ. The model and matching method is designed to represent independent lines of landmarks, therefore parameters of different models are independent. However the repetition of the model in the map of a warehouse, makes it possible to pose a global constraint on its parameters in order to achieve a set of models which are globally consistent.

In next section (2), the model and the matching method are explained in details. The process is demonstrated on a synthesized data in order to sketch the generality of the method where the models in one map are completely independent. Section 3 contains the method we adapt for mapping the environment from a fish-eye camera. Resulting map is modeled by the proposed method, while introducing global constraints for a real world map.

2 Model Generation and Fitting

Let's assume a $\mathcal{MAP}$ consisting of landmarks is given, where each landmark is described by its pose ($\mu = x, y$) and corresponding uncertainty modeled with a 2D normal distribution (with covariance Σ) as defined in equation 1.

$$\mathcal{MAP} = \{lm_i \mid lm_i := \mathcal{N}_i(X)\} \quad , \mathcal{N}_i(X) = \frac{\exp\left(-\frac{1}{2}(X-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(X-\mu_i)\right)}{\sqrt{(2\pi)^2|\Sigma_i|}} \tag{1}$$

The expected pattern in this map was described and motivated in the introduction (section 1). This model as illustrated in fig. 2, represents a set of landmarks (pillars) aligned on one side of a corridor. We call these sets of landmarks $\mathcal{L}$, and we try to fit one model per set.

$$\mathcal{L}_i = \{\{lm_{ij}\}, \theta_i \mid lm_{ij} \in \mathcal{MAP}\} \tag{2}$$

In the definition of $\mathcal{L}_i$ (equation 2) lm_{ij} are landmarks aligned on the side a corridor, and θ_i is the angle of alignment. Model M shown in fig. 2c and expressed in equation 3 is designed to represent sets of landmarks $\mathcal{L}$, and each model is uniquely identified by 5 parameters (n, d, x, y, θ), where n is the number of landmarks in the corresponding $\mathcal{L}$, d is the distance between two consecutive landmarks, (x, y) are the coordination of the 1st landmark, and θ_i is the angle of alignment of the set.

$$M_i(n_i, d_i, x_i, y_i, \theta_i) := \{p_{ij}\} \quad , p_{ij} = \begin{bmatrix} x_i + n_j d_i \cos\theta_i \\ y_i + n_j d_i \sin\theta_i \end{bmatrix}, 0 \leq n_j < n_i \tag{3}$$

First step in generating the models M is to segment the $\mathcal{MAP}$ into the sets of landmarks $\mathcal{L}_i$. For this purpose we have developed a closed-form of Hough transform in combination with a clustering algorithm. This technique not only locates the desired $\mathcal{L}$ by (θ, ρ), but also directly clusters the landmarks into different $\mathcal{L}$. After the segmentation of landmarks into sets, each set is projected into an axis passing through that set (see fig. 4b). This operation will map the 2D normal distributions of the landmarks into a 1D signal. An analysis of the resulting 1D signal in frequency domain will result in an estimation of the n and d of the model. Given the θ, n and d of the model by closed-form Hough transform and frequency analysis, x and y remain for extraction. An optimization would serve this purpose, where the first point of the set $\mathcal{L}$ serves as the initial guess of the optimization process.

2.1 Segmentation by Closed-Form Hough Transform

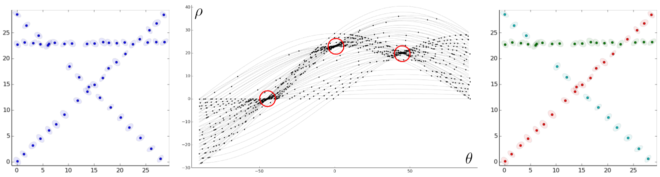
Transformation of a point from Cartesian space (x, y) into Hough space [3] (θ, ρ) is performed by equation 4.

$$\rho = x \cos\theta + y \sin\theta \tag{4}$$

Conventional form of Hough transform is applied to discrete images. Hough space is also a discretized image where the value of each pixel (θ, ρ) is given by the summation of the value of all pixels (x, y) that satisfy the equation 4. The peaks in the Hough space represent lines in original image where points are aligned. But we are interested in more than that. We would like to know which particular points in Cartesian space contributed to an specific peak (θ, ρ) in Hough space. Therefore we introduce a closed-form solution of Hough transform to address that issue. This is realized by representing each point (x_i, y_i), with a corresponding sinusoid as expressed in equation 4. Next step is to intersect all resulting sinusoids and store the intersections with $(\theta_{ij}, \rho_{ij}, i, j)$. Where (θ_{ij}, ρ_{ij}) represent the location of the intersection in Hough space and i, j are the indices of intersecting sinusoids. Outcome of this step is a set of intersection points in Hough space as illustrated in fig. 3b.

Advantages of closed-form approach are, first, clustering the intersection points in Hough space directly corresponds to clustering aligned points in Cartesian space. Secondly, it prevents us from discretization of a continues Cartesian space.

From the Hough space, clustering the intersection points is straightforward. We employed the Density-based spatial clustering of applications with noise (DBSCAN) [4] algorithm. This algorithm requires a value for the minimum number



(a) Cartesian space (b) Hough space (c) clustered sets

Fig. 3: Closed-form Hough transform maps the points from Cartesian space into sinusoids in Hough space. A clustering is performed over intersections in Hough space (red circles) to cluster the aligned points in Cartesian space.

of samples per cluster, which is derived from the number of expected points in each alignment $\mathcal{L}$. Assuming a set of n point aligned in Cartesian space ($\mathcal{L}$), consequently there will be the same number of sinusoids in Hough space intersecting $\frac{n(n-1)}{2}$ times at the same location (θ, ρ). As a perfect alignment in noisy data is unlikely, minimum number of samples are set to half of that ($\frac{n(n-1)}{4}$) to guarantee a successful clustering.

The result of clustering in Hough space is mapped to Cartesian space, generating sets of points $\mathcal{L}_i$ as desired. Angle of alignment $\theta_i \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$, is therefore the first estimated parameter of the model.

2.2 Frequency and Length via Fourier Transform

In order to estimate two parameters of the model n, d, each set of landmarks $\mathcal{L}_i$ is projected to an axis passing through the center of mass, projecting the set of 2D normal distribution functions into a 1D signal. This projected signal is illustrated in fig. 4b and defined in equation 5.

$$\begin{aligned} &signal_i = \sum_{lm_j \in \mathcal{L}_i} f_{lm_j}\left(\mu_j, \sigma_j\right) \\ &f_{lm_j}\left(\mu_j, \sigma_j\right) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \quad , \mu_j = \frac{\overrightarrow{v_1}.\overrightarrow{v_2}}{\|\overrightarrow{v_1}\|}, \sigma_j = \sigma_x \\ &\begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = R(\theta_j)\Sigma_j R^T(\theta_j) \end{aligned} \tag{5}$$

In equation 5 vectors $\overrightarrow{v_1}, \overrightarrow{v_2}$ are those illustrated in fig. 4b. $\overrightarrow{v_1}$ is the projection axis itself and $\overrightarrow{v_2}$ is a vector from starting point of $\overrightarrow{v_1}$ to the point subjected to projection.

Assuming a uniform distribution of points in each set implies that, Fourier transform of the projection signal ($\mathcal{F}(f)$) has a dominant frequency. This frequency relates to the values n, d as expressed in equations 6.

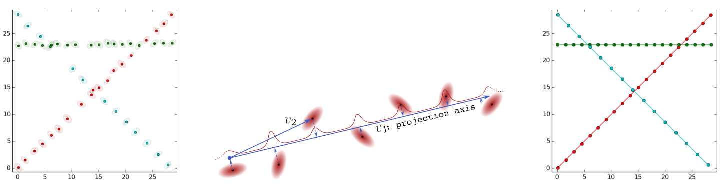
$$\begin{aligned} &n = f_0 \ , d = \frac{\|\overrightarrow{v_1}\|}{f_0} \\ &f_0 = \arg\max \mathcal{F}(f) := \{f_0 \mid \forall f : \mathcal{F}(f) \leq \mathcal{F}(f_0)\} \end{aligned} \tag{6}$$

2.3 Optimization and Model Fitting

Last parameters to uniquely define the model are (x, y). For that purpose, (x, y) are set to the pose (μ) of the first landmark in $\mathcal{L}$. Considering that the angle of alignment θ belongs to $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, first element of each $\mathcal{L}$ is the most left item, or the lowest in case of vertical lines. Then through an optimization process (x, y) are tuned. It should be noted that the first landmark in $\mathcal{L}$ is an initial guess for the optimization. This is based on the assumption that $\mathcal{L}$ does not consist of very far off outliers. Such outliers may bring the optimization into a local minima, causing a shift in model's position. Objective function of the optimization is given in equation 7. This function is a summation of all normal distribution functions of landmarks of the line $(\mathcal{N}_{ik} \in \mathcal{L}_i)$ operating on all the points of the model $(p_{il} \in M_i)$.

$$f_i(X) = \frac{1}{\sum\limits_{l}^{p_{il} \in M_i} \sum\limits_{k}^{\mathcal{N}_{ik} \in \mathcal{L}_i} \mathcal{N}_{ik}(p_{il})} \tag{7}$$

Since in the map of a real environment most of the models share 3 parameters (n, d, θ), for demonstrating the performance and generality of the method, it was applied to a synthesized data instead of a real map. Performance of the proposed method is evaluated over a map from real world in section 3. Unlike real environment the synthesized data in Figure 4a contains 3 $\mathcal{L}$ with all different parameters.



(a) clustered sets (b) 2D to 1D projection (c) models

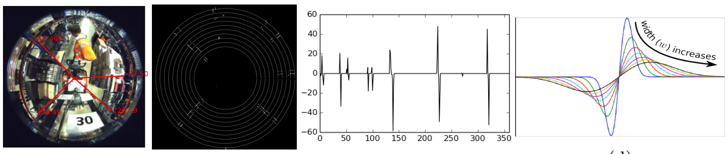
Fig. 4: Modeling 3 segmented line sets $(\mathcal{L})$ from 4a to 4c. 4b shows 2D to 1D projection of a single $\mathcal{L}$ according to equation 5 for frequency analysis

Result of clustering is encoded in colors. While in a real world map a crossing between $\mathcal{L}$ is not expected, in this synthesized data the lines are crossing to demonstrate the performance of the closed-form Hough transform. In addition to a uniform noise added to the position of each landmark, 10% of landmarks are removed to evaluate the result of Fourier transform in estimating n and d. As it is observed in fig. 4c, the model would fill in the position of missing landmarks. Finally the values of the parameters are successfully computed by the method, in fig. 4 $M_1\,(20, 2.12, -0.1, 0.0, 0.78)$ colored red, $M_2\,(20, 1.50, 0.0, 23.0, 0.00)$ colored green and $M_3\,(15, 2.80, 0.0, 28.5, -0.78)$ colored cyan. It should be mentioned that in case of a real map not all minor assumptions are met, so global constraints are introduced over all parameters. However none of the global constraints are preset, but all are extracted from the map. This aspect is explored further in section 3.1.

3 Experimental Results and Discussion

This section describes the procedure of creating and modeling a map of a real warehouse. For this purpose we use a "Panasonic" 185^o fish-eye lens mounted on "Prosilica GC2450" camera, installed on a AGV forklift truck in a warehouse. As the truck is provided with a localization system based on lasers and reflectors, we adopt an Extended Kalman Filter based bearing-only technique for mapping. Then the model developed in this work is employed to represent the map. To this end, a set of global estimations are extracted from the map and posed over the model's parameters as global constraints.

Pillar Detection Considering the common color coding of the pillars in the warehouses, pillar detection starts with segmentation through color indexing [11], followed by calculation of the oriented gradient. The camera is pointing downward and all the pillars are parallel to the camera axis. Consequently pillars are pointing to the vanishing point of the camera in the images, hence gradient vectors of pillars' edges are perpendicular to lines passing through the vanishing point. Any other gradient vectors are considered non-relevant and filtered out. The gradient image is sampled over multiple concentric circles as illustrated in fig. 5b. Results of all sampling are accumulated in a signal as in fig. 5c, capturing the pattern of a pillar's appearance by two opposite peaks representing the edges of the pillar. Detecting the position of such a pattern returns the bearing to pillars. A continuous wavelet transform (CWT) based peak detection technique [2] is adopted for detection of pillars' pattern in the gradient signal. The method is based on matching a pattern encoded in a wavelet, by the help of CWT. The pattern representing two sides of the pillar in the gradient signal (see fig. 5c), could be modeled with a wavelet based on the 1^{st} order derivative of a Gaussian function as illustrated in fig. 5d.



(a) (b) (c) (d)

Fig. 5: Detection of pillars. 5a) original image and the result of pillar detection, 5b) circular sampling of gradient image. 5c) accumulation of circular sampling as a signal, illustrating the occurance of pillars. A wavelet in 5d, resembling the pattern of pillars in gradient signal, is used for pillar detection by CWT technique.

Bearing Only Mapping Mapping is performed in an Extended Kalman Filter (EKF) fashion proposed in [1] for bearing-only SLAM. However, as the pose of the truck is provided through a laser-reflector based system, there is no need to localize the truck through EKF framework. Nevertheless, we include the state of the vehicle in the Kalman filter's state to ease the adaptation and system description. The truck's state is neither predicted nor updated in corresponding phases of the EKF, instead it comes directly from the localization system. This ensures that Kalman gain computation and prediction step are based on the "true" pose of the truck, and leaves the possibility to extend this implementation into SLAM if required.

A map generated through this framework, from the data logged in a real world warehouse is sketched in fig. 6. Colors in this map code the result of clustering of the line sets ($\mathcal{L}$), and blacks are outliers.

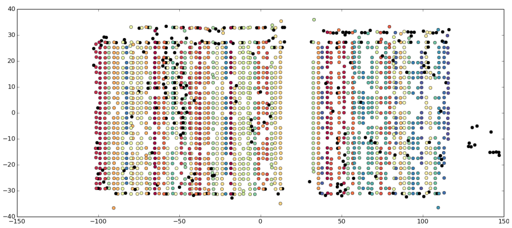


Fig. 6: Mapping 77 lines of pillars of a warehouse of size $250^m \times 72^m$.

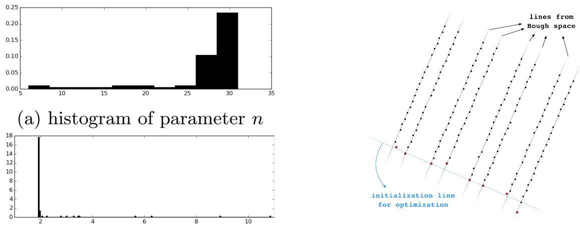
3.1 Global Constraints on the Models

Recalling the assumption of a structured environment, the set of models representing the boundaries of corridors are parallel, and have similar parameters n, d, θ. Even the starting points of the models are aligned on a straight line (like in fig. 6). We exploit this information to pose a set of global constraints, which results in a globally more consistent set of models of the environment.

Dominant Orientation refers to the fact that corridors in an environment have the same direction. This suggest that we could pose a global constraint on θ. This has a crucial importance for the clustering step as well, as the number of intersection points in Hough space (see fig. 3b) increases quadratically by the number of landmarks ($\frac{n(n-1)}{2}$). This means for a map consisting of more than 2000 landmarks like fig. 6, there will be more than 2 million intersection points in Hough space. Such a number of points can not be easily clustered. However, considering the assumption of dominant orientation, it is possible to reject a huge number of intersection points which are not within an acceptable range (5^o) from dominant orientation. This filtering process is handled by discretization of the Hough space and finding the dominant orientations, where most of the peaks occur. After the filtering process, the clustering proceeds as explained earlier in closed-form.

Dominant Frequency Similarly a constraint is posed over parameters n and d. This is also very helpful since in a real map, sometimes landmarks are not visible and hence not included in the map. If the percentage of missing landmarks become too big, it may yield a wrong estimation of the frequency. In order to estimate global constraint of these two parameters, first n and d are computed for all lines ($\mathcal{L}$) without constraint. Then the mode in the histogram of these parameters as shown in fig. 7 provides a global estimation of the parameters. After acquiring the global values from the map, models are generated again, this time with the global constraints.

Initialization Line Assumption of a structured environment implies that the starting points of all $\mathcal{L}$ are located on the same line. This line is then used for initial guess in the optimization process. The initialization line is calculated by a linear regression among first landmarks in all $\mathcal{L}$, shown in fig. 7c. While each $\mathcal{L}$ comes from Hough space with its line equation, the exact position of the optimization's initial guess for each $\mathcal{L}$ is given by the intersection point of mentioned line and initialization line.



(a) histogram of parameter n

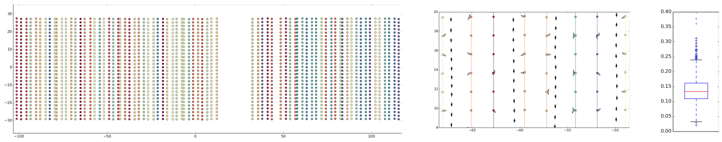
(b) histogram of parameter d

(c) initialization line

Fig. 7: Global constraints; Histograms of (n, d), mode in each shows the dominancy. Initialization line is a linear regression of first landmarks in all $\mathcal{L}$ (red points).

3.2 Discussion

Engaging all the global constraints, result of modeling the map given in fig. 6 is illustrated in fig. 8a. The final result of modeling is useful for geometric mapping of the boundaries of corridors, as well as semantic annotation of pallet rack cell on the side of corridors. A crop of the model augmented with truck's pose and reflectors is presented in fig. 8b. The reflectors are those pre-installed in the environment used for AGVs' localization. We use the position of reflectors as ground truth for accuracy estimation. A box plot in fig. 8c demonstrates the distances between 1143 reflectors and nearest point in the model. Taking into account the width of pillar's and reflector's mount, the distances are representing the errors of the model with an offset of about 1 decimeter.



(a) (b) (c)

Fig. 8: Modeling a warehouse (see fig. 6) with geometric semantic models (all axes' unit in meter). 8b) shows a crop of model augmented with the truck's pose(black arrow) and reflectors (white triangles). 8c) shows distance between models and closest reflectors, with an offset about 1^{dm}.

It must be remembered, although the global constraints improve the global consistency of the models, yet it is only applicable if the infrastructures share similar geometry, that is to say the resulting models have the same parameters. In case of an environment that consists of multiple *regions* with different characteristics, those regions must be subjected to regional constraints.

4 Conclusion

Toward developing an awareness framework for effective management of logistics and inventory in a warehouse environment, a rich geometric-semantic map is set as a goal. In this work we address this objective on the level of infrastructure modeling. We suggest a choice of landmark for bearing-only mapping of a warehouse environment, leading to a map which integrates the infrastructure of the warehouse. This paper presents a canonical geometric-semantic model, along with a method for generating and matching these models into the map. The model is a set of points aligned on a straight line uniquely identified by 5 parameters. Geometrically it represents the boundaries of corridors, and semantically represents pallet rack cells. A series of tools such as closed-form Hough transform, DBSCAN clustering algorithm, Fourier transform and optimization techniques are employed to estimate the parameters of each model and match them with the map. The performance of the method is demonstrated over both synthesized data and real world map.

Proposed model and extracting method have interesting characteristics, such as modularity, generality, and representing both geometric and semantic knowledge. Most of the steps and parameters of the model are independent, therefore it is adjustable according to different scenarios. One example of such adjustment was given as global constraint in this work, where all of the models' parameters adopt to the global characteristic of the map for a better consistency. However the method has its own limits as discussed in 3.2, such as sensitivity to initial guess of the optimization, and consideration of global constraints for consistency.

We plan to develop this work further by fusing other sensory data into this map, such as introducing occupancy notation from range scanners. And also by developing a more dynamic model, allowing harmonics of the dominant frequency to participate, and handling those cases where landmarks are not uniformly distributed. Segmentation of map into regions based on dominant orientations would be helpful, if the environment consists of multiple orientations. Indeed we investigated this aspect in another work based on occupancy maps, and will introduce it to this method as soon as the occupancy notation is fused into this map.

References

1. Bailey, T.: Constrained initialisation for bearing-only slam. In: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on. vol. 2, pp. 1966–1971. IEEE (2003)
2. Du, P., Kibbe, W.A., Lin, S.M.: Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 22(17), 2059–2065 (2006)
3. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. Communications of the ACM 15(1), 11–15 (1972)
4. Ester, M., peter Kriegel, H., S, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. pp. 226–231. AAAI Press (1996)
5. Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernandez-Madrigal, J.A., González, J.: Multi-hierarchical semantic maps for mobile robotics. In: Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on. pp. 2278–2283. IEEE (2005)
6. Liu, Z., von Wichert, G.: Extracting semantic indoor maps from occupancy grids. Robotics and Autonomous Systems (2013)
7. Park, S., Kim, S., Park, M., Park, S.K.: Vision-based global localization for mobile robots with hybrid maps of objects and spatial layouts. Information Sciences 179(24), 4174–4198 (2009)
8. Persson, M., Duckett, T., Valgren, C., Lilienthal, A.: Probabilistic semantic mapping with a virtual sensor for building/nature detection. In: Computational Intelligence in Robotics and Automation, 2007. CIRA 2007. International Symposium on. pp. 236–242. IEEE (2007)
9. Pronobis, A., Jensfelt, P.: Large-scale semantic mapping and reasoning with heterogeneous modalities. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. pp. 3515–3522. IEEE (2012)
10. de la Puente, P., Rodriguez-Losada, D.: Feature based graph-slam in structured environments. Autonomous Robots pp. 1–18 (2014)
11. Swain, M.J., Ballard, D.H.: Color indexing. International journal of computer vision 7(1), 11–32 (1991)
12. Wolf, D.F., Sukhatme, G.: Semantic mapping using mobile robots. Robotics, IEEE Transactions on 24(2), 245–258 (2008)
13. Wolter, D., Latecki, L.J., Lakämper, R., Sun, X.: Shape-based robot mapping. In: KI 2004: Advances in Artificial Intelligence, pp. 439–452. Springer (2004)
14. Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G.J., Burgard, W.: Conceptual spatial representations for indoor mobile robots. Robotics and Autonomous Systems 56(6), 493–502 (2008)

Paper B - Sensor Based Adaptive Metric-Topological Cell Decomposition Method for Semantic Annotation of Structured Environments.

Sensor Based Adaptive Metric-Topological Cell Decomposition Method for Semantic Annotation of Structured Environments

Saeed Gholami Shahbandi
saesha@hh.se
Halmstad University, Sweden
Box 823, 30118 Halmstad

Björn Åstrand
bjorn.astrand@hh.se
Halmstad University, Sweden
Box 823, 30118 Halmstad

Roland Philippsen
roland.philippsen@hh.se
Halmstad University, Sweden
Box 823, 30118 Halmstad

***Abstract*—A fundamental ingredient for semantic labeling is a reliable method for determining and representing the relevant spatial features of an environment. We address this challenge for planar metric-topological maps based on occupancy grids. Our method detects arbitrary dominant orientations in the presence of significant clutter, fits corresponding line features with tunable resolution, and extracts topological information by polygonal cell decomposition. Real-world case studies taken from the target application domain (autonomous forklift trucks in warehouses) demonstrate the performance and robustness of our method, while results from a preliminary algorithm to extract corridors, and junctions, demonstrate its expressiveness. Contribution of this work starts with the formulation of metric-topological surveying of environment, and a generic n-direction planar representation accompanied with a general method for extracting it from occupancy map. The implementation also includes some semantic labels specific to warehouse like environments.**

I. Introduction

The state of the art in autonomous robotics has advanced sufficiently that open implementations of many core technologies are now readily available. Consequently, there is growing research on the design and development of innovative solutions that leverage insights from several specialist domains. The Automatic Inventory and Mapping of Stock (AIMS) project lies in this category targeting the traditional warehouses where no infrastructure is designed or installed for automation. Its goal is to develop a system that seamlessly combines inventory management with autonomous forklift trucks in intelligent warehouses (Fig. 1). Information compatible with human operators, management systems, as well as mobile robots is of particular importance here. A rich and "live" map combining metric and semantic information is a crucial ingredient for effective management of logistics and inventory, especially for autonomous fleets working in the same space as humans and human-operated devices.

This paper focuses on a subsystem developed for AIMS, namely a robust method to automatically analyze an occupancy grid map.

This work was supported by the Swedish Knowledge Foundation and industry partners Kollmorgen, Optronic, and Toyota Material Handling Europe. It was carried out at the Center for Applied Intelligent Systems Research at Halmstad University, Sweden.

It produces a "subdivision" representation, whose elements correspond to a variety of spatial resolutions and orientations that are the most relevant for a given environment and application context. This representation serves the task of surveying a new environment, as input to a first stage of semantic annotation based on spatial features for further analysis. The resulting representation is expected to later feed into CAD softwares for setting up and operating Auto Guided Vehicles (AGV) in warehouses, as well as interact with Warehouse Management Systems (WMS).

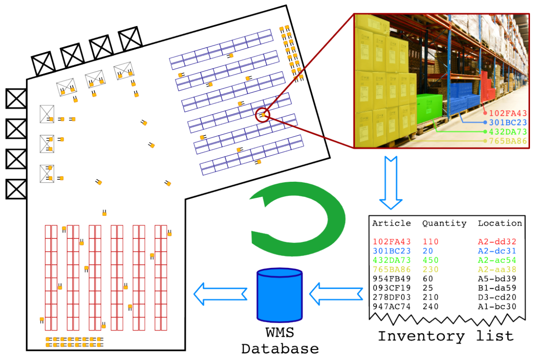


Fig. 1: The AIMS project establishes a set of tools for setting up and operating AGVs such as forklift trucks. It includes a range of aspects, from surveying of warehouses to on-line inventory tracking.

The environment is assumed to have a given number of predominant directions, but the angles between them are not constrained. We also assume that the spatial scale of the semantic features to be detected in the subsequent step is approximately known. To illustrate these points and demonstrate the suitability of our method, we also describe a preliminary algorithm to detect corridors and junctions. Experiments based on two real-world data sets from different sources show the versatility and robustness of our method.

Related Work

There is a significant body of work on semantic analysis of the kinds of environments relevant for AIMS. We discuss only the ones most pertinent to this paper. *Local approaches* base semantic information either on object labeling or the identification of local shapes. For instance, classifiers can be used to detect regions in range data [1]. Image features, and similarly range features, can also be used [2]. To label segments in 3D point clouds, visual appearance, shape, geometry, and geometrical context can be used [3]. *Global approaches* on the other hand derive semantic information via place labeling. For instance, [4] builds a topological map based on the notion of connectivity and adjacency, [5] explores an occupancy grid map through the Voronoi graph, [6] bases a topological map on the connectivity between acquired images, [7] provides a topological graph of the environment based on the concept of virtual door, and [8] employs a series of kernels based on Markov chain Monte Carlo for semantic labeling.

For our purposes, a hybrid metric-semantic approach is most appropriate. Our formulation relies heavily on topological aspects, while grounding the representations in metric sensor-based information, and leveraging both in the abstraction and labeling algorithms.

Contributions and Approach

The common reliance on line extraction [9], [8], [10], [11], [12] indicates the importance of orientation in semantic mapping. However, we cannot readily reuse these results, due to clutter and discontinuities in the physical structures. For instance, the walls of warehouse corridors consist of irregularly stored articles on shelves built with pillars whose horizontal extents are significantly smaller than the goods containers.

Thus, a particular challenge that our method addresses is the discrepancy between the importance of orientation information and the absence of corresponding clean long lines in the range data that can be obtained via laser scanners in warehouses. We achieve this with a combination of *Histogram of Oriented Gradients* [13] and a *Radiogram* [14] operating on an occupancy map. The radiogram is similar to a *Radon Transform* [15], operating on specific angles and associated with a filtering mechanism based on oriented gradients.

The histogram of oriented gradients (Fig. 2) allows to extract a (given number of) dominant directions in the environment. Radiography is then applied to occupied cells in direction of dominant orientations, and the resulting peaks correspond to feature locations in the map. The real-world environments shown in this paper exhibit two perpendicular dominant orientations, but our method is by no means limited to these cases. Employing radiography based technique for feature extraction requires the resolution of the map to be sufficiently high to represent each side of the physical entities with enough observations. The spatial arrangement of features is extracted via *polygonal cell decomposition*, and captured in a *subdivision* and an *adjacency graph* data structure to enable a higher-level topological analysis.

We collect the information on orientations, locations, the subdivision and adjacency graph data structures in our *adaptive* method. To clarify terminology, adaptivity here is with respect to the geometric structure of the environment while preserving the metric information. This is distinct from resolution adaptivity such as found in quadtrees or octrees [16], [17].

Our novel representation characterizes the geometric structure of the environment in addition to providing an abstract metric representation of the map.

Note that this paper does not fully address subsequent semantic analysis, but focuses on an adaptive generic representation to support computing and storing of such info, and one specific implementation of the method for environments with two dominant directions. Developing a more complete and versatile annotation method and representation is part of ongoing and future research.

II. Method

Performing a survey by means of an occupancy grid map demands an understanding of the general structure of the map. We rely on the assumption that there exist physical elements approximately arranged in straight lines according to an arbitrary limited set of cominant directions. Our method consists of the following main steps.

1) Detect dominant directions via histograms of orientated gradients in the occupancy map (Fig. 2, section II-A).
2) Extract line features via radiography along the dominant directions, using wavelet-based peak detection to robustly influence the resulting resolution (Fig. 3a, section II-B).
3) Employ polygonal cell decomposition to compute a planar subdivision and an adjacency graph (Fig. 3b, section II-C).
4) Refine the adjacency abstraction by analyzing occupancy inside faces and along edges (Fig. 5, section II-D).
5) Combine metric and topological information to infer semantic labels using template matching (Fig.8, section II-E).

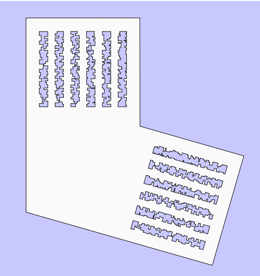
In this paper, we focus on producing the subdivision, which captures the structure of physical entities in the environment, as well as the adjacency graph, which encodes the topology (connectedness of open space). Inference is sketched for the kind of environment of current relevance to practical advances in the AIMS project, looking at two usage scenarios: aiding in setting up AGV operation, and online path planning.

The method is demonstrated first on a map of the *Intel Jones Farms Campus*[1] and later applied to data collected in the warehouse of one of our industrial partners. Performance on the warehouse map is discussed in more detail in section III.

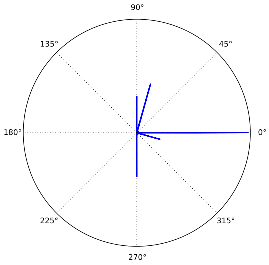
A. Dominant Orientations Detection

The method starts by detecting the major orientations of lines in the map. Recalling the assumption of straight line in the environment, some specific orientations are expected

[1]This data set was obtained from the Robotics Data Set Repository (Radish) [18]. We thank *Maxim Batalin* for providing this data.

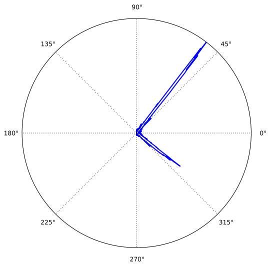


(a) synthesized map



(b) 4 major orientations

(c) real-world map



(d) 2 major orientations

Fig. 2: Dominant orientations detection from Histogram of Oriented Gradients (HOG)

to appear more frequently in the oriented gradients, caused by long lines or collinear line segments in the map. These unidirectional features create peaks in the histograms of orientations as illustrated in Fig. 2. Due to symmetry, the angles are confined to the interval $[-\pi, \pi]$.

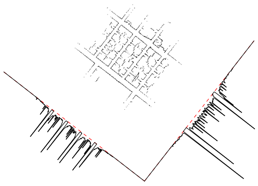
Histograms shown in Fig. 2 are weighted by the magnitudes of oriented gradients, therefore the strength of each gradient vector votes for dominant orientations, too. With an admissible assumption that the number of dominant directions is known, finding them from the histogram is straightforward. Dominant orientations are represented by the set Θ.

$$\Theta := \left\{\theta_i \mid i \in \mathbb{N},\ \theta_i \in [-\frac{\pi}{2}, \frac{\pi}{2}]\right\}$$

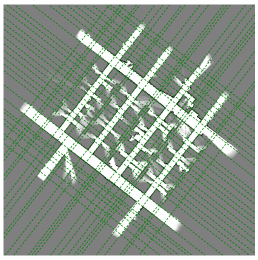
While many environments of practical interest consist of two dominant perpendicular orientations, Fig. 2 shows the ability of this technique in handling more general scenarios.

B. Radiogram and Peak Detection

The set of dominant orientations (Θ) entails the directions of long or most frequently aligned elements in the map. Projecting occupied cells of the occupancy map along the direction of $\theta_i \in \Theta$, generates signals referred to as radiograms [14]. This is demonstrated in Fig. 3a. This step is equivalent to computing a Radon Transforms [15] of the image in direction of the dominant orientations. Radiography as implemented in this work employs a filtering mechanism based on the orientation



(a) radiography of edge points



(b) cell decomposition

Fig. 3: Radiography of edge points in direction of dominant orientations. Resuting signal is used for line detection and cell decomposition of the map.

of points, rejecting if their gradients are not perpendicular to projection axes.

A set of line equations ($\mathcal{L}_{ines}$) constructs a low level spatial representation of the environment's structure.

$$\mathcal{L}_{ines} := \{y = m_i \times x + b_i \mid i \in \mathbb{N},\ x, y, m_i, b_i \in \mathbb{R}\}$$

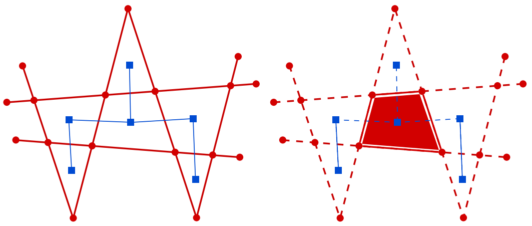
As the slope m_i is already encoded in the dominant orientations Θ, only the offset b_i needs to be estimated. This is done via the radiogram, where peaks correspond to lines in the occupancy map. The peaks have widths that are characterised by the appearance of the entities that constitute a particular line, and this can be leveraged to influence the resolution of the resulting set of extracted lines.

Du et al. developed a technique based on continuous wavelet transform for detecting peaks in noisy signals [19], with a wavelet resembling the shape of target peaks. In this case a Ricker wavelet has been used to perform the continuous wavelet transform over the signal. Including a range of wavelets with different widths (σ) leads to robust peak detection. This technique returns only the peaks that match the characteristic of the provided wavelet restricted by the range of width. In practice, we found that $\sigma \in [1, 20]$ is a sufficient range of wavelet's width scale for detecting peaks.

Magnitude of the peaks represent the length of the lines, interpreted as the importance of that particular set of elements. The magnitude values of these peaks can be used for adjusting the resolution of the cell decomposition in later stage, and thus the level of abstraction. It's possible to coarsen subdivision by rejecting less important peaks. This aspect is discussed more in section III.

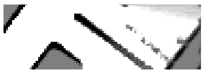
C. Cell Decomposition

Given the equation of lines by the set $\mathcal{L}_{ines}$ which represents structure of the environment, a polygonal cell decomposition ables us to extract higher level informations from the map. Cell decomposition in this work has a similar implementation as the polygonal decomposition explained by *Kloetzer* and *Ghita* [20]. This decomposition results in polygon faces (cells) with different shapes and sizes and is stored in a *subdivision*



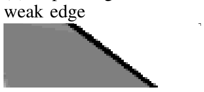
(a) duality (b) validity

Fig. 4: Illustration of subdivision and adjacency graph, duality and validity. Thick red lines and red dots are edges and vertices of the subdivision. Thin blue lines and blue squares are links and nodes from adjacency graph. Solid lines represent valid edges and links, while dashed lines encode invalidity.

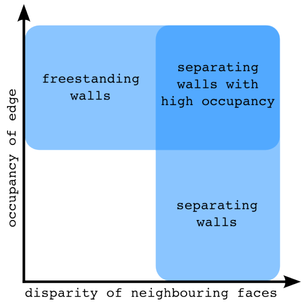


(a) freestanding wall

(b) separating wall with weak edge



(c) separating wall with strong edge



(d) Feature space of edges from subdivision

Fig. 5: Feature space of edges from subdivision. This space illustrates possible circumstances where an edge is identified as a wall.

(red in Fig. 4a) and an *adjacency graph* (blue in Fig. 4a). Subdivision is composed of 3 sets of $\mathcal{V}_{ertices}$, $\mathcal{E}_{dges}$ and $\mathcal{F}_{aces}$. 4 functions are defined over those sets, namely neighbors of edge (N), cogency of edge (C) and occupancy of edge and face (O). Adjacency graph has 2 functions, neighbors of node (N) and connectivity of link (C) operating on the two sets of $\mathcal{N}_{odes}$ and $\mathcal{L}_{inks}$.

Subdivision Data Structure: Data structure S is defined as:

$$\begin{aligned} &\mathrm{S} = (\mathcal{V}_{ertices}, \mathcal{E}_{dges}, \mathcal{F}_{aces}) = (\{v_i\}, \{e_i\}, \{f_i\}) \\ &v_i = (x_i, y_i),\ e_i = (v_{il}, v_{ik}),\ f_i = \{e_{ij}\} \end{aligned}$$

The set $\mathcal{V}_{ertices}$ lists intersection points between lines from $\mathcal{L}_{ines}$ as vertices. Each edge in $\mathcal{E}_{dges}$ represents line segments between two consecutive vertices from same line ($l_i \in \mathcal{L}_{ines}$). Each face f_i in the set $\mathcal{F}_{aces}$ is itself a set of edges $\{e_{ij}\}$ that bound f_i. Combination of faces is not considered as a face.

Neighbor function of an edge $\mathrm{N}(e_i)$ returns those two faces neighboring through the e_i. Occupancy function of a face $\mathrm{O}(f_i)$ returns the average occupancy of all pixels bounded by the edges of the f_i. Occupancy function of an edge $\mathrm{O}(e_i)$, returns the average occupancy of all pixels corresponding to the e_i.

$$\begin{aligned} &\mathrm{N}(e_i) = (f_{il}, f_{ik}),\ (e_i \in f_{il}\ \wedge\ e_i \in f_{ik}) \\ &\mathrm{O}(e_i) = \text{average occupancy of all cells relative to } e_i \\ &\mathrm{O}(f_i) = \text{average occupancy of all cells bounded in } f_i \end{aligned}$$

Adjacency Graph Data Structure: Adjacency graph is a complementary data structure, proposed to formulate the connectivity between faces in the subdivision, and represents the metric-topological structure of the environment and defined as:

$$\begin{aligned} &\mathrm{A} = (\mathcal{N}_{odes}, \mathcal{L}_{inks}) = (\{n_i\}, \{l_i\}) \\ &n_i = (x_i, y_i),\ l_i = (n_{i1}, n_{i2}) \end{aligned}$$

Where each node n_i is located at the center of a face f_i. A neighborhood function (N) returning all the nodes connecting to n_i through a single link, is defined as:

$$\mathrm{N}(n_i) = \{n_j \mid l_k \in \mathcal{L}_{inks},\ l_k = (n_i, n_j)\}$$

Duality of subdivision and adjacency graph: Each edge in subdivision is neighboring two faces, which correspond to a link and two nodes from adjacency graph. Nodes in adjacency graph represents faces in subdivision and the link between two nodes encodes the connectivity between them. This duality is represented by the dual function D, where:

$$\mathrm{D}(n) \in \{f\},\ \mathrm{D}(f) \in \{n\},\ \mathrm{D}(l) \in \{e\},\ \mathrm{D}(e) \in \{l\}$$

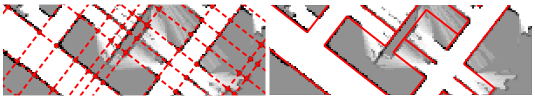
Edges in subdivision have a property called *cogency*, which means there is a physical separation (e.g. wall) between neighboring faces. Dual to edges' cogency, links have a property of *connectivity* indicating the connectivity between nodes (i.e faces). A "cogent" edge in subdivision is dual with a "disconnected" link in adjacency graph. In Fig 4b, assuming only the middle face is occupied, valid edges and links are represented by solid lines, where invalid edges and links are visualized via dashed lines.

D. Subdivision and Abstraction

Edges in subdivision could be candidates of being physical elements like walls in the map. It should be noted that here a wall is any entity perceivable by laser scanner as a straight line segment. Through the process of cogency evaluation, candidate edges are either accepted or rejected as walls. This process will lead to an abstract representation of the map. Cogency of an edge $\mathrm{C}(e_i)$ is evaluated upon two features, *face disparity* and *edge occupancy* $\mathrm{O}(e_i)$ (visualized in Fig. 5d) and is defined:

$$\begin{aligned} &\text{face disparity} = |\mathrm{O}(f_1) - \mathrm{O}(f_2)| \\ &\mathrm{C}(e_i) = (\text{face disparity} > t_1) \vee (\mathrm{O}(e_i) > t_2) \end{aligned}$$

where $(f_1, f_2) = \mathrm{N}(e_i)$ and t_1, t_2 are thresholds. If the disparity of two neighboring faces is high, it means one being occupied and the other being free space. This implies there must be a physical element separating these two faces, hence the edge is representing a *separating walls* as in Fig. 5. Sometime there are walls in the middle of free space of the environment with both sides known to be free, where



(a) subdivision (b) abstraction

Fig. 6: Subdivision encodes informations for abstraction.

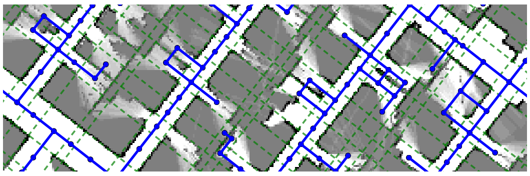


Fig. 7: The subdivision in green, and adjacency graph (only valid links) in blue. Nodes of adjacency graph are located at the center of faces and the links in this graph represent connectivity between faces.

disparity of faces would be low. For this scenario the feature of *edge occupancy* is taken into account (*freestanding walls* in Fig. 5). Fig. 6 demonstrates a portion of the *subdivision* and the resulting line abstraction with separating as well as freestanding walls.

E. Adjacency Graph and Semantic Labeling

As mentioned earlier nodes and links in adjacency graph are dual to faces and edge from subdivision. When a cogent edge represents an isolation between neighboring faces, connectivity of the dual link is defined as the complement of cogency.

$$\mathrm{C}(l_i) = \neg\mathrm{C}\left(\mathrm{D}(e_i)\right)$$

Fig. 7 illustrates that the valid links in the adjacency graph correspond to a connectivity map of the open space.

Metric attributes of the adjacency graph make it suitable for matching semantic templates that are metric-topological. Algorithm 1 presents an example for matching different types of junction patterns. In this implementation, $\mathcal{T}$ models two types of junctions, T-type and X-type, as illustrated in Fig. 8. It should be noticed that neither the templates in Fig. 8 nor Algorithm 1 presents a comprehensive set of semantic labels and matching algorithms. Corridors as illustrated in Fig. 8c are detected with a similar technique. The result of semantic analysis is shown in Fig. 8d.

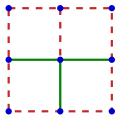
III. Results and Discussion

The method proposed in this work is tested on data from a real-world warehouse of bathroom accessories. The occupancy map (in Fig. 9a) was built using an open implementation of a SLAM algorithm based on *Rao-Blackwellized particle filter* called "GMapping" [21]. Performance of the method is demonstrated through semantic labeling and map abstraction (Figures 9e and 9d).

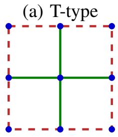
Algorithm 1 template matching in adjacency graph: example of "junction" detection

INPUT: adjacency graph
for $n_i \in \mathbb{N}_{\text{odes}}$ **do**
  $valid\_junction \leftarrow TRUE$
  for all $(n_l, n_k) \in \mathcal{N}(n_i),\ l \neq k$ **do**
    if $\mathcal{N}(n_l) \wedge \mathcal{N}(n_k) \neq \{n_i\}$ **then**
      $valid\_junction \leftarrow FALSE$
    end if
  end for
  $n_i = \begin{cases} \mathcal{T}\left(size\left(\mathcal{N}(n_i)\right)\right) & if\ valid\_junction \\ \emptyset & otherwise \end{cases}$
end for

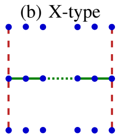
$$\mathcal{T}(n) := \begin{cases} \text{T type} & n = 3 \\ \text{X type} & n = 4 \\ \emptyset & otherwise \end{cases}$$



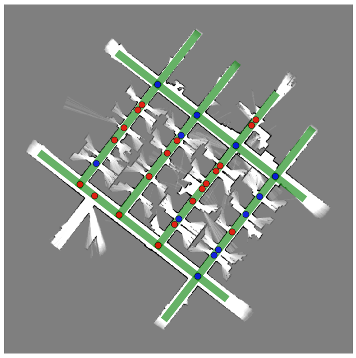
(a) T-type



(b) X-type



(c) Corridor



(d) semantic labeling in a real-world map, corridors in green, X-type junctions in blue and T-types in red.

Fig. 8: Some templates for pattern detection from adjacency graph are presented along with result of applying to a real-world map. Green lines in templates are valid connection between nodes, on the contrary red lines imply *blocked access*.

Choice of Resolution: Degree of the cell decomposition's resolution comes from spatial characteristic of the physical entities in the environment. If the elements of the environment are distributed homogeneously with similar length, they would provoke similar peaks in radiogram signal. In reality peaks have different amplitude and width, which represent how many and how well things line up in the environment over this particular line in the model ($\mathcal{L}_{ines}$). Tuning the sensitivity of peak detection technique with these two values (amplitude and width) will result in different resolutions in the cell decomposition. Figures 9b and 9c show how the change of resolution affects the decomposition. The optimum resolution depends on the application. For instance a high resolution decomposition preserving more details from the environment would be suitable for path planning (Fig. 9e).

Coarser resolution appear to be more appropriate for semantic labeling of regions. For example in Fig. 9d, corridors are marked in green, X-junctions with blue circle, T-junctions with red circles, blue lines show spatial connectivity, and red lines are the map abstractions.

Deficiency of Semantic Labels: Each label is the result of a possible match between some predefined templates and the adjacency graph. As it is noticeable in Fig. 8d and Fig. 9d, there exist few misclassifications of X-type and T-type junctions. This imperfection is rooted in the degeneracy of adjacency graph. A map of partially explored environment leads to some degeneracies in the adjacency graph, such as stub-like links or missing links at the frontier of unexplored areas. While these degeneracies could be minimized with a fitting choice of resolution, this would not fully eliminate them. Current work targets this shortcoming via improved graph matching and richer templates, which is beyond the scope of this paper.

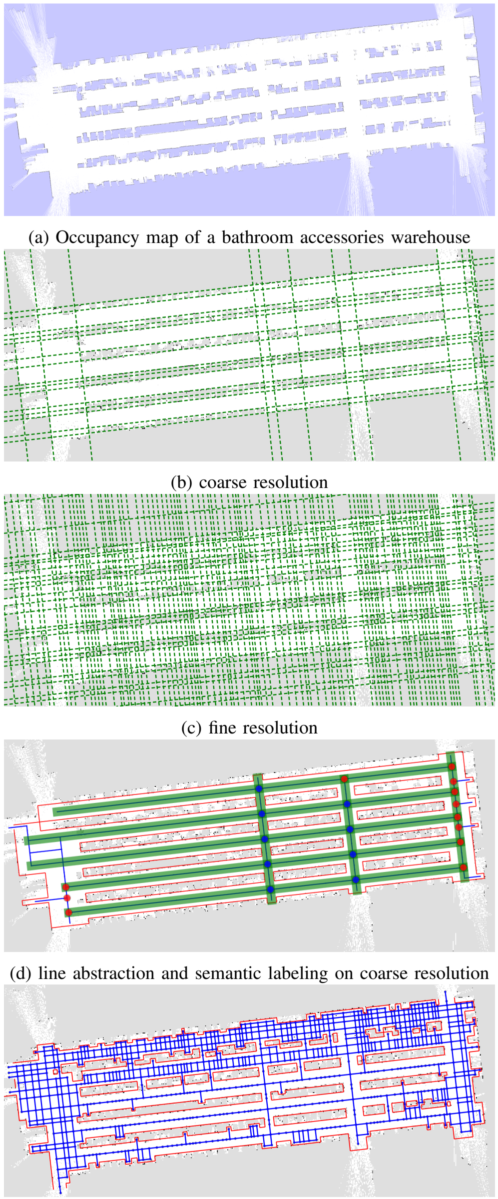
IV. Conclusion and Future Work

Robots with a heightened awareness of their surroundings are key for novel applications in automation, such as intelligent warehouses with integrative inventory management. In addition to increasing the reliability and robustness of operation, it can provide crucial background knowledge for adaptive learning or reasoning in novel situations. The presented work provides a concrete method for increasing the level of awareness beyond the state of the art.

The proposed method interprets occupancy maps to bring out underlying spatial characteristics in a format readily usable by machines and humans. The problem is approached from a global perspective by extracting dominant orientations, assuming the presence of straight lines that are important yet hard to detect. We employ radiogram analysis to fit independently oriented and spaced lines to the occupancy map. After cell decomposition over those fitted lines, two corresponding data structures capture topological characteristics and ease subsequent semantic labeling.

We present a case study of two real-world maps, one from a commonly available dataset and the other from data we collected in the warehouse of an industrial partner, to demonstrate the effectiveness of the method. Furthermore we show that a preliminary semantic labeling algorithm reliably detects corridors, and junctions.

The key features of the proposed method are the flexible and robust direction detection and the adaptive spacing of lines. At this stage, the number of orientations is given in advance. Clustering techniques could be used to generalize this aspect. No restriction is placed on the relative angle between dominant orientations. Control of the spatial resolution at which grid lines should be inferred (scale of the Ricker wavelet) influences the abstraction level of the resulting topological information. This is currently a (rather straightforward) manual process, inferred from the type of environments subject to our experiments.



(a) Occupancy map of a bathroom accessories warehouse

(b) coarse resolution

(c) fine resolution

(d) line abstraction and semantic labeling on coarse resolution

(e) line abstraction and connectivity map on fine resolution

Fig. 9: Cell decomposition of a real-world warehouse of bathroom accessories, with coarse and fine resolution. While coarse decomposition is more suitable for semantic labeling and extracting the layout of the environment, fine decomposition is more appropriate for detailed abstraction and connectivity maps.

Improvement of the proposed method will include a study of spatial frequency to better understand and detect features (such as pillars) that occur with some underlying regular pattern. This is expected to help with low-quality maps and enable automatic adaptation of the spatial resolution. An important next milestone in the project is the inclusion of visual perception. In [22] we proposed a modeling technique for structured environment by the means of vision sensors. That study starts with mapping a warehouse by its infrastructural landmarks (pillars). It is followed by modeling the map with a "repetitive canonical geometric-semantic" model. While the model provides an abstraction of the geometry of the environment by its landmarks and corridor boundaries, it also implies semantics on a conceptual levels. Next goal in our research would be to develop a combination technique, that is able to associate maps and their inferred knowledge from different modalities.

References

[1] O. M. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using adaboost," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 1730–1735.

[2] A. Rottmann, O. M. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *proceedings of the national conference on artificial intelligence*, vol. 20, no. 3. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005, p. 1306.

[3] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, "Semantic labeling of 3d point clouds for indoor scenes," *Proceedings of the Advances in Neural Information Processing Systems*, 2011.

[4] E. Fabrizi and A. Saffiotti, "Extracting topology-based maps from gridmaps," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 3. IEEE, 2000, pp. 2972–2978.

[5] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: Extracting the topological structure of indoor environments via place labeling," in *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 35, 2007.

[6] Z. Zivkovic, O. Booij, and B. Kröse, "From images to rooms," *Robotics and Autonomous Systems*, vol. 55, no. 5, pp. 411–418, 2007.

[7] K. Joo, T.-K. Lee, S. Baek, and S.-Y. Oh, "Generating topological map from occupancy grid-map using virtual door detection," in *Evolutionary Computation (CEC), 2010 IEEE Congress on*. IEEE, 2010, pp. 1–6.

[8] Z. Liu and G. von Wichert, "Extracting semantic indoor maps from occupancy grids," *Robotics and Autonomous Systems*, 2013.

[9] S. An, J. Kang, L. Lee, and S. Oh, "SLAM with salient line feature extraction in indoor environments," in *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on*. IEEE, 2010, pp. 410–416.

[10] K. Arras and R. Siegwart, "Feature extraction and scene interpretation for map-based navigation and map building," in *Proceedings of SPIE, Mobile Robotics XII*, vol. 3210, 1997, pp. 42–53.

[11] S. T. Pfister, S. I. Roumeliotis, and J. W. Burdick, "Weighted line fitting algorithms for mobile robot map building and efficient data representation," in *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, vol. 1. IEEE, 2003, pp. 1304–1311.

[12] A. Garulli, A. Giannitrapani, A. Rossi, and A. Vicino, "Mobile robot SLAM for line-based environment representation," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*. IEEE, 2005, pp. 2041–2046.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[14] J. Bigun, S. Bhattacharjee, and S. Michel, "Orientation radiograms for image retrieval: an alternative to segmentation," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 3. IEEE, 1996, pp. 346–350.

[15] S. R. Deans, *The Radon transform and some of its applications*. Courier Dover Publications, 2007.

[16] E. Einhorn, C. Schroter, and H.-M. Gross, "Finding the adequate resolution for grid mapping-cell sizes locally adapting on-the-fly," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1843–1848.

[17] G. K. Kraetzschmar, G. P. Gassull, and K. Uhl, "Probabilistic quadtrees for variable-resolution mapping of large environments," in *Proceedings of the 5th IFAC/EURON symposium on intelligent autonomous vehicles*. Citeseer, 2004.

[18] A. Howard and N. Roy, "The Robotics Data Set Repository (Radish)," 2003. [Online]. Available: http://radish.sourceforge.net

[19] P. Du, W. A. Kibbe, and S. M. Lin, "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching," *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006.

[20] M. Kloetzer and N. Ghita, "Software tool for constructing cell decompositions," in *Automation Science and Engineering (CASE), 2011 IEEE Conference on*, Aug 2011, pp. 507–512.

[21] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *Robotics, IEEE Transactions on*, vol. 23, no. 1, pp. 34–46, 2007.

[22] S. G. Shahbandi and B. Å strand, "Modeling of a large structured environment," in *Advances in Autonomous Robotics Systems*, ser. Lecture Notes in Computer Science, M. Mistry, A. Leonardis, M. Witkowski, and C. Melhuish, Eds. Springer International Publishing, 2014, vol. 8717, pp. 1–12. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10401-0_1

Paper C - Semi-Supervised Semantic Labeling of Adaptive Cell Decomposition Maps in Well-Structured Environments.

Semi-Supervised Semantic Labeling of Adaptive Cell Decomposition Maps in Well-Structured Environments

Saeed Gholami Shahbandi
Halmstad University, Sweden
Box 823 - 30118 Halmstad
Email: saesha@hh.se

Björn Åstrand
Halmstad University, Sweden
Box 823 - 30118 Halmstad
Email: bjorn.astrand@hh.se

Roland Philippsen
Halmstad University, Sweden
Box 823 - 30118 Halmstad
Email: roland.philippsen@hh.se

***Abstract*—We present a semi-supervised approach for semantic mapping, by introducing human knowledge after unsupervised place categorization has been combined with an adaptive cell decomposition of an occupancy map. Place categorization is based on clustering features extracted from raycasting in the occupancy map. The cell decomposition is provided by work we published previously, which is effective for the maps that could be abstracted by straight lines. Compared to related methods, our approach obviates the need for a low-level link between human knowledge and the perception and mapping sub-system, or the onerous preparation of training data for supervised learning. Application scenarios include intelligent warehouse robots which need a heightened awareness in order to operate with a higher degree of autonomy and flexibility, and integrate more fully with inventory management systems. The approach is shown to be robust and flexible with respect to different types of environments and sensor setups.**

I. Introduction

Semantic labeling has been the nucleus of much robotics research in different application areas. The objective behind *semantics* is to elevate the robot's understanding by linking its world model to human related meanings. Semantics vary by context, such as human-robot interaction [1], [2] and object semantics for manipulation [3]. Arguably, the term "semantic map" is most commonly associated with semantic labeling of places in mobile robotics. A robot's understanding solely from the sensor perspective is limited to the level of appearance-based distinctions (e.g. from vision, laser, haptics). On the other hand, human knowledge often derives semantics from the functionality of the subject (objects or places). In general, a robot cannot acquire semantics unless there is a link between human knowledge and a robot's understanding. That is to say, human semantics must be introduced to robots. More intelligent behavior is enabled by exploiting the links with human knowledge embedded in a robot's semantic model.

Automatic Inventory and Mapping of Stock (AIMS) is a project that targets the domain of intelligent warehouses, by leveraging a synergic application of logistics management and intelligent vehicles. In AIMS, a set of tools is developed for setting up and operating Automatic Guided Vehicles (AGV) in warehouses. The project covers a variety of tasks, from layout surveying to on-line inventory tracking. A central element for fulfilling the project objectives is the establishment of a robust and flexible semantic mapping framework. Our previous work [4], [5] focuses on spatial modelling methods well suited for semantic annotation and inference. In this present paper, we address the problem of providing human knowledge to robots for the purpose of semantic mapping.

This work was supported by the Swedish Knowledge Foundation and industry partners Kollmorgen, Optronic, and Toyota Material Handling Europe. It was carried out at the Center for Applied Intelligent Systems Research at Halmstad University, Sweden.

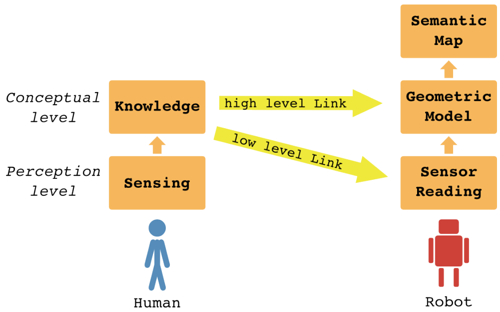


Fig. 1: Levels of semantic linkage between human knowledge and robot world model. Our method lifts the input required from humans to a higher level of abstraction by combining unsupervised place categorization with automatic spatial abstraction.

Related Work: There is a significant body of work on semantic analysis of the kinds of environments relevant for AIMS. We discuss only the ones most pertinent to this paper, emphasizing the abstraction level where semantic links are introduced (Fig. 1).

Low level delivery of human defined semantic to robots covers research that provides training data with semantic labels. Examples of such work includes supervised semantic labeling of cell in an occupancy map [6], using image features and similarly range features for place labeling [7], through generation of a visual vocabulary based on features and employing Latent Dirichlet Allocation (LDA) [8], or learning of feature hierarchies at different abstraction levels [9].

On the other hand, some works provide semantic notation to the robot when it already has a map of the environment (high level link in Fig. 1). In such a case, either place segmentation and semantic annotation is performed manually (as it is provided on the location), or place detection and labeling is performed with a template matching approach. Examples of manual selection and annotation are; [10] explores an occupancy grid map through the Voronoi graph, [11] uses natural language based annotation on location, [12] segments and annotates of decomposed occupancy map, [13] annotates place with semantic label on the location using a laser pointer. Examples of matching a template with the model are; [14] builds a geometric model of the world, and looks for templates (e.g. walls, floors, or objects), [15] proposes an inference model over a 3D correlating objects framework with semantic labels, [16] provides a topological graph of the environment based on the concept of virtual door, [17] embeds semantics in kernels and the map is refined and segmented accordingly, [5] presents semantic labels as graph templates and matches to the connectivity map of the open space, [3] performs a segmentation and semantic labeling of objects in 3D data. The last example is the probabilistic framework for semantic inference proposed by [18] which covers a large variety of techniques on different levels.

The provided examples deliver reliable performance in specific application areas such as office environments. Application area of the AIMS project requires semantic labeling of the structure of the environment. Approaching semantic labeling through objects found in places seems less suitable. In addition to that, our solution is required to be generic, independent of training and robust to different configuration of environment and mapping. That explains why we are interested in postpoing the delivery of semantics to robot's model into higher level.

Our Approach: The focus of this work is the construction of a semantic map without providing semantic training data as an input to the system. The robot autonomously builds a high level spatial model of the world and instantiates it, without prior knowledge of the environment. This allows an abstraction of human semantics, instead of providing this knowledge through a training set. Human defined semantics are introduced at a high level, as annotations of the model's instances. These semantics are provided as labels accompanied with their functionality and possible inter relations between them. This allows the robot to perform an inference over its model and provided semantics, resulting in the desired semantic map. The contribution of this work is the autonomous generation of a high level representation that allows an efficient abstraction of the semantics based on human input.

II. Method

The end goal of this work is the generation of a semantic map. For this purpose, an abstract spatial representation of the environment is provided and fused with an unsupervised place categorization of each cell of the occupancy map. While the robot generates and carries a high level spatial-conceptual model of the environment, human semantics is introduced. The semantics notion is provided in a high level and abstract form. The semantics relates the place categories in the robot's model with human-defined functionalities and attributes of the places. This connection enables an inference over the robot's spatial-conceptual model, into a semantic map. The developed semantic map maintains the abstract spatial model of the environment along with the semantic labels of each entity in the model, accompanied with inferred inter relation between entities and their functionalities.

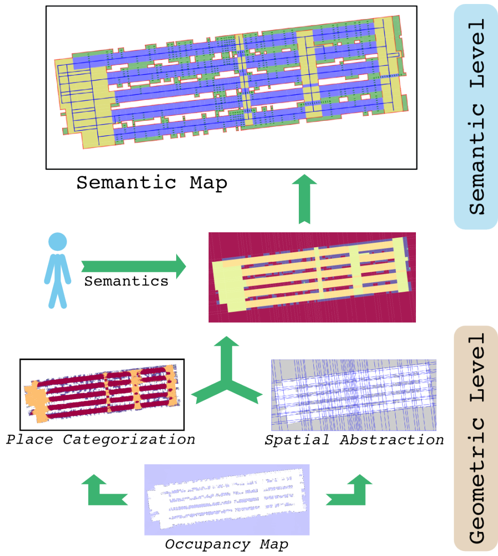


Fig. 2: Overall flow of the proposed method. The geometric level runs without human input and produces a spatially abstracted map annotated with place categories. The semantic level leverages human refinement of the geometric information to infers a higher-level semantic map.

This process is depicted in Fig. 2. For each map the two processes of spatial abstraction and place categorization is done independently. Spatially abstract representation of the map is achieved by an Adaptive Cell Decomposition represented by two data structures of subdivision and adjacency graph (see our previous work[4]). The other process performs an unsupervised place categorization of the map, by clustering places according to their geometric shape represented by a feature set proposed by [6]. This feature set is based on the assumption that places are identifiable based on their shape as reflected in the sensor reading. Furnishing the spatial model with place categories, the model holds instances ready to receive semantic annotations, designed according to the environment and in correspondence to the place categorization. With simple inference rules the robot is able to realize relations among neighboring instances of the model. The inference over the semantic labels in the context of the robot's model results in the desired semantic map.

A. Unsupervised Place Categorization

Mozos et al. proposed a method of supervised semantic labeling of places [6] based on geometric descriptors and

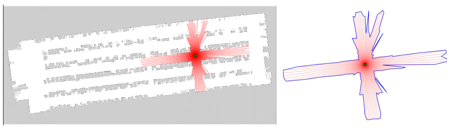


Fig. 3: Raycast example in a warehouse.

Mean and standard deviation of beams' length.
Mean and standard deviation of differences between consecutive beams' length.
Area (A) and perimeter (P) of the surface covered by the raycast, and the ratio (P^2/A).
Number of gaps between consecutive beams, defined by a threshold.
Kurtosis, a measure of elongation of the covered area.
$PC1$ and $PC2$ of the PCAs from the raycast. The ratio ($PC1/PC2$) and the offset between origin of the PCA frame and raycasting point.

TABLE I: Thirteen features extracted from raycast proposed by [6].

bootstrapped classifiers. From each cell of the occupancy map, laser readings are simulated by raycasting (Fig. 3), and simple geometric features (table I) are extracted from the raycasts. In their method, the map is divided into two parts, one for training classifiers and one for evaluation. The performance of this feature set is evaluated in office-like environments.

Place categorization in our work is inspired by their feature set. However, instead of classifying map cells based on training from manually labelled data, our work relies on unsupervised clustering. By using this set of features we rely on the same assumption as [6], namely that places in the map are distinguishable based on their shape as perceived by the sensor. Our unsupervised approach liberates it from the need for specific and predefined semantic labels.

One of the advantages of using raycasts is that they sees far beyond the origin location. More importantly, there is no need for an assumption on the underlying structure of the environment for categorization of places (as opposed to training sets). Fig. 6a demonstrates the result of place categorization of an occupancy map from a real warehouse. In our current implementation, we use k-means clustering of the features presented in table I, and the *k* value is set according to type of the environment (i.e. more than the number of expected categories). However, note that our method does not rely on the specific clustering algorithm, and we expect to go beyond simple k-means in future work. Our current implementation assigns a single category label $\mathrm{L}(p_i)$ to each cell $p_i = (x_i, y_i)$ in the occupancy map. Future work may extend this to more expressive schemes, such as determining distributions over possible categories.

B. Adaptive Cell Decomposition

In previous work we developed an *Adaptive Cell Decomposition* of maps [4], a novel approach of abstraction and capturing the latent structure of the environment by adapting the decomposition to the spatial resolution of the structure. This structure could be generated by walls in an office environment, or the alignment of physical entities (crates containing articles) in a warehouse which typically lack walls where AGVs operate. This subsection summarizes the decomposition method

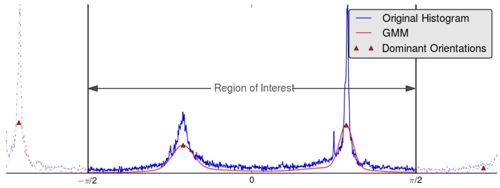


Fig. 4: Detection of dominant orientations, using Continuous Wavelet Transform (CWT) of weighted Histogram of Oriented Gradients (HOG), modeled by Gaussian Mixture Model (GMM).

and describes improvements and modifications introduced for this paper.

Dominant Orientations: Decomposition starts by finding the dominant orientations of the structure through the peaks in the weighted Histogram of Oriented Gradients (wHOG), given the number of expected orientations. We have removed the previous need to provide the number of dominant orientations by modelling the wHOG with a Gaussian Mixture Model (GMM). Dominant orientations are detected from the GMM modelled signal by a peak detection technique based on Continuous Wavelet Transform (CWT). Although CWT is known to be robust against noise by employing a range of wavelets with different scales, it is not sufficient for this case. This is due to cases where spikes appear close to strong peaks which correspond to dominant orientations. While it is desired to reject this kind of noise, there are similar stand-alone peaks that actually represent other dominant orientations, and which should not be filtered out. Therefore if CWT is used solely, the range of wavelet scales must be kept very broad. Here GMM acts as an aggressive Low Pass Filter, to merge noises into dominant orientations, while not completely removing important stand-alone peaks. That is to say, GMM simplifies and captures the outline of the histogram (See Fig. 4), resulting in a smoother signal. This allows wavelet scales to be more narrow, resulting in increased robustness. Dominant orientations are represented by the set Θ

$$\Theta := \left\{ \theta_i \mid i \in \mathbb{N},\ \theta_i \in [-\frac{\pi}{2}, \frac{\pi}{2}] \right\}$$

Cell Decomposition: A set of lines ($\mathcal{L}_{ines}$) constructs a low level spatial representation of the environment's structure. This set underlies the cell decomposition.

$$\mathcal{L}_{ines} := \{ y = m_i \times x + b_i \mid i \in \mathbb{N},\ x, y, m_i, b_i \in \mathbb{R} \}$$

As the gradient m_i is already computed in the dominant orientations Θ, the constant term b_i (y-intercept) remains to be determined. This is achieved by radiography. Radiography is a projection of occupied cells of the map along the angles in Θ. Peaks in the radiogram signals correspond to alignment of physical elements in the occupancy map along the angles in Θ. For more details see [4].

Data Structures: Cell decomposition is generated over the line set $\mathcal{L}_{ines}$ and represents a higher level abstraction

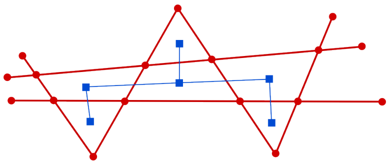


Fig. 5: Subdivision and adjacency graph. Thick red lines and red dots are edges and vertices of the subdivision. Thin blue lines and blue squares are links and nodes from adjacency graph.

of the map. The cell decomposition is conveyed by two data structures, subdivision and adjacency graph (see Fig. 5).

The subdivision S is composed of 3 sets of $\mathcal{V}_{ertices}$, $\mathcal{E}_{dges}$ and $\mathcal{F}_{aces}$. A neighborhood function of edges $\mathrm{N}(e)$ and a membership function of faces $\mathrm{M}(f)$ are defined over those sets of subdivision.

$$
\begin{aligned}
&\mathrm{S} = (\mathcal{V}_{ertices}, \mathcal{E}_{dges}, \mathcal{F}_{aces}) = (\{v_i\}, \{e_i\}, \{f_i\})\\
&v_i = (x_i, y_i),\ e_i = (v_{il}, v_{ik}),\ f_i = \{e_{ij}\}\\
&\mathrm{N}(e_i) = \{(f_{il}, f_{ik}) \mid (e_i \in f_{il}\ \wedge\ e_i \in f_{ik})\}\\
&\mathrm{M}(f_i) = \{(x_i, y_i) \mid (x_i, y_i) \text{ inside } f_i\}
\end{aligned}
$$

The adjacency graph A has a neighborhood function of nodes $\mathrm{N}(n)$ operating on two sets of $\mathcal{N}_{odes}$ and $\mathcal{L}_{inks}$.

$$
\begin{aligned}
&\mathrm{A} = (\mathcal{N}_{odes}, \mathcal{L}_{inks}) = (\{n_i\}, \{l_i\})\\
&n_i = (x_i, y_i),\ l_i = (n_{i1}, n_{i2})\\
&\mathrm{N}(n_i) = \{n_j \mid l_k \in \mathcal{L}_{inks},\ l_k = (n_i, n_j)\}
\end{aligned}
$$

The two data structures, subdivision and adjacency graph, are dual in nature (see Fig. 5). That is to say, each face and edge in S are dual to a node and link in AG respectively. Duality is addressed by a set of *dual* functions $\mathrm{D}(\cdot)$.

$$\mathrm{D}(n) \in \{f\},\ \mathrm{D}(f) \in \{n\},\ \mathrm{D}(l) \in \{e\},\ \mathrm{D}(e) \in \{l\}$$

An adaptive cell decomposition of a warehouse map is illustrated in Fig. 6b only by its decomposing lines ($\mathcal{L}_{ines}$).

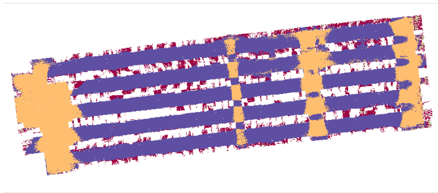
Subdivision Annotation: Originally, in [4], the occupancy of the faces are referred to as their labels. Here, we rely on the unsupervised place categorization instead. In the current implementation, faces in the subdivision inherit their labels $\mathrm{L}(f)$ from underlying cells (pixels) of the occupancy map in a voting process (winner takes all). Future work may extend this to more elaborate schemes, for example to increase robustness and flexibility.

$$\mathrm{L}(f_i) = \text{Statistical Mode}\,\{L(p_i) \mid p_i \in \mathrm{M}(f_i)\}$$

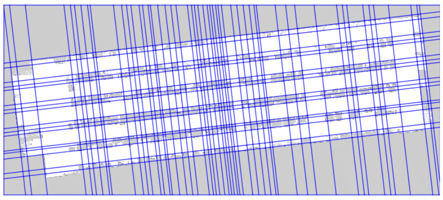
Inheritance of the labels from occupancy map to the subdivision is demonstrated in Fig. 6c.

C. Introducing Human Semantics

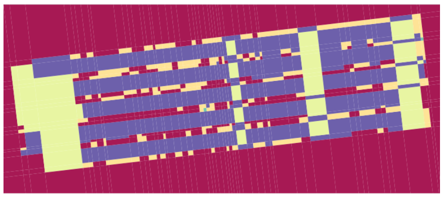
At this point the robot has a rich representation model of the environment. Yet this model requires semantics from human knowledge in order for the robot to benefit from its model's full functionality. The result of this link heightens the robot's awareness of its environment.



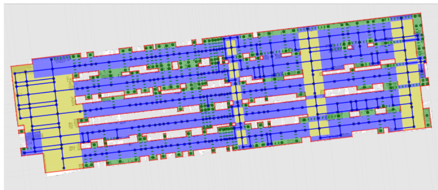
(a) place categorization



(b) adaptive cell decomposition



(c) subdivision annotation



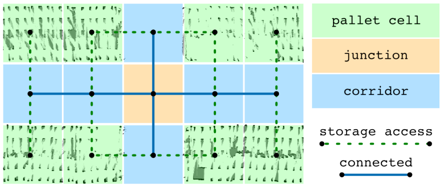
(d) semantic map

Fig. 6: 6b demonstrates the decomposition of an occupancy map, 6a shows unsupervised place categorization, 6c demonstrates the annotation of the decomposition (6b) with categories from 6a, followed by 6d demonstrating one possible inference (table II) over the linked knowledge of human to robot's model. Note that only colors from 6d have semantic meanings, and they correspond to those colors of figure 7a.

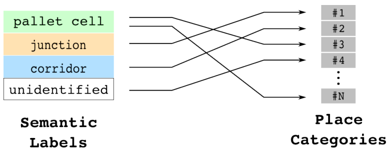
An example of realizing this objective is to deliver to the robot a set of semantic labels, their functionalities and corresponding relations. We propose such labels for a warehouse in table II and visualized in Fig. 7. This is one way, and to our experience, the most straight forward way of labeling a warehouse environment. While table II is highly abstract, it is sufficiently descriptive to enable the robot inferring human related semantic labels from its model. For example type of relations between neighboring instances of the model could be

Label	Functionality	Connect-ability
co: *corridor*	drivable & storage access	co & ju & pc
ju: *junctions*	drivable & connection	co & ju
pc: *pallet cells*	storage	co
un: *unidentified*	—	None

TABLE II: Semantic labels, their attributes and functionality *manually* designed for warehouses.



(a) semantic labels



(b) assigning semantics to place categories

Fig. 7: In this figure, human semantics designated to places in warehouse are demonstrated according to table II. Types of connections between different places as annotated, could be inferred from the same table.

inferred.

Edges and links in the subdivision and adjacency graph have attributes, indicating the type of connections between neighboring faces/nodes. The attributes depend on the labels of neighboring faces, and can be inferred from the "Connect-ability" column of table II. A *cogent* edge implies a physical obstruction between neighboring faces (and dual nodes), which means that the dual link has a *disconnected* attribute. An *invalid* edge implies an open access between neighboring faces (and dual nodes), and consequently the dual link is *connected*. The third type of edge/link attribute, *storage access*, indicates that a drivable face (corridor) is neighboring a storage face (pallet cell). One generic realization of inference over table II is given by the *Cogency* function of edges $\mathrm{C}(e)$ and *connectivity* function of links $\mathrm{C}(l)$.

$$\mathrm{C}(e_i) = \begin{cases} \text{cogent} & if\ \text{un} \in N(e_i) \\ \text{cogent} & if\ N(e_i) = \{\text{ju} \wedge \text{pc}\} \\ \text{storage access} & if\ N(e_i) = \{\text{co} \wedge \text{pc}\} \\ \text{invalid} & otherwise \end{cases}$$

$$\mathrm{C}(l_i) = \begin{cases} \text{disconnected} & if\ \mathrm{D}(\mathrm{l_i}) = \text{cogent} \\ \text{storage access} & if\ \mathrm{D}(\mathrm{l_i}) = \text{storage access} \\ \text{connected} & if\ \mathrm{D}(\mathrm{l_i}) = \text{invalid} \end{cases}$$

Fig. 7 visualizes the interpretation of the cogency and connectivity functions. The decomposition of faces and their

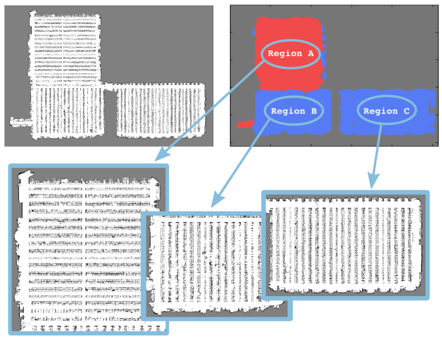


Fig. 8: Region segmentation using Gabor decomposition. The map represents a warehouse environment of size $\sim 190m \times 270m$.

place categories in Fig. 7 is already available in the robot's model of the environment. The link between human semantics (table II) and robot model allows the robot to perceive its surrounding similar to humans. At this stage of our work, the links between human semantics and robot's model of the environment are assigned manually. However it is possible within the same framework presented in this paper to learn the suitable links from characteristics of labels and the occupancy values.

D. Region Segmentation

This work focuses on semantic labeling and a high level human semantic involvement. However some additional techniques that improve the result are presented alongside. One noteworthy technique is "region segmentation". Regions in a map are those blocks of the map that have different patterns. Differences in patterns could be rooted in different structures, or even similar structures with different dominant orientations (see Fig. 8). If a map is composed of multiple regions, line sets $\mathcal{L}_{ines}$ from one region might interfere with cell decomposition of other regions, as cell decomposition is a global operation. Although theoretically the final result is still valid in such a case, yet it causes a higher decomposition and consequently degrading the abstraction level. To avoid this undesirable effect of multi regions, we exploit an unsupervised texture segmentation method based on Gabor filters proposed by J. Bigun [19]. The method proposes a Gabor filter for decomposition and feature extraction, followed by clustering of pixels according to their features. In this implementation we set the Gabor filter with 5 frequency channels and 8 orientation channels, resulting in a 40 dimension feature space. Dimensionality is reduced by Principal Component Analysis, and a k-means clustering is employed as the final step of texture segmentation. Performance of the texture segmentation method over a multi region map of a real warehouse of size $\sim 190^m \times 270^m$, fit with bounding boxes, is presented in Fig. 8.

III. Results and Discussion

Our peoposed method improves the state of the art methodologically. The improvement is made possible by abstracting

Region	"Total Pallets"	Correct Detection	Systematic Error
A	2628	82.6%	11.0%
B	1891	77.4%	18.2%
C	2172	92.9%	3.2%

TABLE III

Label	Functionality	Connect-ability
co: *corridor*	drivable & access	co & ju & of
ju: *junctions*	drivable & access	co & ju & of
of: *office*	work place	co & ju & of
un: *unidentified*	—	None

TABLE IV: Semantic labels for office environments.

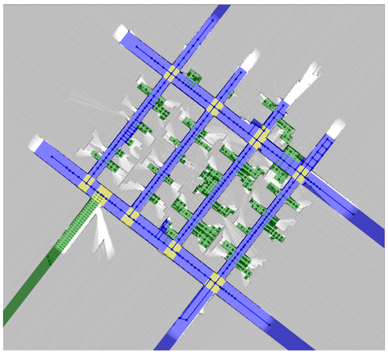


Fig. 10: Semantic map of an office-like environment obtained from the Radish data set, according to table IV.

the delivery of human concept of environment to the robot, and hence avoinding a time consuming low-level training phase.

We applied our method to two different maps acquired in the warehouse of one of our industrial partners, and to one office environment map available from the Radish data set. The warehouse data was acquired in different environments and with different configurations of the sensor, resulting in different maps. In one case (Fig. 6) the range scanner is pointing at the articles stored in the pallet cells, therefore the pallet cells are mostly either occupied or unexplored. The range scanner in the other configuration is pointing to the pallets in the cell, recording the pallet footprints in the map (Fig. 7).

While all regions of the map in Fig. 8 have been tested, due to the lack of space only one region (C) is shown in Fig. 9. Table III presents a qualitative measure of the performance over the pallet cells. We skip the margin areas, as they are partially explored and the information in those regions is unreliable. Systematic error refers only to the errors caused by violation of the assumption that any place in the environment is identifiable based on its shape (i.e. any cell in the decomposition). For instance, storage cells can be empty instead of containing a pallet, or storage cells can be in unexplored areas where raycasts do not penetrate (such as when a crate is placed directly on the floor instead of on a pallet). Note that systematic error and correct detection do not add up to 100%.

A. Beyond Warehouses

We showed that the method handles warehouses under differing sensor configurations. Here, we furthermore demonstrate the applicability of the concept to other kinds of environments, such as office buildings. For this purpose the method is applied to the *Intel Jones Farms Campus map*[1]. Corresponding to the change of the environment, we defined a new set of semantics (table IV). The result is shown in Fig. 10.

B. Discussion

In this paper we demonstrate a method of semantic labeling of the environment, relying on a high level link between human knowledge and robot world model, using an abstract form of human semantics. The final result contains semantic annotations that are not explicitly explained to the robot, but inferred according to the linkage with human knowledge. This introduces the semantics at a higher level, which is significantly less onerous and allows the semantics to be abstracted more efficiently. We have also shown that, the method adapts with minimum effort to new kinds of environments and sensor configurations.

Outlook: The performance achieved with the current implementation corresponds to a proof of concept. Some shortcomings remain, leaving room for future improvements, some of which have been sketched in the methods section already. Some of these limitations are rooted in the shortcomings of underlying tools (e.g. k-means) and can be addressed by working on individual components. Other are more systematics, and we plan to employ combinations with other methods to overcome them.

One element of the method that significantly affects the result is the feature set, clustering, and consequently the place categorization. In [6], probabilistic relaxation is used to compensate for misclassifications. We plan to incorporate a similar scheme as well. Also we employed k-mean for clustering which does not support multi-labeling of observations in its original form. We expect multi-labeling to enable better place categorization, as some area might share similarities. For instance, corridors and junctions are both drivable and may often be very similar in the makeup of the features extracted via raycasting.

Concerning the systematic shortcomings, we plan to improve our results by augmenting a landmark map generated from vision sensor from our previous work [5]. The landmark map generated in [5] models the infrastructure of the environment (pillars) providing a more accurate spatial decomposition of the model. Here, we expect to take inspiration from research on hierarchical multi-modal methods for semantic analysis.

IV. Conclusion

Robots that have an increased understanding of their environment are a key enabler for novel applications in automation, such as intelligent warehouses. In addition to increasing the reliability and robustness of operation, it can provide crucial

[1]This data set was obtained from the Robotics Data Set Repository (Radish) [20]. We thank *Maxim Batalin* for providing this data.

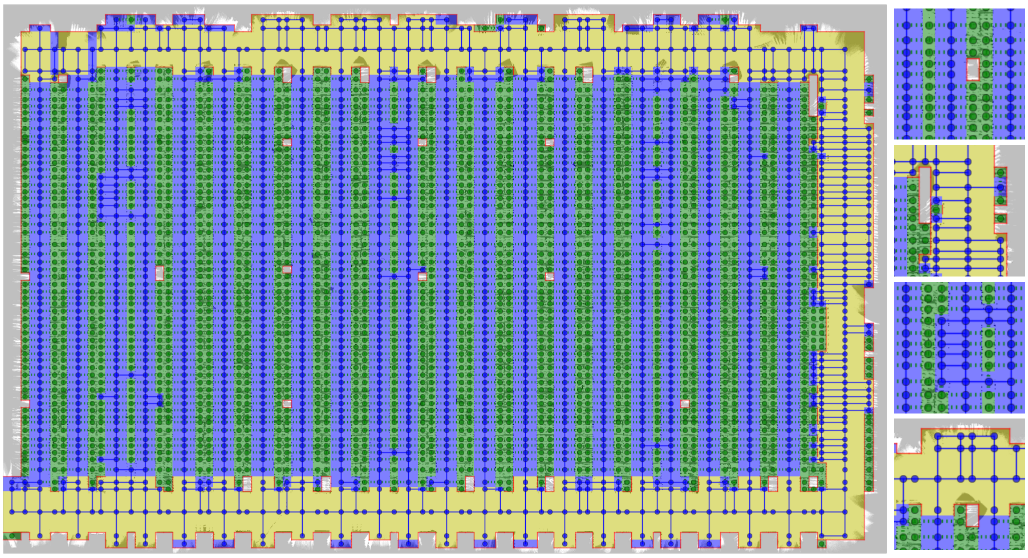


Fig. 9: Semantic map of a real warehouse, junctions in yellow, pallet cells in green and corridors in blue. Four cropped images on the right illustrate some failure cases.

background knowledge for adaptive learning or reasoning in novel situations. The presented work provides a reliable method of semantic interpretation of maps, supporting robots' awareness of their surroundings.

The proposed method generates a semantic map of the environment, suitable for operating autonomous robots. The problem is approached by spatial modelling of a given map, and delivering the semantics to the robot on a higher level in an abstract form. We employ a method of adaptive cell decomposition for modelling the map according to the structure of the environment. Place categorization is performed with clustering methods employing a set of simple geometric features extracted from raycasting. We present a case study of three real-world maps, one from a commonly available dataset and the others from data we collected with two different sensor setups in the warehouses of an industrial partner. The result demonstrates the simplicity and effectiveness of the method in adapting to new environments.

The key features of the proposed method are the independence from training data for semantic labels. Delivery of human semantics to the robot is done on a higher level, and directly supports semantic inference. Places are categorized in a semi-automatic unsupervised manner and are independent from the type of environments.

Improvement of the proposed method will include enhancements of place categorization and the incorporation of other modalities.

References

[1] V. M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori, "Hri in the sky: Creating and commanding teams of uavs with a vision-mediated gestural interface," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 617–623.

[2] S. Pourmehr, V. M. Monajjemi, R. Vaughan, and G. Mori, "you two! take off!: Creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 137–142.

[3] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," *KI-Künstliche Intelligenz*, vol. 24, no. 4, pp. 345–348, 2010.

[4] S. G. Shahbandi, B. Åstrand, and R. Philippsen, "Sensor based adaptive metric-topological cell decomposition method for semantic annotation of structured environments," in *to appear in 13th International Conference on Control, Automation, Robotics & Vision (ICARCV)*. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-26597

[5] S. G. Shahbandi and B. Åstrand, "Modeling of a large structured environment," in *Advances in Autonomous Robotics Systems*, ser. Lecture Notes in Computer Science, M. Mistry, A. Leonardis, M. Witkowski, and C. Melhuish, Eds. Springer International Publishing, 2014, vol. 8717, pp. 1–12. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-26316

[6] O. M. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using adaboost," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 1730–1735.

[7] A. Rottmann, O. M. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *proceedings of the national conference on artificial intelligence*, vol. 20, no. 3. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005, p. 1306.

[8] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, "Unsupervised semantic labeling framework for identification of complex facilities in

high-resolution remote sensing images," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010, pp. 273–280.

[9] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1915–1929, 2013.

[10] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: Extracting the topological structure of indoor environments via place labeling," in *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 35, 2007.

[11] M. R. Walter, S. Hemachandra, B. Homberg, S. Tellex, and S. Teller, "A framework for learning semantic maps from grounded natural language descriptions," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1167–1190, 2014.

[12] R. Capobianco, G. Gemignani, D. D. Bloisi, D. Nardi, and L. Iocchi, "Automatic extraction of structural representations of environments," in *Proceedings of the 13th Intelligent Autonomous System conference*, 2014.

[13] E. Bastianelli, D. D. Bloisi, R. Capobianco, F. Cossu, G. Gemignani, L. Iocchi, and D. Nardi, "On-line semantic mapping," in *Advanced Robotics (ICAR), 2013 16th International Conference on*, Nov 2013, pp. 1–6.

[14] A. Nüchter and J. Hertzberg, "Towards semantic maps for mobile robots," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 915–926, 2008.

[15] A. Ranganathan and F. Dellaert, "Semantic modeling of places using objects," in *Proceedings of the 2007 Robotics: Science and Systems Conference*, vol. 3, 2007, pp. 27–30.

[16] K. Joo, T.-K. Lee, S. Baek, and S.-Y. Oh, "Generating topological map from occupancy grid-map using virtual door detection," in *Evolutionary Computation (CEC), 2010 IEEE Congress on*. IEEE, 2010, pp. 1–6.

[17] Z. Liu and G. von Wichert, "Extracting semantic indoor maps from occupancy grids," *Robotics and Autonomous Systems*, vol. 62, no. 5, pp. 663–674, 2014.

[18] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3515–3522.

[19] J. Bigun and J. H. du Buf, "N-folded symmetries by complex moments in gabor space and their application to unsupervised texture segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 1, pp. 80–87, 1994.

[20] A. Howard and N. Roy, "The Robotics Data Set Repository (Radish)," 2003. [Online]. Available: http://radish.sourceforge.net