Semantic Maps for Robotics

Dagmar Lang and Dietrich Paulus
Active Vision Group
University of Koblenz-Landau
Universitätsstr. 1, 56070 Koblenz, Germany
{dagmarlang, paulus}@uni-koblenz.de

***Abstract*— Most mobile robotic systems use internal representations of the gathered information that is not intuitively understandable by humans and that is inadequate for learning from commonly available sources. The combination of object/place classification and common-sense knowledge to semantic maps found its way into indoor semantic mapping approaches in order to improve human-robot interaction. The aim is to pass complex task settings to the robot, so that it guides the search for the solution itself. In this paper, we present a common formal definition of semantic robotic maps as an extension of hybrid maps introduced by Buschka [1]. We discuss different criteria for design, classification and challenges of semantic maps. Furthermore, we present an evaluation template based on the definition, properties and challenges of semantic maps for three well known semantic mapping approaches.**

I. INTRODUCTION

In recent years, simultaneous localization and mapping algorithms for mapping the environment with cameras, 2D and 3D laser range finders have been developed and different methods have been established, e. g., [2], [5], [7], [10]. Most of the created maps are represented as a mixture of metrical and topological data structures. For complex tasks such as path and task planning, the representation of these maps has to be simplified and adapted to the scenario that the robot has to deal with.

Consider the task of fetching a cup of coffee. A robot with a metric map representation and the ability of object recognition has to search in a brute force manner in all open space. Introducing some common-sense knowledge about rooms including the probability of objects in a certain room, the search could be guided to places with high probability to places with lower probability.

The semantic mapping community can be subdivided on the basis of the application domain indoor, urban and cross-country environments. In literature, e. g., in [4], [9], [12], [13], [15], different methods are available to create indoor semantic maps with several sensors, such as cameras and/or laser range finders and map representations. By contrast, for outdoor environments, mainly place and object classification on 3D laser range data and cameras have been presented. Only a few publications present approaches establishing the connection between object classification and common-sense knowledge, e. g., [3], [8], [11], [14].

We first present a general formal definition of semantic maps based on the definition of hybrid maps of Buschka [1] in Section II. In Section III we discuss different criteria for designing and classifying semantic maps and the challenges of designing semantic maps. Based on an evaluation scheme, we present well known semantic mapping approaches in Section IV. We draw our conclusion of this work in Section V.

II. DEFINITION OF SEMANTIC MAPS

In this section, we present the formal definition of semantic maps. We first present the basic definitions of how and of what we are mapping. Then, we present the definitions of different (robotic) maps and their probabilities. At the end, we present the definition of semantic maps. The presented definitions are rather nonrestrictive. The aim is to allow for discussing different varieties of semantic maps in literature and their mechanisms, without committing to any specific formalism. We define maps and semantic maps in common, as (semantic) maps are not limited to the robotic domain, since the deployed sensors are also used in other domains.

A. Basic Definitions

When we talk about mapping and the formal definition of maps, we first define what we want to map and how to describe it. The idea of the following definitions in this subsection is based on Niemann [6]. In this context, we have a look at the definition of the world first and will be more precise and detailed in the following.

The world consists of physical entities such as objects, phenomena, etc. These entities are associated to a pose and they are described by attributes.

***Definition** 1 (World):* The world W is a physical space defined as tuple $W = \langle Y, P, \Sigma \rangle$, where Y is a set of entities, P is a set of poses and Σ is a set of attributes. Each entity $y \in Y$ will be associated with one pose $p \in P$ by a function $\mathrm{p} : y \longrightarrow p$. Furthermore, each entity y can be associated a subset of attributes by a function $\mathrm{a} : y \longrightarrow 2^{\Sigma}$.

For robotic tasks, the pose can be defined in an measurable space and entities are usually objects or things such as *houses*, *cars*, *people*, *rooms*, *tables*, *cups* and so on. Entities will be assigned different attributes with the function a. For example, the entities *path* or *grass* can be assigned the attributes *good* or *less drivable*, *cupboard* can be assigned *contains cup*. Attributes can be specialized, if necessary, e. g., to linguistic terms or real-valued measures.

Regarding the perception of humans or recordings of sensors, these modalities only observe a limited field of view. Therefore, the environment has to be defined as follows:

***Definition** 2 (Environment):* The environment E is a subset of W which describes a local area such as the field of vision of humans or sensors. It is given by $W \supseteq E = \langle Y', P', \Sigma' \rangle$ and consists of a subset of entities $Y \supseteq Y'$ located in the local area, a subset of corresponding poses $P \supseteq P'$ and a subset of attributes $\Sigma \supseteq \Sigma'$.

The design and use of semantic maps differs on the intended application, such as for in- and outdoor maps. In order to describe the intended application, the task domain has to be defined as:

***Definition** 3 (Task domain):* A task domain D is a subset of E. It contains only entities, poses or attributes belonging to a strictly limited application or subsection of the world. It is given by the set $E \supseteq D = \{E_i \mid i = 1, \ldots, n\}$.

The task domain can be observed by different sensors. Based on the properties of different physical devices, we define observations as:

***Definition** 4 (Observation):* An Observation O is defined as a tuple $O = \langle Y', P', \mathrm{s} \rangle$, which can be measured by physical devices in E defined by D. Function $\mathrm{s} : \langle Y, P, \Sigma \rangle \longrightarrow \langle Y', P' \rangle$ maps the entities and poses of E by representing sensor readings of an arbitrary physical device.

B. Definition of Different Maps

Buschka [1] defines general and specific maps as in Definition 5 - 8. We use his terminology and define his terms using our previous Definition 1 - 4. In contrast to Buschka, we changed Definition 6 from *robot map* to *observation map*, because maps can be generated independent of any robotic system or robotic application. First, we have to define a map in common sense. If we look at the different types of maps available, a set of entities observed will be mapped to a subset of the environment and is limited to a task domain, e. g., a road map or an underground map.

***Definition** 5 (Map):* A map of E limited to D is a pair $M = \langle Y', \mathrm{g} \rangle$, where Y' is a set of entities and a function $\mathrm{g} : Y' \longrightarrow 2^E$ associates each element in Y' with a subset of E.

When we talk about robotic maps or other map representations, the set of entities measured by the sensor system is represented in a specific mathematical structure $\mathcal{S}$, which associates each element in Y' with a subset of E. Thus, an observation map is defined as:

***Definition** 6 (Observation map):* An observation map for E limited to D is a tuple $M_{\mathrm{R}} = \langle Y', \mathrm{f}, \mathcal{S}, \mathrm{pos} \rangle$, where Y' is a set of entities, $\mathcal{S}$ is a mathematical structure. A function $\mathrm{pos} : Y' \longrightarrow 2^{\mathcal{S}}$ associates each entity in Y' with a subset of $\mathcal{S}$ and a function $\mathrm{f} : \mathcal{S} \longrightarrow 2^E$ associates each element in $\mathcal{S}$ with a subset of E.

In practice, we map a task domain. Among others in literature, metrical and topological representation types of maps are presented. According to these representations, [1] presents the definition of metrical and topological maps as:

***Definition** 7 (Metric map):* A metric map for E limited to D is a tuple $M_{\mathrm{C}} = \langle Y', \mathrm{f}, \langle S, \mathrm{d} \rangle, \mathrm{pos} \rangle$as in Definition 6, where structure $\mathcal{S} = \langle S, \mathrm{d} \rangle$ is ametric space with a distance function $\mathrm{d} : S \times S \longrightarrow \mathbb{R}$ and associates a metric space $S \times S$ to a real number.

***Definition** 8 (Topological map):* A topological map for E limited to D is a tuple $M_{\mathrm{T}} = \langle Y', \mathrm{f}, \langle \mathcal{N}, \mathcal{E} \rangle, \mathrm{pos} \rangle$ as in Definition 6. The structure $\mathcal{S} = \langle \mathcal{N}, \mathcal{E} \rangle$ is a graph with a set of nodes $\mathcal{N}$ and a set of edges $\mathcal{E} \in \mathcal{N} \times \mathcal{N}$.

Note that both definitions are an extension to Definition 6 and only the definition of $\mathcal{S}$ is different. Buschka introduced a formalized definition which combines two or more maps of different types of maps into one, called hybrid map.

***Definition** 9 (Hybrid map):* A hybrid map for E limited to D is a pair $\mathcal{H} = \langle \mathcal{M}, \mathcal{L} \rangle$, where $\mathcal{M} = \{M_{\mathrm{C_1}}, \ldots, M_{\mathrm{C_n}}, M_{\mathrm{T_1}}, \ldots, M_{\mathrm{T_m}}\}$ is a set of maps for E, such that $M_{\mathrm{C_i}} = \langle Y_i', \mathrm{f}_i, \langle S_i, \mathrm{d}_i \rangle, \mathrm{pos}_i \rangle$ or $M_{\mathrm{T_i}} = \langle Y_i', \mathrm{f}_i, \langle \mathcal{N}_i, \mathcal{E}_i \rangle, \mathrm{pos}_i \rangle$. $\mathcal{L} = \{L_1, \ldots, L_p\}$ is a set of links, where $L_i = \langle x_{j_l}, x_{k_o} \rangle$ such that $x_{j_l} \subseteq Y_j'$, $x_{k_o} \subseteq Y_k'$ and $j \neq k$.

The applicability of Definition 9 to various established map representations has been demonstrated in [1].

C. Properties of Hybrid Maps

Buschka defines three essential properties which have to be satisfied for the usability of hybrid maps applied to real world applications. Firstly, there has to be a function, which enables a robot to localize itself in the consisting $\mathcal{S}$ representing the map. Secondly, there has to be a function, which allows for determining, if there is traversable space in $\mathcal{S}$ of the map in order to move from pose A to B. And thirdly, there has to be a map building function, which allows for creating a map in a $\mathcal{S}$ based on actions of a robot. The formal definition of this functions can be found in [1]. In contrast to Buschka, we treat a robot as a physical entity of the task domain. The presented functions can be adapted to this definition.

According to Buschka, one major advantage of hybrid maps is the ability to exploit the individual advantages of each map component. Therefore, maps can be classified according to their component co-operation into *injection* and *synergy*. Injection evaluates if one component of the hybrid map does not have a given usability, e. g., it is not separable with respect to that usability. Synergy increases the performance to a level that would be hard or impossible to achieve by transferring information to another map, using one single map component.

Three different dimension of hybridization of hybrid maps, which apply also for semantic maps, were introduced by [1]. Firstly, he classified hybrid maps according to their heterogeneity. If at least two components of $\mathcal{M}$ of a hybrid map are of essential different types, a map is *heterogeneous* and otherwise *homogeneous*. Secondly, a map can be classified according to its hierarchy. A hybrid map is *hierarchical* if its components are hierarchically ordered and *flat* otherwise. Thirdly, according to its separability, a hybrid map can be classified as *separable* if each component can be used independently from other components. A map is *integrated* if each component needs the other ones in order to operate.

D. Definition of Semantic Maps

Next, we define our own semantic map in a common sense such as Buschka for hybrid maps. We have to consider the task domain and the goal of a semantic map. As described above, we want to combine the classification of entities in the map with common-sense knowledge to obtain knowledge about a scene for better understanding and human-robot interaction. In order to describe different steps to obtain the semantic map, we first have to define classes that can be obtained by classification. Niemann [6] defines classes for the pattern recognition domain and the definition is adapted to our approach as:

***Definition** 10 (Class):* A set of entities Y can be described by a set of classes Ω, which can be partitioned into k subsets Ω_κ, $\kappa = 1, \ldots, k$. It is required that

$$\Omega_\kappa \neq \emptyset, \quad \Omega_\kappa \cap \Omega_\lambda = \emptyset, \kappa \neq \lambda \quad \text{and} \quad \bigcup_{\kappa=1}^{k} \Omega_\kappa = \Omega.$$

The function $\mathrm{c} : y \longrightarrow \Omega_\kappa$ assigns each $y \in Y'$ of a set of entities exactly one class Ω_κ. As a partition implies a equivalence relation, classes correspond to entities that have some similarity with respect to a certain property. If necessary, different partitions may exist to establish taxonomies or hierarchies of concepts with respect to other properties.

During semantic mapping, entities are perceived by a sensor and sensor data will be classified into classes. These classes could be, for example, *building*, *car*, *path*, *street* or *grass*. *Good* or *less drivable* normally are attributes, but in case of specific applications, such as path planning, they can also be classes.

The combination of classes and attributes can lead to high-level functionality. In order to design algorithms dealing with high-level functionality, semantic maps have to be defined as a foundation.

***Definition** 11 (Semantic map):* A semantic map for E limited to D is a tuple $\mathcal{M}_{\text{sem}} = \langle \mathcal{M}, \mathcal{L}, \mathcal{A} \rangle$, where $\mathcal{M}$ and $\mathcal{L}$ are defined in Definition 9. $\mathcal{A}$ is a structure, which represents knowledge about the relation between entities, classes and attributes, also known as common-sense knowledge about D. Generally, $\mathcal{A}$ can be defined in an arbitrary way and has to allow for inference.

In literature, $\mathcal{A}$ is defined, e. g., as a graph based structure or an ontology. The application of the definition to semantic maps in literature will be presented in Section IV.

III. Properties and Challenges of Semantic Maps

The formal definitions of the three essential functions presented in Subsection II-C apply also for semantic maps as they are an extension of the definition of hybrid maps. In the following, we present different criteria for the design and evaluation of semantic maps.

Semantic maps can be distinguished based on common-sense knowledge. On the one hand, the acquisition of the common-sense knowledge can be different. The knowledge can be modeled by the user or by the application of a common-sense database of the task domain. Additionally, common-sense knowledge about the task domain can be acquired by human-computer interaction and/or can be inferred during the mapping process based on acquired classification results and knowledge. On the other hand, the point in time when the knowledge is available for the robotic system can be modeled differently. In most of the cases, the knowledge is available at the beginning of the mapping process. On the other side, the knowledge acquisition can be modeled as first step of the mapping process, such that the knowledge will be acquired by the system before the mapping by human-computer interaction. If the semantic knowledge is available at the beginning of the mapping process, the knowledge can be adapted by the system through the mapping process by human-computer interaction or new classification results with reliable probability. The aim of common-sense knowledge is to infer from classified entities in the environment to attributes or vice versa. Here, the representation of the common-sense knowledge and the applied inference algorithms are distinguishable. One established method is to model a semantic mapping system with different layers such as sensor data, map information, classified entities and common-sense knowledge. In contrast, different layers with different abstraction levels can be modeled and the common-sense knowledge can be represented over different abstraction levels and not as separate layers. The knowledge can be modeled in a probabilistic and non-probabilistic manner. Graph based representations, such as directed, undirected and mixed graphs, or ontologies and corresponding inference methods are established methods for knowledge representation.

Semantic maps can be build from several sensor data. Multi-model and uni-modal approaches can be differentiated. The sensors used for map building, entity classification etc., can be divided into different classes. Most of the approaches use depth information, obtained by laser range finders, time of flight, structured light sensors or stereo data, or camera information obtained by one or more cameras or RGB-D sensors such as the Kinect.

The introduced semantic map from Definition 11 enables a lot of different applications. In the following, some basic functionalities, applications and queries will be presented, which could be supported by a semantic map that satisfies Definition 11. The semantic mapping system should support the recognition of entities in particular parts of the map and the classification of entities by assigning an entity a class or an attribute. Additionally, the system should allow for inferring from a given class of an entity to an attribute or a set of attributes or to infer between an attribute or set of attributes of an entity to the corresponding class. Given a point in the world or a position in the data structure of the map, it should be possible to obtain an entity at this position. The presented definition also supports complex tasks, such as to infer a probable pose of an entity or a set of entities which has not been previously observed according to the class or attributes of a part of the map. It could also be possible to search for special entities in a predefined part of the map. Furthermore, the definition allows for detecting localization errors in the map by reasoning about the expected location of entities and to deal with ambiguities in the map or ambiguous classifications of attributes or entities in a part of the map.

In addition to the presented properties, functionality and queries many principled design decisions have to be made, when considering the creation of a semantic map and one has to deal with a lot of challenges which arise during this process. One basic decision is to decide how many layers should be modeled for the semantic mapping system and how the common-sense knowledge will be integrated. That depends in principle on the applications based on the semantic map. During the modeling of the system, one has to decide how many topological and metric components are used for map creation and representation and how are these structures connected with each other. A challenge is to combine common-sense knowledge from different sources and different time points during the mapping process. If it is acquired for the application of the semantic map, the design of the system has to deal with the increasing amount of information over time during the mapping process. Systems applied for complex task settings have to deal with large-scale semantic information of the common-sense knowledge. The challenge is to use the most appropriate knowledge representation and suitable inference algorithm. An appropriate solution has to be found based on the knowledge representation and inference, if the system has to handle ambiguous observations in relation to the common-sense knowledge during the mapping system. One issue could be how to learn new information for the common-sense knowledge about the environment.

The presented properties and challenges are neither exclusive nor complete for all application domains. Hence, an extension or improvement for complex and more detailed scenarios is possible.

IV. APPLICATION OF SEMANTIC MAPS

The evaluation of hybrid maps is based on the following properties: usability, co-operation and hybridization. The application of the definition to hybrid maps was presented in [1]. In the following, we present an evaluation template according to the definition, properties and challenges of semantic maps presented in Section II and III. The evaluation criteria of [1] is out of the scope of this evaluation, since the main usabilities are self-evident for semantic mapping.

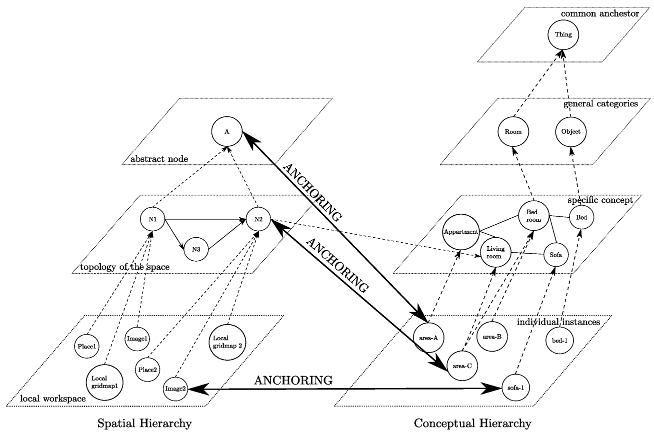

Fig. 1: Multi-hierarchical semantic maps presented by Galindo et al. [4] (Figure reproduced and adapted from [4]).

In the following, we will apply this evaluation scheme:

Description: General introduction of the semantic mapping system and its components.

Representation: Detailed description of the components of the semantic map and application to Definition 11.

Properties: Description of the presented properties of the semantic mapping system based on Section III.

We present the application to the evaluation template in the following for a selection of papers.

A. Multi-Hierarchical Semantic Maps

Description: Galindo et al. [4] introduce a multi-hierarchical approach dealing with a spatial and semantic representation of indoor office environments. They present a spatial and conceptual hierarchy which are connected by an anchoring process. An overview sketch is shown in Figure 1. The spatial hierarchy enables to reliably plan and execute robotic tasks by storing spatial and metric information from the robot environment. The spatial hierarchy is divided into three dimensions. On the lowest level, images of objects and local grid maps are stored. In the second level, the topology of the space is represented by nodes (open areas) and edges (navigable space between open areas) and in the upper level the whole spatial environment is represented in one node. The conceptual hierarchy provides the spatial perspective with a human-like interface and inference capabilities on symbolic reasoning by modeling semantic knowledge about the robot environment. All concepts derive from the common ancestor in the upper level. In the level below, general categories and in the level below that level, specific concepts are derived from the upper level. In the lowest level, individual instances are represented.

Representation: A semantic map is defined for the task domain indoor environment with the aim to infer about room types according to the appearance of objects and rooms. The map is defined as $\mathcal{M}_{\mathrm{sem}} = \langle \mathcal{M}, \mathcal{L}, \mathcal{A} \rangle$, where $\mathcal{M} = \{M_{\mathrm{C}_1}, \ldots, M_{\mathrm{C}_\mathrm{n}}, M_{\mathrm{T}_1}, \ldots, M_{\mathrm{T}_\mathrm{m}}\}$ is a set of maps for E. In case of multi-hierarchical maps, there are local

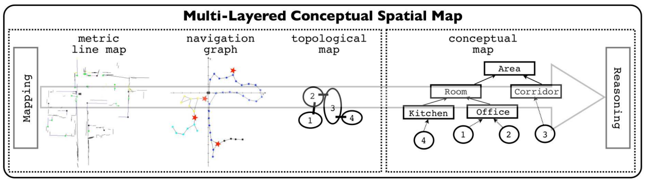

Fig. 2: Conceptual spatial map representation of Zender et al. [15] (Source: [15]).

metric components $M_{\mathrm{C_i}}$ (occupancy grid maps) associated one per node to a topological component in the spatial hierarchy. There are topological components $M_{\mathrm{T_i}}$ with different abstraction levels. In the spatial hierarchy, the second layer represents the topology of space and in the conceptual hierarchy, the layer defines specific concepts. $\mathcal{L}$ is liable to create links between entities in different (types of) maps. For the presented map representation, there are links between topological and metric components, e. g., between the local workspace and the topology of space in the spatial hierarchy. Links between different topological components are also present between the topology of space and the specific concept layer. The representation of the semantic common-sense knowledge about the task domain is defined as graph based structure $\mathcal{A} = \langle \mathcal{N}_{\mathcal{A}}, \mathcal{E}_{\mathcal{A}} \rangle$. The knowledge is represented in the conceptual hierarchy, in the specific concept layer as well as in links between the different layers. In the specific concept layer, classes as well as properties are linked together. The connection between the individual instances and the specific concept links entities with classes and properties. Between the specific concept and general category classes are linked with attributes.

Properties: Multi-hierarchical semantic maps can be applied to map building, navigation and detection of localization errors. The conceptual hierarchy provides the spatial perspective with a human-like interface and inference capabilities on symbolic reasoning by modeling semantic knowledge about the robot environment.

B. *Multi-Layered Conceptual Spatial Map Representation*

Description: The work of Zender et al. [15] extends the previous presented system by a multi-layered spatial representation consisting of metric, navigation, topological and conceptual map layers. The different layers are presented in Figure 2. The first three layers are used for mapping and the last one for reasoning. There are three main subsystems involved in constructing, maintaining and using the spatial representation: the perception subsystem for evaluation of sensory input, the communication subsystem for situated spoken dialog and the subsystem for multi-layered conceptual spatial mapping, that bridges the gap between sensor-based maps and the human-like spatial representation. The metric map used in the conceptual spatial map is a line based map and represents the part of the space that can be described by lines. The navigation map connects nodes and edges representing the trajectory of the robot and nodes are labeled by the classes *room*, *corridor* and *doorway*. The topological map divides a set of nodes in the navigation map into areas. The conceptual map on the one hand contains an innate conceptual ontology that defines abstract categories for rooms and objects and how they are related. On the other hand, information extracted form sensor data and/or given through situated dialogue about the actual environment is represented as tokens that instantiate abstract concepts.

Representation: A multi-layered conceptual spatial map is defined as $\mathcal{M}_{\mathrm{sem}} = \langle \mathcal{M}, \mathcal{L}, \mathcal{A} \rangle$, where $\mathcal{M} = \{M_{\mathrm{C_1}}, \ldots, M_{\mathrm{C_n}}, M_{\mathrm{T_{nav}}}, M_{\mathrm{T_{top}}}\}$ is a set of maps for the task domain of office environments. The navigation graph is a topological map $M_{\mathrm{T_{nav}}}$, which links small sets of local metric line maps $M_{\mathrm{C_i}}$ into a global frame work and also builds the basis for the topological map $M_{\mathrm{T_{top}}}$. $\mathcal{L}$ links entities of the metric, navigation and topological map. In the presented approach, the sensor data is classified into different classes in the navigation map and in the topological map into different topological areas. Attributes in this example are the concepts represented in the conceptual map, which can be derived by previously human-generated models or human-computer interaction. The common-sense knowledge about indoor environments is represented as a OWL-DL ontology. It connects sets of classified entities of the metric, navigation and topological map to attributes which allows for connecting the conceptual map and the other maps.

Properties: The conceptual knowledge is encoded as an OWL-DL ontology and a description-logic reasoner is used to classify spatial areas into corridors and different rooms such as kitchen and office. The combination and evaluation of acquired knowledge by human computer interaction and the asserted knowledge within the context of the initiate conceptual ontology enables the reasoner to infer more specific categories for known areas. The conceptual spatial map can be applied to map building, object recognition, door detection, place classification, navigation and reasoning.

C. *Large-scale Semantic Mapping and Reasoning*

Description: Pronobis et al. [9] present one main approach to develop an autonomous indoor service robot, which is able to create large-scale semantic maps and plan complex tasks. The semantic mapping system is divided into four layers on three hierarchy levels. A sketch of the layers is illustrated in Figure 3. The first hierarchy consists of the sensory layer, which represents camera and laser data, a metric map and the spatial information of features calculated from the data. The second hierarchy consists of two layers, the categorical layer and the place layer. The categorical layer represents room shape, appearance, objects and landmark models and the place layer contains the places, paths and place holders. The highest hierarchy represents the conceptual layer, which compromises common-sense knowledge about concepts, relations between those concepts and instances of spatial entities.

Representation: The large-scale semantic map is defined as $\mathcal{M}_{\mathrm{sem}} = \langle \mathcal{M}, \mathcal{L}, \mathcal{A} \rangle$, where $\mathcal{M} = \{M_{\mathrm{C_{sen}}}, M_{\mathrm{T_{sen}}}, M_{\mathrm{T_{pla}}}, M_{\mathrm{T_{cat}}}\}$ is a set of maps

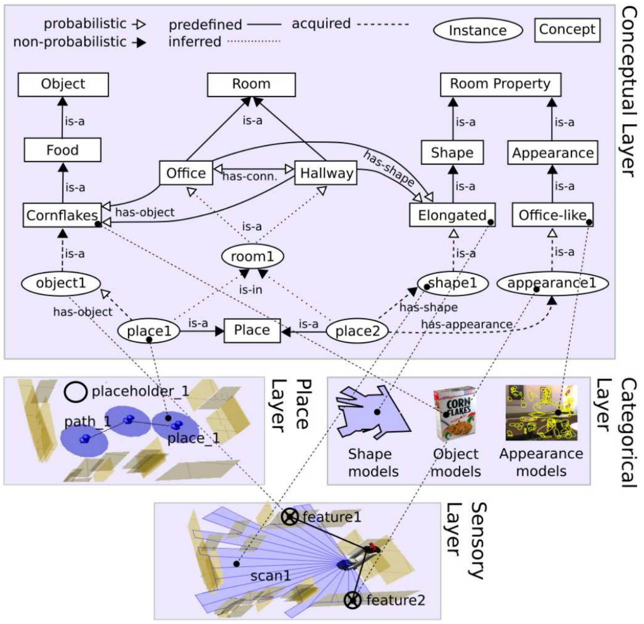

Fig. 3: Approach for large-scale semantic mapping and reasoning of Pronobis et al. [9] (Source: [9]).

for the task domain of office environments. The sensory layer is a hybrid map, where the camera and laser data and the spatial information of features calculated on the data are represented in a topological map $M_{\mathrm{T_{sen}}}$ and the corresponding grid maps in a metric map $M_{\mathrm{C_{sen}}}$. The place layer is represented as topological map $M_{\mathrm{T_{pla}}}$, where the nodes contain places and place holders and the edges contain the paths between the places and place holders. The representation of the categorical layer is not uniquely described in [9], but is assumed as topological, $M_{\mathrm{T_{cat}}}$, according to the links between the categorical and conceptual layer. There are no links $\mathcal{L}$ between sensory, place and categorical layer. The common-sense knowledge is modeled as graph $\mathcal{A} = \langle \mathcal{N}_{\mathcal{A}}, \mathcal{E}_{\mathcal{A}} \rangle$ by nodes and directed and undirected edges as relations between concepts, and describing instance knowledge as relation between either instances and concepts or instances and other instances. Entities are associated with attributes of the categorical layer and classes, such as room types and objects.

Properties: The inference between the conceptual and the other layers is realized by a chain graph model. The graph models both directed casual as well as undirected symmetric or associative relationships including circular dependencies originating from possible loops in the topological graph and enables the classification of parts of the topological map into room categories by inference on the chain graph. Relations in the common-sense knowledge are either predefined, acquired or inferred. The probability of existence of an object of a certain category in a certain type of room is first bootstrapped using a part of the *Open Mind Indoor Common Sense* database. The probabilistic relational conceptual representation is capable to perform uncertain inference about some concepts solely based on their relations to other concepts rather than direct observations. Hence, spatial reasoning about the unexplored space is now possible. The semantic mapping system can be applied to map building, object recognition and localization, door detection, place classification and reasoning.

V. Conclusion

In this paper, we presented a high-level overview of existing work on semantic mapping and a principle formal definition of semantic maps. We described requirements, properties and challenges of semantic maps, which have to be fulfilled to allow the combination between classification, mapping and common-sense knowledge. Additionally, we presented an evaluation scheme and applied it to well-known semantic mapping approaches.

VI. Acknowledgments

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG) under research contract PA 599/11-1.

References

[1] P. Buschka. *An Investigation of Hybrid Maps for Mobile Robots*. PhD thesis, Örebro University, Department of Technology, 2005.

[2] M. Cummins and P. M. Newman. Appearance-only SLAM at Large Scale with FAB-MAP 2.0. *The International Journal of Robotics Research*, 30(9):1100–1123, 2010.

[3] B. Douillard, D. Fox, F. Ramos, and H. Durrant-Whyte. Classification and Semantic Mapping of Urban Environments. *International Journal of Robotics Research*, 30(1):5–32, 2011.

[4] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernández-Madrigal, and J. González. Multi-Hierarchical Semantic Maps for Mobile Robotics. In *Proc. of IROS*, pages 3492–3497, 2005.

[5] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A General Framework for Graph Optimization. In *Proc. of ICRA*, pages 3607–3613, 2011.

[6] H. Niemann. *Pattern Analysis and Understanding*. Springer Verlag, 1990.

[7] E. Olson. Real-Time Correlative Scan Matching. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4387–4393, 2009.

[8] I. Posner, D. Schroeter, and P. M. Newman. Online Generation of Scene Descriptions in Urban Environments. *Robotics and Autonomous Systems*, 56(11):901–914, 2008.

[9] A. Pronobis and P. Jensfelt. Large-scale Semantic Mapping and Reasoning with Heterogeneous Modalities. In *Proc. of ICRA*, pages 3515–3522, 2012.

[10] J. Sprickerhof, A. Nüchter, K. Lingemann, and J. Hertzberg. A Heuristic Loop Closing Technique for Large-Scale 6D SLAM. *Automatika - Journal for Control, Measurement, Electronics, Computing and Communications, Special Issue with selected papers from the European Conference on Mobile Robots*, 52(3):199–222, 2011.

[11] R. Triebel, P. Rohan, D. Rus, and P. M. Newman. Parsing Outdoor Scenes from Streamed 3D Laser Data Using Online Clustering and Incremental Belief Updates. In *Proc. of AAAI*, 2012.

[12] S. Vasudevan and R. Siegwart. Bayesian Space Conceptualization and Place Classification for Semantic Maps in Mobile Robotics. *Robotics and Autonomous Systems*, 56(6):522–537, 2008.

[13] P. Viswanathan, D. Meger, T. Southey, J. J. Little, and A. K. Mackworth. Automated Spatial-Semantic Modeling with Applications to Place Labeling and Informed Search. In *Proc. of CRV*, pages 284–291, 2009.

[14] X. Xiong, D. Munoz, J. A. Bagnell, and M. Hebert. 3-D Scene Analysis via Sequenced Predictions over Points and Regions. In *Proc. of IRCA*, pages 2609–2616, 2011.

[15] H. Zender, Ó. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard. Conceptual Spatial Representations for Indoor Mobile Robots. *Robotics and Autonomous Systems*, 56(6):493–502, 2008.