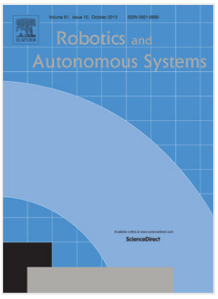
Accepted Manuscript



Semantic mapping for mobile robotics tasks: A survey

Ioannis Kostavelis, Antonios Gasteratos

PII: S0921-8890(14)00303-0
DOI: http://dx.doi.org/10.1016/j.robot.2014.12.006
Reference: ROBOT 2413

To appear in: *Robotics and Autonomous Systems*

Received date: 6 February 2014
Revised date: 7 December 2014
Accepted date: 12 December 2014

Please cite this article as: I. Kostavelis, A. Gasteratos, Semantic mapping for mobile robotics tasks: A survey, *Robotics and Autonomous Systems* (2014), http://dx.doi.org/10.1016/j.robot.2014.12.006

Semantic Mapping for Mobile Robotics Tasks: A Survey

Ioannis Kostavelis, Antonios Gasteratos

Laboratory of Robotics and Automation,
Production and Management Engineering Dept.,
Democritus University of Thrace,
Vas. Sophias 12, GR-671 00 Xanthi, Greece
Phone: +30 2541 079330
http://robotics.pme.duth.gr

Abstract

The evolution of contemporary mobile robotics has given thrust to a series of additional conjunct technologies. Of such is the semantic mapping, which provides an abstraction of space and a means for human-robot communication. The recent introduction and evolution of semantic mapping motivated this survey, in which an explicit analysis of the existing methods is sought. The several algorithms are categorized according to their primary characteristics, namely scalability, inference model, temporal coherence and topological map usage. The applications involving semantic maps are also outlined in the work at hand, emphasizing on human interaction, knowledge representation and planning. The existence of public available validation datasets and benchmarking, suitable for the evaluation of semantic mapping techniques is also discussed in detail. Last, an attempt to address open issues and questions is also made.

Key words: mobile robots, semantic map, topological map, temporal coherence, object recognition, place recognition, human-robot interaction, knowledge representation, planning

Email addresses: gkostave@pme.duth.gr (Ioannis Kostavelis), agaster@pme.duth.gr (Antonios Gasteratos)

A local villager knows his way by wont and without reflection to the village church, to the town hall, to the shops and back home again from the personal point of view of one who lives there. But, asked to draw or to consult a map of his village, he is faced with learning a new and different sort of task: one that employs compass bearing and units of measurement. What was first understood in the personal terms of local snapshots now has to be considered in the completely general terms of the cartographer. The villager's knowledge by wont, enabling him to lead a stranger from place to place, is a different skill from one requiring him to tell the stranger, in perfectly general and neutral terms, how to get to any of the places, or indeed, how to understand these places in relation to those of other villages.
¬ *Gilbert Ryle "Abstractions"*

1. Introduction

The above quoted metaphor was used by the coiner of the phrase *logical geography* in his attempt to elucidate the term [1], however today's robotics specialists have realized that they face the same problem as the local villagers, yet the other way round. Nowadays one may argue that the problem of *simultaneously localization and mapping* (SLAM) has been solved, still the output of such a process is only perceivable by a man bearing compass and units of measurement. Accordingly, contemporary mobile robots behave like machine cartographers, unable to liaise with local villagers, that is the human inhabitants, who know by wont to navigate through the own environment. Thus, the majority of the existing mapping approaches aim to construct a globally consistent metric map of the robot's operating environment. The robots bear state of the art instrumentation that allows, on the one hand, the construction of the map and, on the other hand, the own localization with respect to this map and, thus, to determine their global pose with remarkable accuracy. Based on this capability, the robots can plan a path and navigate towards a goal, which should be also a specified metric position in the global map reference frame. Howbeit, for a robot to apprehend the environment the way a human does and, consequently, to lead a stranger from place to place, a different skill than any geometrical map can provide is required. The robots to come should be endowed with capacities to understand their surroundings in a human-centric term, i.e. to be able to tell the difference between a room and a corridor or to discriminate the different functionality a kitchen and a living room have. Therefore, the formation of maps augmented by

semasiological attributes involving human concepts, such as types of rooms, objects and their spatial arrangement, is considered a compulsory attribute for the future robots that should be designed to operate in environments inhabited by humans.

A solution to this problem is offered by semantic mapping, a qualitative description of the robot's surroundings, aiming to augment the navigation capabilities and the task-planning, as well as to bridge the gap in *human-robot interaction* (HRI), see e.g. [2], [3] and [4]. Especially the work in [4] addresses semantic mapping with emphasis on HRI by using natural language, thus enabling the most direct way for robots to socialize with humans. Thence, semantic mapping is a flourished pioneering area encouraging the elaboration of several doctoral dissertations [5], [6]. The term semantic derives from the Greek word $\sigma\eta\mu\alpha\nu\tau\iota\kappa\grave{o}\varsigma$ [*sēmantikos*], standing for significant, which in turn derives from the verb $\sigma\eta\mu\alpha\grave{\iota}\nu\epsilon\iota\nu$ [*sēmainein*], meaning to signify, that successively stems from the noun $\sigma\tilde{\eta}\mu\alpha$ [*sēma*], that is sign. Thus, semantics is related to the study between signs and the things to which they refer, that is their meaning. The latter is oriented to the identification of the way that two or more entities interact, behave toward, and deal with each other [7]. Thereby, the semantic mapping targets to the identification and the record of the signs and the symbols that contain meaningful concepts for humans, during the robot's wander in human-inhabited areas. Consequently, a semantic map is an enhanced representation of the environment, which entails both geometrical information and high level qualitative features. Speculating the ability of the artificial agents to semantically perceive the own environment and accurately recall the learned spatial memories, the fundamental communication link between human and robots can be established. Therefore, for a successful HRI the robots must retain cognitive interpretation capacities about space, i.e. they should involve semantic attributes about the objects and the places encountered, in association with the geometrical perception of the surroundings. Moreover, the semantic information existing in the abient need to be organized in a such a fashion that the artificial agent can appropriately perceive and represent its environment. The most suitable way to organize all these information is by means of a map, namely a semantic map. Due to the fact that contemporary robots use to navigate in their environments by computing their pose within metric maps, the vast amount of the semantic mapping methods reported in the literature use these metric maps to add semantic information on top of it [2], [4]. Therefore, a semantic map comprises high level features that model the human concepts about places,

objects, shapes and even the relationships of all these, whilst a metric map retains all those geometrical features the robot should be aware of in order to safely navigate within its surroundings. Yet, it should be further noted that works have been reported on semantic mapping, which do not use a metric map to determine the type of a place, specially the ones using vision [8], [9].

The goal of the review paper in hand is to provide insights of the semantic mapping, to study the distinct components encompassing, to give a categorization of the related literature, to mention the possible applications in mobile robotics and, lastly, to refer to the methods and databases available for benchmarking. In order to support this goal, a quality-based taxonomy of the existing mapping strategies is attempted here, which should highlight the dominant attributes such methods retain. An illustrative representation of the described taxonomy of the most frequent components the semantic mapping approaches possess is depicted in Fig. 1. The primary characteristics constitute the condiciones sine quibus non a method producing a complete semantic map should satisfy. Of such are the modalities utilized to reason about the observed scene constituting an element apt to distinguish the abundance of different methods. In particular in many methods only single cues -e.g. objects- are utilized to infer about a place, while some other methodologies exploit multiple cues -such as objects, places and shapes- to produce semasiological clues about an area. Another frequented feature in many semantic mapping techniques is the temporal coherence such a map reveals, which renders it useful for high-level activities, viz. task planning or HRI. An additional important attribute a typical semantic mapping method possesses is the existence of a respective topological map, that is an abstraction of the explored environment in terms of a graph. The nodes of such a graph are organized in a geometrical manner, so as to simultaneously preserve conceptual knowledge about the explored places. These graphs could be either unconstrained ones retaining only geometrical characteristics or they could posses several constrains in accordance with the semantic attributes that they enclose. The existence of a 2D or a 3D metric map of the explored environment -either indoors or outdoors- is a complementary component, which frequently supplements the attributes implemented by the semasiological methods. According to the scale, to which each method is expanded, the metric map could be either a single scene or a progressively created map, that is the pose is referred to a local or a global coordinate system, respectively.

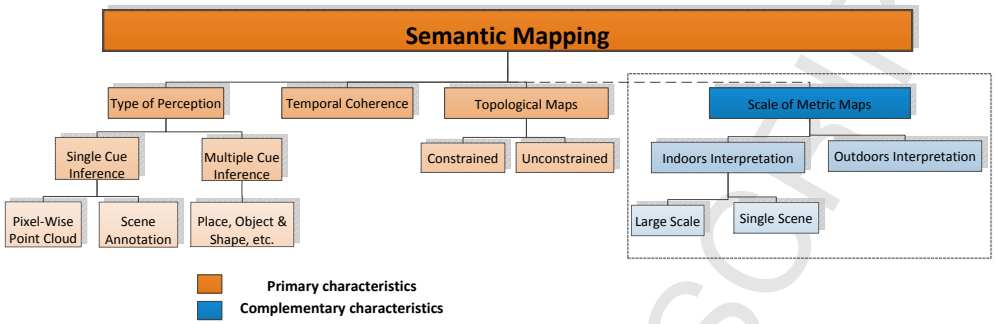


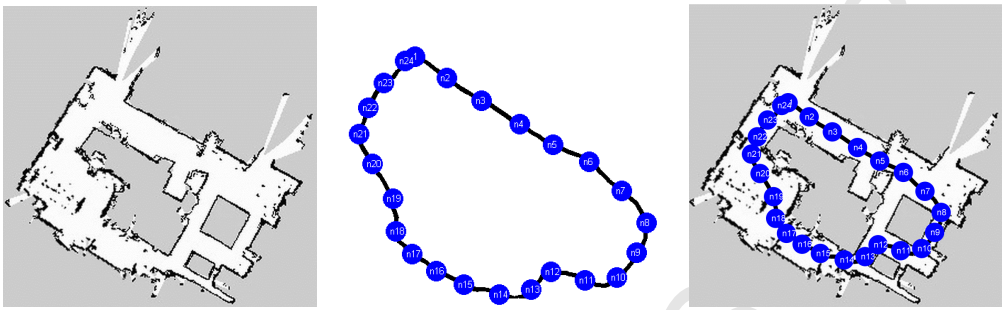
Figure 1: Taxonomy of the semantic mapping methodologies. Note that the metric mapping is considered complementary attribute for the semantic mapping

2. Antecedents

Among the several modalities employed in robot navigation, vision is the most dominant. This is mainly due to the fact that scholars are able to straightforwardly replicate the own vision-based navigation experiences onto their experimental agents. The first twenty years of vision-based robot navigation are surveyed in [10]. This work was conducted over one decade ago and it concludes that although then (2002) there were an adequate accumulated expertise to send a mobile robot from one coordinate location to another, there were not enough to perform *function-driven navigation*, such as to locate and bring a fire-extinguisher from somewhere within a hall, for instance. Moreover, in the same paper it is stated that navigation paradigms in which the robot needs to be conscious of the meaning of the objects it runs across are cumbersome. Any solution to such navigation situations need to be associated with the overall problem of computer vision, i.e. the automatic scene interpretation. Referring to Kuipers' pioneering works [11], [12], [13], the several semantic mapping and navigation methods introduced during the last decade aspire to fill this gap. Of course, in order for a robot to be able to navigate efficiently, a consistent geometrical map has to be built. Therefore, we could state that no progress would have been made in the area of semantic mapping, unless a prior advancement in SLAM had been made. In the last decades a profusion of laborious research has been conducted in the respective field, yielding remarkable results in the area of

mobile robot navigation and mapping [14]. With the purpose of accurately localizing themselves [15], [16], mobile robots construct a consistent representation of the spatial layout of their working environment. The representative works described in [14], [17], [18] and [19] prove the necessity for an accurate representation of the robot's surroundings as well as the development of efficient mapping methods. More precisely, SLAM provides solution to the problem, according to which a mobile robot placed at any unknown location in an unexplored area incrementally builds a consistent map of the environment while simultaneously determines its location within this map. Seeking to engineer an efficient solution to this problem, several successful research attempts have been carried out, an analytical summary of which is presented in a two-part review paper [20], [21].

Nevertheless, a deeper apprehension of the SLAM requires a further decomposition of the problem. An attempt to provide a taxonomy of the existing geometrical mapping methods, according to the way the environment is perceived, results in three classes, namely the metric, topological and topometric mapping. The metric mapping comprises a geometrical representation (see e.g. Fig. 2 (a)), where every pose is strictly related to a global coordinate system. This is typically either a 3D map or a 2D occupancy grid, which permits precise robot spotting. Besides, topological mapping involves a graph, each of the nodes of which corresponds to a distinguished place in the real environment [22], [23], such as the maps of the subway placed above the wagon doors. The respective topological map for the metric one shown in Fig. 2(a) is depicted in Fig. 2(b). The most recent approach is the topometric mapping, comprising a combination of topological and metric mapping, as shown in Fig. 2 (c), which facilitates faster and more accurate robot localization. An early sample of this approach introduced a topometric method with the aim to reconstruct robot's path in a hybrid continuous-discrete state space, which combines metric and topological maps [24]. In a more recent method the problem of SLAM is addressed by combining visual loop-closure detection with metrical information available at a real-time created topometric map of an unknown environment [25]. In spite of the fact that all the methods developed so far proved to be adequate to afford robot navigation to specific target positions, they lack high-level attributes suitable for operating in typical environments. Therefore, the turn into the construction of anthropocentric maps endowed with cognizant capacities was inevitable, since the contemporary trend in robotics is to design agents to operate in human environments close to living beings.



(a) (b) (c)

Figure 2: a) An example of a 2D metric map of an explored indoors environment, b) the respective topological graph and c) the hybrid topometric map, where each node in the topological graph is registered with a spatial specific region of the occupancy grid

3. Trends in Semantic Map Building

3.1. Scale Based Categorization

Owing to the fact that in many occasions a semantic map is built on top of a metric one, a straightforward clustering of the existing techniques can be based on the scale the underlying method retains. Thus, semantic mapping paradigms have been employed both in indoors and outdoors cases. Moreover, the methods developed for indoors situations are further distinguished into single-scene and large-scale ones. The single-scene class gleans those methods that reason about an instance frame with respect to a local coordinate system, also providing conceptual attributes about the observed objects of the scene. Besides, large-scale approaches progressively construct a metric map, with respect to a global coordinate system, simultaneously annotated with high-level features, such as object types, place labels and shape interpretation. Concerning the outdoors methods, it is noticeable that there is hardly any single-scene one. A summary of the scale based clustering concerning the methods reported in the literature -presented in the subsections bellow- is illustrated in Fig. 3, where it is noticeable that the large scale indoors cases are the ones on which most ink has been spilled.

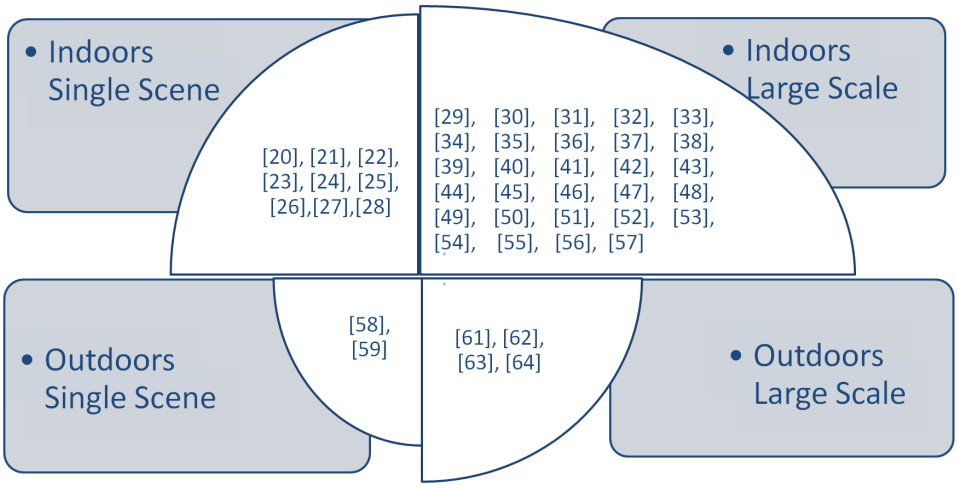


Figure 3: Diagrammatic representation of the scale based categorization

3.1.1. Indoors Single Scene Interpretation

Nielsen *et al.* [26] approximated the semantic mapping as an interface between robots and humans. They introduced a single-frame snapshot application as a means to grab real-world pictures and store them with the aim to augment a metric map. In particular the map advancement is accomplished by means of icons or symbols, thus providing signification of places and objects of interest. In an early work presented by Kostavelis et al. [27] an SVM based memorization algorithm has been utilized to semantically infer about the traversability of the scene. This work proved suitable to operate in indoor post-disaster environments as the semantic inferences was further taken advantage by a local path planing algorithm. The methodology described in [28] utilizes stereo vision and operates on the image plane with the purpose of classifying the traversability of the scene. Note that this work exhibited remarkable performance both on indoors and outdoors environments. Rusu *et al.* [29] presented a domestic robot equipped with a stereo camera and a SICK laser scanner, able to reason about objects in a kitchen. In this method, the individual sensorial inputs are fused to obtain the essential information about the perceived environment, while the robot learns from demonstration. Viswanathan *et al.* [30], [31] proposed a solution to the visual place recognition problem by utilizing the LabelMe dataset [32].

]

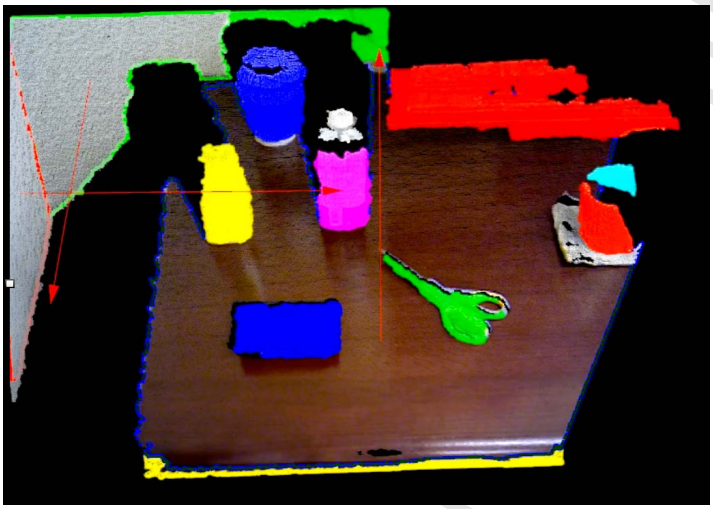


Figure 4: An example of indoors single scene interpretation [33]

The latter is an online database being consisted of user annotated images. In LabelMe, the user can annotate an object in an image by selecting a polygon region and associating it with an appropriate label. In a more recent work, Trevor *et al.* [33] introduced a single scene point cloud segmentation utilizing connected components practices through RGB-D data. Firstly a planar segmentation step is performed on the point cloud data to distinguish the dominant planes in the scene. Then an L2 norm based clustering and a connected component labeling mask are applied on the color image, in order to detect objects on a tabletop, as shown in Fig. 4. Swadzba *et al.* [34] introduced spatial 3D feature vectors suitable for individual scene classification, that operate on pre-captured frames. In another single scene interpretation work, Mozos *et al.* [35] employed the Microsoft Kinect sensor for visual place categorization. Moreover, the authors in [36] utilized visual input to infer about the category label of the detected objects during the robot's perambulation. This piece of information was treated in a hierarchical fusion manner to further to characterize the observed scenes in accordance with the existing objects.

3.1.2. Indoors Large Scale Interpretation

Considering the indoors large-scale interpretation methods one can distinguish them on the basis of the sensor and the strategies utilized to construct the metric map. Accordingly, the works described in [37], [38], [39], [40] and [41] employ laser scanners mounted on mobile robots to reconstruct the 3D environment. More precisely, Nüchter *et al.* [37] exploited a SICK laser scanner to capture a 360^o map of the scene. The correspondences of the successively acquired point clouds are established via semantic labels and are then registered by the *iterative closest point* (ICP) algorithm to obtain a globally consistent map. In a similar fashion, Blodow *et al.* [38] made use of progressively acquired laser scans merged with a 2D-3D registration routine to form the metric map. Segmentation techniques were applied to generate initial hypotheses about the significance of objects, such as furniture drawers and doors. Rusu *et al.* [39], [41] augmented the geometric map by processing large input datasets and extracting relevant objects. The objects modeled are kitchen serviceable ones, such as appliances, cupboards, tables, and drawers, being of specific significance for a domestic assistant robot. Hokuyo UTM-30LX measurements, combined with a rotary unit and the odometry estimations were used in [40] to construct the 3D map of the explored environment. The feature based map involves information concerning the locations of horizontal surfaces, such as tables, shelves, or counters, detected in the 3D point clouds, an example of which is depicted in Fig. 5, where the representative planar areas are shown in different colors. Additionally, Trevor *et al.* [42] utilized the GTSAM method to produce a metric map of the explored environment. This method defines a variety of feature types that can be used for SLAM and semantic mapping. In a swarm-robotics paradigm that produces semantic inferences about the explored environment the authors in [43] utilize a laser scanner to produce a 3D metric map.

Both the methods described in [3] and [44] employed RGB-D sensors to obtain the 3D map of the environment. In the first one [3] a hierarchical strategy was applied to create a global consistent 3D metric map. Firstly, the consecutive acquired point clouds are merged using visual odometry, followed by a refinement step based on ICP alignment of the dominant planes. Then, a *bag of features* technique along with a *support vector machine* (SVM) is applied to accurately recognize multiple dissimilar places, as shown in Fig 6. The second RGB-D based method [44] adopted the SLAM6D toolkit to register the subordinate point clouds into a consistent full-scene point

cloud.

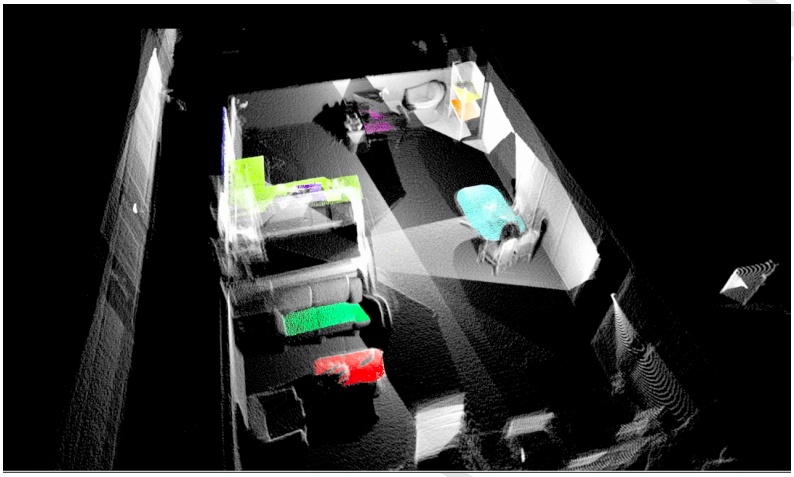


Figure 5: An example of indoors scene interpretation with annotation of serviceable objects [40]

Similarly to [40], the feature based map employs different types of furniture and their poses, while in this case the results are enhanced by the respective *computer aided desing* (CAD) models of the furniture. Civera *et al.* [45] applied a monocular SLAM algorithm in order to create the metric map of the perceived environment. The proposed algorithm runs an *extended Kalman filter* (EKF) monocular SLAM parallel to an object recognition thread, which is utilized to semantically annotate the scene.

An additional cluster of indoors large scale semantic mapping methods is the one utilizing laser scanner to form 2D occupancy grids modeling with the aim to explore the environment. Mozos *et al.* [46] utilized two different robots equipped with SICK laser scanners and simulated the laser scans in the different maps by using the software CARMEN [47]. Moreover, the method used AdaBoost to boost simple features extracted from range data into a strong classifier. Furthermore, in [2], [48], [49] and [4] geometric primitives from laser range scans are extracted and an EKF is applied for the integration of feature measurements. The authors in [50] utilized a robot equipped with a 2D laser scanner to build an occupancy grid of the environment by means of a standard SLAM method and retained it as a basis to build the semantic model upon it. The geometric features employed in all these approaches are lines,

which typically correspond to walls and other straight structures appearing as a line segment at the height of the laser scanner. A graphical model

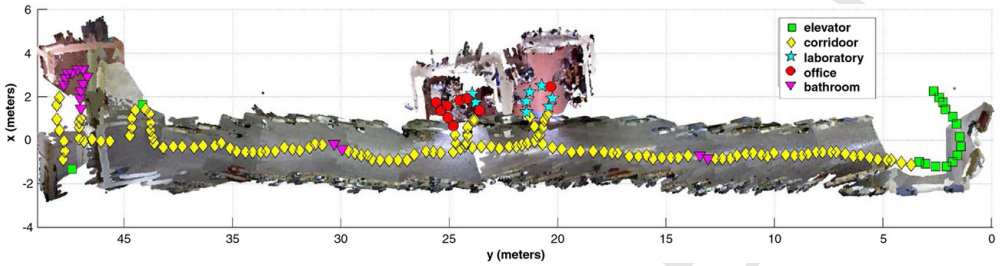


Figure 6: An example of indoors large scale interpretation [3]

was implemented in [2] to represent the semantic information and achieve inferences, while in [48] reasoning is accomplished by an SVM based cue integration scheme. In [49] the generated map was augmented by local and global information about existing objects. In a similar fashion, the method described by Zender *et al.* [4] recognizes places and objects by means of laser and visual data, respectively, with the aim to enhance the metric map constructed. Additionally, the work described in [51] captures laser scans to detect transitions on the map, which are then combined in a global metric map under a loop closure detection rule. More recently, Luperto *et al.* [52] utilized laser scanners placed back-to-back so that they cover an area of 360^o around the robot. This metric map was then employed during the semantic portioning of the explored area. Pronobis *et al.* [53] presented a multi-layered semantic mapping algorithm combining multiple visual and geometrical information. The metric map is build by exploiting the M-space feature representation.

A further cluster of the same indoors category consists of research endeavors exploiting stereo vision to acquire the depth information of the scene, subsequently utilized to solve the SLAM problem, thus obtaining a global and consistent metric map [54], [55]. Particularly in [54] the map generated by SLAM is augmented by object labels, recognized by means of SIFT features. In [55] the SLAM based map is extended by including text detection at an office environment annotated with text labels, such as room numbers and the names of office occupants. Moreover, Nieto-Granda *et al.* [56] employed the SLAM mapping module embedded in *robot operating system* (ROS) [57] - which is based on the Rao-Blackwellized particle filter technique- to partition

the resulting map into labeled spatial regions by means of a Gaussian model. Feng *et al.* [58] proposed a framework for mobile robot localization in an indoor environment, using concepts like homography and matching borrowed from the context of stereo and content-based image retrieval techniques. The work described in [59] utilizes a Visual SLAM system to create a long range metric map, which consists of 3D locations of distinct features observed by the robot during its traverse.

Finally, it is worth mentioning a number of other papers that focus solely on the semantic mapping problem and, therefore, they utilize metric maps existing in the literature [60], [61], [62], [63]. The authors in [60] suggested a method to extract an outlined floor plan of the maps using Bayesian reasoning, resulting to a probabilistic generative model of the environment, defined over abstractions. This work is also augmented with a rule-based context knowledge reasoning. Galindo *et al.* [61] coded the semantic knowledge by means of the, so called, *conceptual hierarchy*, in which the two general categories are objects and rooms. Fasola and Matarić [63] presented a method which permits to the service robots the communicate with humans by natural language and, towards this direction they suggested a semantic field model of spatial prepositions allowing the representation of dynamic spatial relations. In a similar manner the authors in [64] assumed CAD models of the environment scaled accordingly, however they exploited the robot's trajectory in order to semantically annotate the explored places. In [62] semantic-topological maps for indoor environments are produced by using a wearable catadioptric vision system.

3.1.3. Outdoors Interpretation

Multiple approaches have already been proposed to confront with the problem of semantic mapping in outdoors scenarios. Some of these works operate in an open loop fashion inferring semantic attributes about the observed scenes. The methodology described in [28] utilizes stereo vision and operates on the image plane with the purpose of classifying the traversability of the scene. The work presented in [65] exploits multiple sensors to analyze the scene in terms of its primitives e.g. ground, vegetation, structures, obstacles, etc. In a more sophisticated method introduced in [66], the authors perform large-scale 3D mapping of the environment. More precisely, the street level images are automatically labeled using *conditional random fields* (CRF) operating on stereo images, while the estimated labels are aggregated to annotate the 3D volume in a robust fashion. Additionally, the authors in

[67] applied supervised multiclass *Gaussian process* (GP) classification on 3D point cloud data, targeting on the semantic interpretation of the scene. In more detail, a feature extraction and segmentation is applied on the 3D point cloud and, then, the feature vectors are fed in a latent function represented by a GP which is the kernel of the classifier. The main advantage of this method is that when the 3D point cloud becomes denser, the uncertainty of the scene objects already categorized is diminished. Steder *et al.* [68] presented an algorithm based on 3D range data suitable to reliably detect previously seen places of the environment and, at the same time, calculate an accurate transformation between the corresponding scans. It combines *bag of words* (BoW) for the loop closure detection and point-feature-based estimations of relative poses to determine a consistent metric map of the environment. This method exhibited remarkable results both on ground and aerial vehicles. Similarly, in [69] the generation of semantic labels for the street images by directly segmenting the frames is reported. The output is then aggregated with multiple successive frames to produce a large-scale semantic map, as shown in Fig. 7. Singh and Kosecka [70] utilized Lady-Bug multi-camera system for long range semantic mapping of street scene imagery. They clustered the outdoor scenes into specific regions with their respective labels. They introduced an informative feature which characterizes the layout of the perceived environment, while they trained a classifier to recognize street intersections in urban inner city scenes. In the method described in [71] an *unmanned aerial vehicle* (UAV) was utilized to draw semantic inferences from the observations on the ground. More precisely, an online gradient boost algorithm is designed to interactively interpret context dependent detectors used in a video-domain adaptation method operating on on-board-camera images. Katsura *et al.* [72] proposed a vision based outdoor navigation method endowed with object recognition attributes. The great novelty of this system is that the recognition part remains robust to changes of weather and seasons. The authors proposed a comparison method in which the robot firstly recognizes objects in images using object models which allow appearance variations, and then compare recognition results of learned and target images, achieving a great generalization capability.

3.2. Topological Maps

Metric maps are organized in a geometrical arrangement, thus favoring the spatial information, while the associated conceptual one remains concealed. A way to reveal the hidden information is to organize it in terms

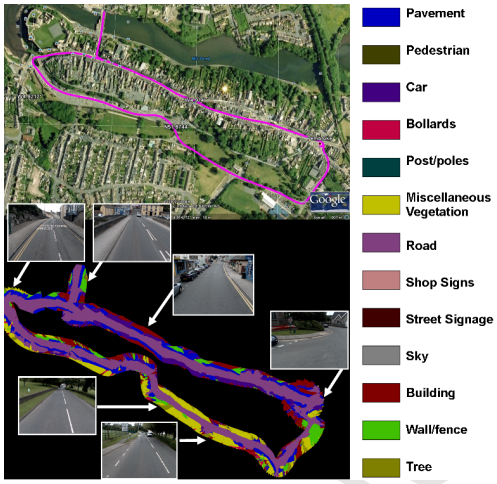


Figure 7: An example of outdoors semantic mapping reconstruction [69]

of a topological map, i.e. a graph, the vertices and edges of which should correspond to locations and the pathways among them, respectively. The term topological map derives from *topology*, the branch of mathematics that studies shapes and spaces, in particular the properties of space remaining unchanged under continuous deformations. Consequently, by means of a topological map, the environment should be structured so as to retain both geometrical information for the arrangement of the places visited and conceptual information about the category they belong. Therefore, such graphs constitute one of the fundamental characteristics in semantic mapping, due to the fact that they enable abstraction to metric maps as well as to the conceptual cues. Additionally, topological maps can be either unconstrained -retaining spatial and semasiological characteristics- or they can posses several constrains in accordance with their semantic attributes, such as the uncertainty about the currently visited place or the transitioning feasibility among the mapped areas (see Fig. 1).

Considering the *unconstrained* topological maps, each node is added to the map whenever the robot has traveled a certain distance or the inference mechanism has converged to a specific class label. In [61] a hierarchical semantic mapping method was introduced, where the link between spatial and semantic information is established via anchoring. Additionally, in [51] a graph is constructed at the top level of the metric map. Each graph's vertex

stands for a semantic construction or label, such as a room or a corridor, whilst the edges represent a transition spot, for example a doorway, linking the two semantic entities. This method utilizes passage detection, seeking to differentiate among the explored places. Ranganathan and Dellaert [9] employed object recognition to form object maps with metric information. Further to this work Viswanathan *et al.* [73] developed a technique to recognize objects registered with a specific place label, e.g. "living room". The clustering of these objects with respect to their spatial arrangement results in a graph, the nodes of which correspond to remembrances of rooms. A similar architecture was proposed in [74], nonetheless, the limits of the annotated places are less consolidated. Akin to this notion, Nieto-Granda *et al.* [56] employed automated recognition and classification of spaces into separate semantic (Gaussian) regions and used such information for the generation of a topological map of the environment. Vasudevan *et al.* [54] introduced a hierarchical probabilistic representation of space based on objects. In this work, a global topological representation of places with object graphs serving as local maps was proposed. Forerunners of these works are the ones described in [2], [48], [53], [75] and [76], in which the topological graphs are endowed with probabilistic assertions, resulting thus in more intuitive semantic maps, as shown in Fig. 8. Moreover, the work described in [3], introduced a semantically annotated topological graph, which relies both on geometrical and cognitive constraints to differentiate among multiple rooms of the same label. The methodology reported in [77] proposes the construction of semantic maps retaining global landmarks similar to those of the human cognitive navigation mechanism.

Considering the *constrained* topological maps, they are typically further abstracted by navigation graphs. The latter are conceptual representations of the semantically annotated topological graphs expressing the connectivity and the transition feasibility among the detected places. Such an approximation is the one described by Mozos and Burgard [78], presenting a method that couples the topological maps with the semantic information of the explored places. Specifically, Adaboost has been utilized as a supervised learning algorithm in order to classify the metric map into semantic classes, e.g. corridor, room, etc. Afterwards, a probabilistic segmentation step that filters out the classification errors is applied and the topological map results by combining the geometric and the semantic knowledge. This map is expressed as a graph the nodes and edges of which correspond to the semantically annotated regions and their connections respectively. Moreover, the

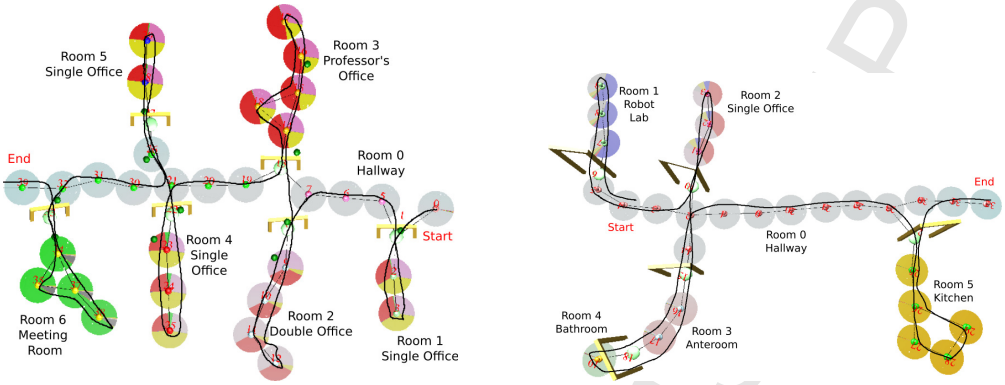


Figure 8: Topological maps -including probabilistic assertions- of the environment associated with a metric map [2]

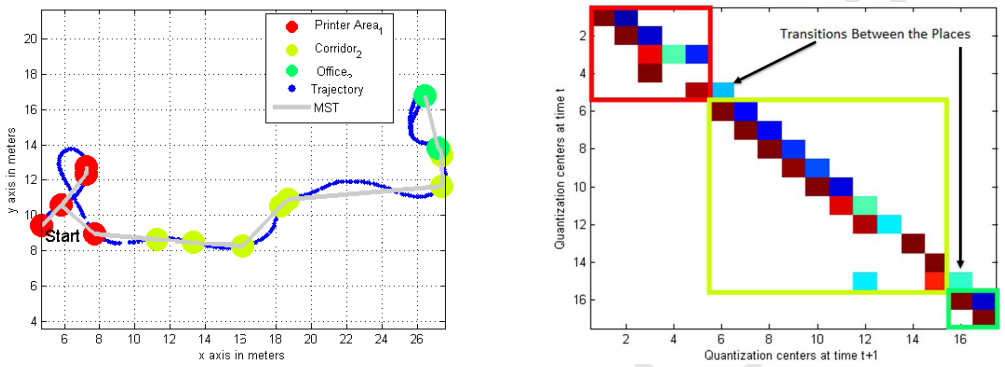
authors in [46] employed *hidden Markov models* (HMM) to semantically infer about the nodes in the topological graph. Additionally, in [4], [49] and [79] the metric line map turned into a sparse navigation graph by exploiting the robot's trajectory. Each node is dropped following specific geometrical constraints, while at the same time it bears a specific place label. The doors and passages comprise separate nodes, while the entire map is further abstracted by the deduced labels in a higher level navigation graph. Last, in a more contemporary work [80] a sparse topological map was introduced, in which each node is accompanied by a place label. HMMs were also utilized to build an augmented navigation graph retaining physically constrained connectivity information for the recognized places during the robot's exploration.

Moreover, in an outdoors scenario [70] the authors exhibited how the evidence of different semantic concepts can induce useful topological representation of the environment, which can aid navigation and localization tasks. In a paradigm introduced in [62] the authors employed a wearable catadioptric system for semantic labeling of topological maps by grouping the formulated Markov models.

3.3. Temporal Coherence

Temporal coherence is another common attribute among the semantic mapping methods. It can be taken into account either during the construc-

tion of the metric maps or during the conceptual formulation of the models. A paradigm of the conceptual formulation of the model that exploits temporal coherence is the work presented by Ranganathan [81]. The author introduced the *Place Labeling through Image Sequence Segmentation* (PLISS) method which retains two major attributes: (i) operation on video streams to exploit time proximity of the frames and to infer about the semantic attributes of the scene and (ii) the ability to detect places for which no prior knowledge is available. In this work the place modeling is performed by exploiting a GP classifier and the resulting uncertainty is further utilized for the detection of the unknown labeled places. The change point detection is performed by applying a projection of the Multivariate Polya measurement model to a lower dimensional space ensuring speed and accuracy simultaneously. A pure place recognition method [82] suitable for semantic mapping worths to be mentioned here as it employs BoW integrated with temporal attributes. Having in mind that the robot acquired frames share great time proximity such a method could boost the performance of place recognition during robot navigation. Moreover, in an example described in [38], temporal difference registration is used to segment out the front furniture faces, during the point cloud annotation of the global metric map. Moreover, Cadena *et al.* [83] introduced a system which examines the temporal consistency of the visual memories, to converge in a loop closure detection, thus facilitating robust metric mapping. On the other hand, in the methods described in [2] and [48] it is supported that the developed systems should posses integration over space and time, due to the fact that the information acquired at a single point rarely provides enough evidence for reliable categorization of places or objects. Therefore, a voting procedure was considered in these work during the inference of learning mechanisms. In a different work. the authors in [62] utilized the result of the place labeling under a probabilistic model to account for temporal consistency along the robot's trajectory. In a more sophisticated manner in [46] and [80] HMM were utilized with the aim to record the temporal proximity and the physical transitions among the explored places. Specifically, in [80], the temporal proximity has been utilized in two phases: (i) during the system's inference concerning the place currently visited, considering a neighborhood of the observed frames; (ii) during the robot's advancement from a specific recognized place to the next one, thus resulting to a physical constrained augmented navigation graph. A visualization of the time proximity, by means of the time adjacency matrix, is provided in Fig. 9. Additionally, Kostavelis *et al.* [84] exploited the time



(a) (b)

Figure 9: a) The topological graph superimposed over the robots trajectory expressed as the minimal spanning tree among the existed nodes; b) the normalized time adjustency matrix, where the partitioning of the quantization centers according to their class label and the transitions among the different groups are illustrated (adopted from [80])

proximity of the acquired frames during the robot's motion. Towards this direction, a temporal window was determined within which a voting procedure during the SVM inferencing mode was performed to label the observed scenes.

3.4. Cue Based Perception

A primitive intent of the semantic mapping methods is to express the surroundings in terms of human signs. That said, such methods take advantage of the plentiful research conducted in the area of place and object recognition and categorization, producing remarkable results, such as those described in [85], and [86]. The solutions produced by such investigations shown to be adequate to be applied on real robots, in order for them to enhance their navigation and mapping competences. Contemporary mapping algorithms are accompanied with place and object recognition capabilities, allowing them to draw conclusions about the observed scenes. This fact comprises a basic solution to the semantic mapping issue and the methods developed might integrate single or multiple cues, with the aim to converge in a more consistent solution. A categorization of the semantic mapping

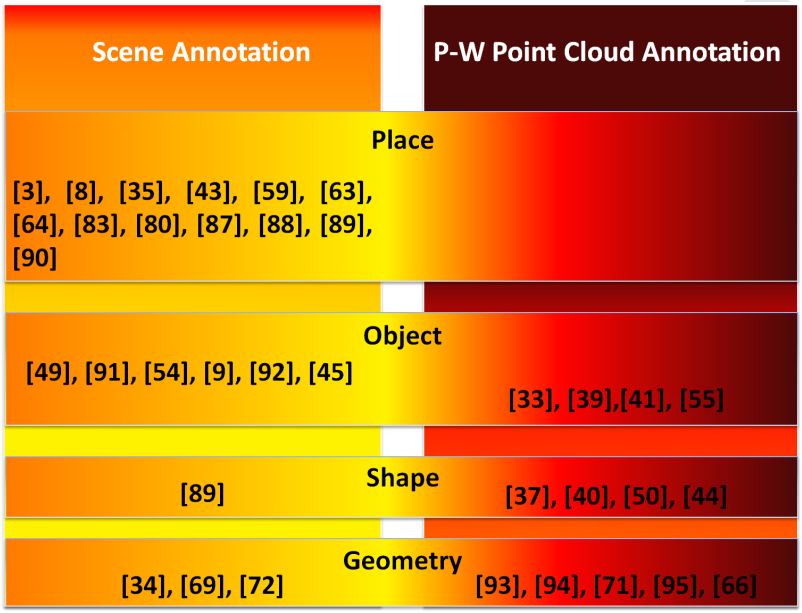


Figure 10: Perception based on type of cue

methods according to the type of cues employed is outlined in Fig. 10.

3.4.1. Inference by Single Cue

In this group we have gathered all those methods sharing the common characteristic to make deductions about environment elements based on a sole perceived component, e.g. place, area or object. The output of such a component is typically employed to annotate the areas observed by the robot. According to the sensory input and the architecture employed, this bunch of methods can be further clustered into *scene annotation* and *pixel-wise point cloud annotation* based, a summary of which is presented in Fig. 10. A preliminary statement is that the pixel-wise point cloud annotation methods may directly recognize a place, albeit, they take advantage of the ample information, captured by the respective sensory mechanism. Thus they have to memorize knowledge in a more anthropomorphic way, i.e. type of objects, their characteristics, their functionality, etc.

Scene Annotation: In many works proposed so far, the place recognition is utilized as a sole attribute to semantically augment a metric map. Pronobis *et al.* [8] proposed a discriminative approach for place recognition

based on visual information. A global descriptor operating directly on the robot acquired images assisted by an SVM produced meaningful appearance based interpretations. This method was engaged on a robot platform and tested for indoors scenarios yielding noticeable results on the place recognition problem. Towards the same direction, the authors in [87] combined an incremental extension of SVMs with a method that reduces the number of support vectors needed to build the decision function, without any performance loss. Moreover, in [83] a stereo vision place recognition algorithm was presented, which considers both appearance and geometric information of points of interest in the images. Emphasis is given in the loop closure verification, which is addressed by comparing CRF. In [35], the authors transformed the depth and color images into histograms of local binary patterns, the dimensionality of which is further controlled by a uniform criterion. The performance of these histograms is examined both on SVMs and random forests in a supervised fashion and proved to be adequate for remarkable place recognition. Wang *et al.* [64] focused on the development of a specific descriptor to visually recognize places. This descriptor integrates image features and color information via the *convex hull census transform* then utilized for the supervised training of SVMs. This method is appropriate for indoor scenarios and operates with images captured by omnidirectional cameras. Feng *et al.* [63] extracted interest points from contrast images to provide perceptually consistent measurement of image content. The image matching is performed by modeling the robot's behavior under cross correlation to determine the corresponding points of interest, which is followed by a RANSAC homography optimization criterion. Fazl-Ersi and Tsotsos [88] utilized *histograms of oriented uniform patterns*, providing strong discriminative capacities and allowing to solve the place recognition problem in a competitive manner. The authors in [59] proposed a method to label areas in a pre-built map using information from camera images. This method labels the area that is viewed in the camera image rather than just the current robot location. The labeling of the viewed scenes is integrated in a CRF which also considers adjacency and place boundaries. Additionally, in [3] a novel appearance based, histogram-oriented approach is recommended, with the aim to address the place categorization issue. The same solution has been employed for the place categorization task in [80], in order to build semantically annotated topological maps that exploit the temporal coherence of the produced labels during the robot's exploration. In an outdoors scenario, Weiss and Biber [89] performed a rough analysis of the scene with the purpose of building features

to be utilized for place classification. This method proved capable to semantically annotate the observed scene with predetermined types of places. In a swarm-robots oriented method [90], the authors suggested a solution for the automated assignment of target locations to the individual robots. They applied a classifier learned with AdaBoost, which additionally considers spatial dependencies between nearby locations, to determine the type of a place a robot should head to. Exploration within topometric and semantic maps with swarm-robots was also addressed by Cowley *et al.* [43]. In this work, highly cluttered environments can confound exploration strategies that rely solely on occupancy grid frontier identification and classification methods keyed on geometric features.

Concerning the object recognition, one straightforward technique is the one implemented in [49], targeting to build a robot that localizes its position on a metric map, recognizes objects on its way and assigns them on the map. In a more sophisticated approach [91], the authors developed an object recognition algorithm supported by saliency attentional model. More precisely, in order to perform successful recognition in a real world scenario, they combined a peripheral-foveal vision approach, a bottom-up visual saliency with structure from stereo and a metric map. The results of the object recognition process on the computed map are shown in Fig. 11. In [54] a hierarchical probabilistic representation of space based on objects was suggested. A SIFT-based object recognition system was adopted, while a stereo camera was used to capture the object and obtain its coordinates on the metric map. Ranganathan and Dellaert [9] presented a model for places using objects as the basic unit of representation. Stereo range data was employed to compute the 3D locations of the objets. Moreover, the Swendsen-Wang algorithm, a *Markov chain Monte Carlo* cluster method, was adopted to solve the correspondence problem between image features and objects. Jebari *et al.* [92] applied a BoW to detect and recognize the various obstacles in the scene. After an object has been recognized it is associated with the semantic map constructed. Additionally, Civera *et al.* [45] employed an object recognition strategy which informs for the presence of an object in the sequence by searching for SURF correspondences and checking their geometric compatibility. The recognized objects are inserted on the exact position measured on the metric map and hence refined by the SLAM algorithm in subsequent frames.

In an outdoors scenario [69], a method for producing a semantic map from multi-view street-level imagery is analyzed. The street images are augmented

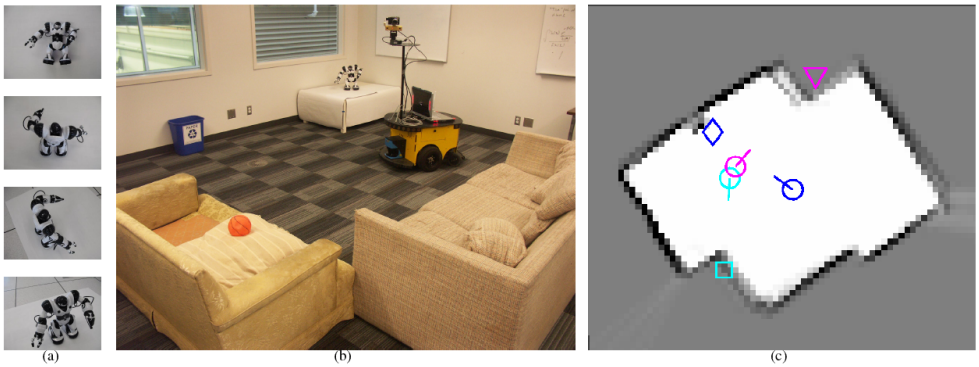


Figure 11: An example of combining the metric map with object recognition, by assigning meaningful object labels to locations on the map: (a) exemplar training data for object robosapien; (b) a view of the explored environment and (c) the resulting map annotated with three objects and the respective locations from which they were observed (□ basketball, ⋄ recycling bin, ▽ robosapien) [91]

with labels of objects, such as cars, pavements, buildings etc, producing a semantic map. It exploits CRF firstly to model the image segmentation in each frame and then to combine the segmented parts over a set off consecutive images defined over a ground plane. Additionally, Swadzba *et al.* [34] introduced 3D spatial feature vectors i.e. the Gist features, which found to be suitable for individual scene classification. In a more generalized model the authors in [72] proposed recognition of entities such as sky, buildings, ground, etc. during robot navigation, by characterizing the detected regions based on the variations in the uniformity, achieving thus robustness even under different weather conditions.

Pixel-Wise Point Cloud Annotation: This bunch of methods exhibit the common characteristic of semantically partitioning and labeling arranged point clouds which are either acquired directly from laser scanners and RGB-D sensors or computed from stereo images. In all the cases, the 3D information can be expanded in scale, so as to facilitate robot's localization. Trevor *et al.* [33] focused on the detection of objects placed over a tabletop, yet this method retains deteriorated semasiological information of the observed scene. Nüchter *et al.* [37] matched the acquired 3D laser scans with semantic features. The basic idea of labeling 3D points with semantic

information is to use the gradient between neighboring points to differentiate between three categories, viz. floor-, object- and ceiling-points. These features are then used for the accurate registration of consecutive full-circle scans. In [39], [41] emphasis is given on the semantic object annotation of 3D point cloud data in kitchen environments. It is assumed that the majority of the surfaces in a kitchen are planar, thus the 3D point cloud is segmented into planes. The geometrical features are computed to be further utilized for object class learning achieved via CRF. In a similar work [40] a feature based mapping technique is introduced including information about the locations of horizontal surfaces on the map. The resulting scans of surfaces are analyzed and segmented into distinct surfaces, which may include measurements of a single surface across multiple scans. The authors in [50] proposed a method to extract an abstracted floor plan from typical grid maps using Bayesian reasoning. Through this procedure a probabilistic generative model of the environment defined over abstract concepts is revealed. Multiple *random sample consensus* (RANSAC) steps are utilized to over-segment the set of point clouds, while the matching of the sought objects obeys geometrical constraints. Günther *et al.* [44] introduced an RGB-D based system to reconstruct the surfaces in the point clouds, detect different types of furniture and estimate their poses. The result is a consistent mesh representation of the environment enriched by the CAD models of the detected pieces of furniture. In [55] *Hough forests* -a variant of random decision and regression trees- were utilized, thus enabling pixel categorization and voting for 3D object position and orientation. Cadena and Košecka [93] parsed the indoor environments into four categories, namely ground, structure, furniture and props. Accordingly, they achieved separation of the instances encountered in the scene as objects and non-object categories. Once again this method is based on the CRF for the semantic segmentation.

In the past, several investigations concentrated in outdoors scenarios have utilized *pixel-wise point cloud annotation* techniques. For example Wolf and Sukhatme [94] combined SVM with HMM to produce terrain and activity-based maps. The acquired point clouds were segmented and annotated in terms of their navigability. Additionally, the work introduced in [71] comprises a semantic mapping of data captured by a UAV. Once the metric map is created, a set of classifiers designed for specific targets using online gradient boost is utilized. These classifiers are geometrically adapted to the onboard camera image domain and sent-back to the flying UAV. Another outdoors scenario [95] make use of stereo image pairs to produce dense depth maps.

In turn, these are combined into a global 3D reconstruction, using camera poses from stereo visual odometry. At the same time, the 2D semantic segmentation using a nonparametric scene parsing method is fused into the 3D model. In a similar method [66], the depth-maps -generated from the stereo pairs across time- are fused into a global 3D volume in an online fashion to accommodate arbitrary long image sequence. The street level images are automatically labeled using CRF tactics exploiting stereo images, while the label estimates are aggregated to annotate the 3D volume, as depicted in Fig. 12.

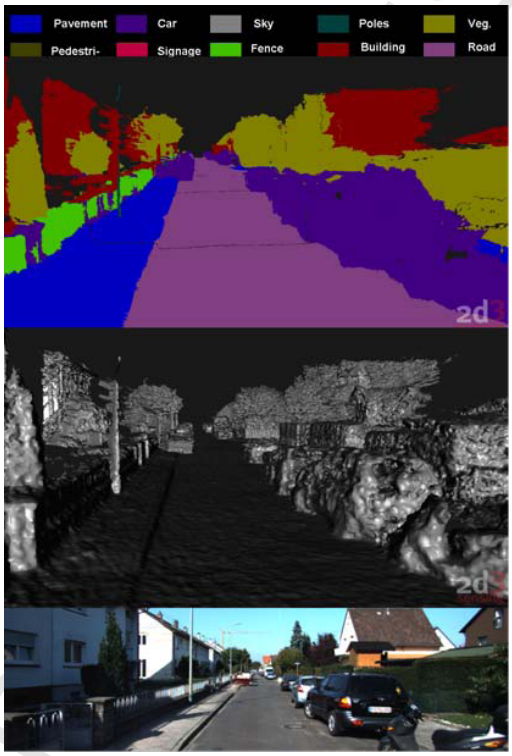


Figure 12: An example of pixel-wise point cloud annotation in outdoors semantic mapping reconstruction: the top row depicts the dense 3D semantic reconstruction with the associated class labels; the middle one shows the dense surface reconstruction, whilst the bottom one is one of the input images [66]

3.4.2. Inference by Multiple Cues

While many techniques were presented in the previous paragraph relying the inference on a single cue, there are also a sufficient number of applications that exploit multiple cues to draw semantic conclusions. These methods might either combine different modalities to apprehend the robot's surroundings or exploit multiple perception methodologies of the same sensory input to deduce about the observed scenes. The existing methods combine human concepts such as types of places, objects and even the geometry of these two, resulting to metric maps augment with semasiological information. An example of a method utilizing visual data to produce multiple cues for semantic place recognition is the one described in [96]. A transfer learning tactic is described, affording the robot to automatically decide whether its internal knowledge is useful for a place, about which no prior knowledge is available. Krishnan and Krishna [51] employed two different modalities to produce semantic cues, namely 2D laser scans and color images. The laser scans allow detection of the transition regions in an area under exploration, while the color images are exploited for the place annotation and loop closure detection, through a BoW scheme. In a couple of papers [73], [76] visual place and object recognition (*occurrence frequencies*) are simultaneously utilized, with the aim to construct a spatial-semantic model. Specifically, in [76] the authors used the semantic map in order to search for the memorized objects under a plan based rule. Viswanathan *et al.* [30], [31] combined local and global characteristics of images in order to semantically infer about the observed scenes. This is an object-centric method which fuses global properties in indoor scenes and produces good results for place recognition. Furthermore, Ko *et al.* [77] constructed consistent semantic maps exploiting place categorization accompanied by object matching techniques. Geometrical information is also employed to augment the spatial relationships among the detected objects. Trevor *et al.* [42] produced semantic maps by employing a variety of feature types as landmarks, including planar surfaces such as walls, tables, and shelves, as well as objects such as door signs. This work also investigates the aspect where these landmarks can be optionally labeled by a human for later reference. In addition, Blodow *et al.* [38] exploited 3D laser data firstly to extract geometry information about the planes in kitchen environments and secondly to infer about the object labels within the organized point clouds. Moreover, in [74] an attempt to establish a link between the robot's sensors, the objects and the places was described. It

employed a place classification algorithm which -assisted by an object recognition one- empowers the semantic knowledge about a place, through a *naive Bayes classifier*. Similarly, Anand *et al.* [97] proposed a semantic labeling and search of 3D point clouds based on RGB-D data. Considering the case of RGB-D data Kostavelis *et al.* [84] utilized a visual vocabulary for place recognition and *hierarchical temporal networks* combined with an attentional model for object recognition. These two cues where fused under a decision making algorithm to semantically infer about the place that a robot stands. It applied 2D images to detect objects and the corresponding depth data to obtain results about their shape. The method discussed in [55] focuses on the creation of a semantic map with text data, including room numbers, people's names and other written descriptions associated with rooms and offices. The robot captures pictures of corridors and queries a trained text detector to classify image regions. By employing depth data the detected signs are placed with specific orientation within the metric map (see Fig. 13). In

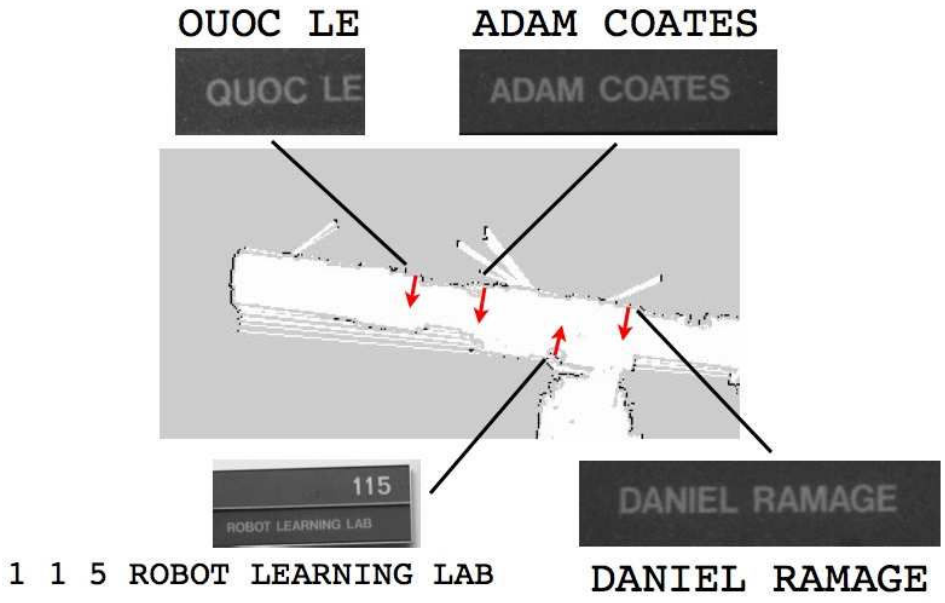


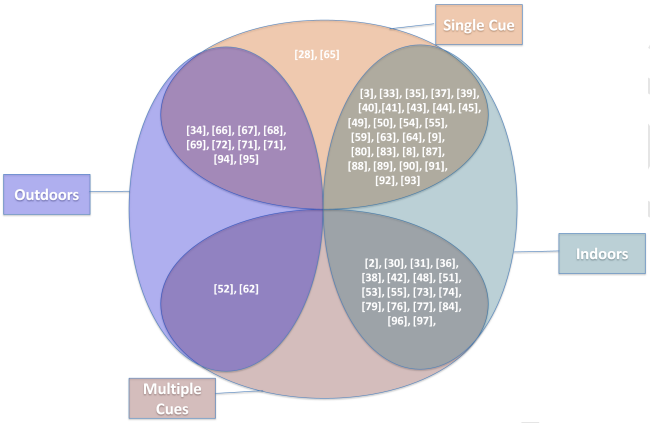
Figure 13: A metric map annotated with cues (semantic labels) read from signs posted on walls [55]

addition, the authors in [36] introduced a generative probabilistic hierarchical model, where object category classifiers were used to associate low-level visual features to objects, and contextual relations to associate objects to places. The detection accuracy was increased by using a 3D range sensor that allowed to the developed attention mechanism successful recall of the geometric and structural information. In more sophisticated works, such as those described in [2], [48] and [79] the authors utilized multiple sensors to make deductions about different characteristics of the scene. For example place and object recognition is performed through vision, while the shapes of the rooms are extracted by utilizing laser scanners. All the retained concepts are fused under generalized SVM models to produce probabilistic inferences about the explored areas. Similar tactics exploited also by Pronobis *et al.* [53] to build semantic maps that enclose information about the existence of objects in the environment with knowledge about the topology and semantic properties of space such as room size, shape and general appearance. In a different work, Luperto *et al.* [52] assumed that the buildings on which the robots are designed to operate have a specific *typology*. This method semantically labels portions of a metric map using different classifiers that are trained on data belonging to different building typologies. In this way, by knowing the typology of the building where a robot operates, the right classifier can be selected. Two different types of features have been utilized; those that describe the shape of a place and those that represent the structure of the building and the connections of the room with the rest of the environment. Of course, in order to obtain such information prior knowledge for the building is required. The authors in [62] proposed an interesting idea in order to augment the semantic information of a typical indoor topological map. In this work basic types of indoor regions has been considered, such as places and transitions, to semantically segment the environment.

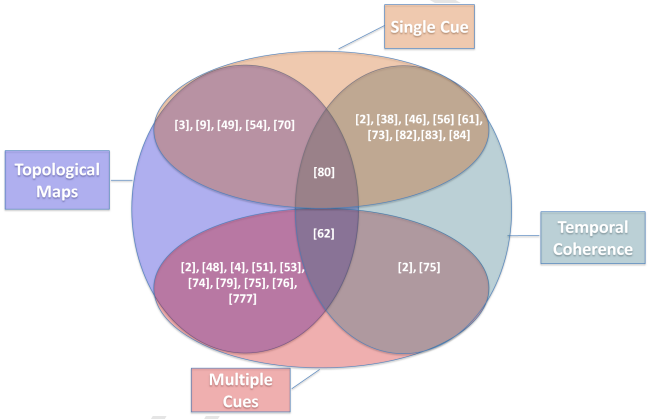
A summary of the categorization for the examined methods relative to the type of inference and the operation environment, as well as to the type of inference and the existence of temporal coherence or topological maps is illustrated in Fig. 14(a) and Fig. 14(b), respectively.

4. Applications

How do you put into practice semantic maps? This section intends to review the application areas reported hitherto. The principal ambition of semantic mapping, as outlined earlier in this survey, is to provide the robots



(a)



(b)

Figure 14: a) Diagrammatic separation of the methods according to the type of inference -single/multiple- and their operation environment -indoors/outdoors- and b) diagrammatic separation of the methods according to the type of inference -single/multiple- and the existence of topological maps or temporal coherence

with a depiction comprehensible by humans. Yet, with the aim to include a human in mobile robot tasks, remote information should be efficiently presented to human beings. The various works examined in this section aspire to cover this aspect from different perspectives. In addition, in the same

section we examine how the semantic mapping is applied in several integrated frameworks involving humans and robots. Knowledge representation schemata mixed up with semantic maps enable efficient task planning, path planning and navigation.

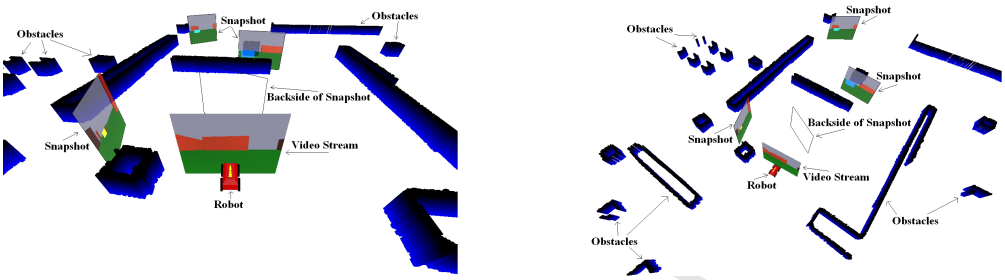
4.1. Human in the Loop

When the human operator enters the semantic map loop, it is either in order to augment the map building procedure by indicating objects or places, or to get assistance by the robot to find places or objects. In the first case, Nieto-Granda *et al.* [56] suggested a technique for automated recognition and classification of spaces into distinguished semantic regions, so as to produce a semantic map of the environment. In this technique, the relationship between semantic labels and spatial regions is based on *human augmented mapping* [98], which provides a parting of regions relating a user's view on the surroundings and detection of transitions between those regions. The setup specifying the regions is an interactive one, providing distinctiveness for space segmentation and localization purposes. A clear example proving that the semantic mapping conforms the HRI is the work presented in [4], where conceptual representations of human-made indoor environments are drawn, using mobile robots. More specifically, a situated dialogue is introduced, which is a functionality that enables the robot to carry out a natural language dialogue with a human [99]. The system retains the *combinatory categorial grammar* (CCG) parser in order to interpret spoken language. The latter is processed by the situated dialogue functionality, which examines the previously mentioned objects and events, and the previous utterances in terms of speech acts. Furthermore, a human-robot collaboration method is presented in [100], aiming to extract the 3D shapes of objects indicated by a human operator. In the same spirit Trevor *et al.* [101] introduced a semantic mapping framework which is augmented with HRI attributes. It examines the case of service robots that should apprehend human commands to fulfil fetching tasks. The user aids the construction of such a semantic map by labeling the detected surfaces. On the other hand, Dayoub *et al.* [102] presented an object-based semantic memory for service robots that enhances topological and metric maps. The overall objective of this work was to give the robot the means to assist humans by proposing the most likely locations of particular objects on the map. Sharing the same concept, Nielsen *et al.* [26] introduced a snapshot technology as an instrument to acquire real world images and store them in a semantic map, which relies on an occupancy-grid.

Moreover, the map is rendered by means of a mixed reality 3D interface allowing the combination of real and virtual elements of the robot's environment, enabling its presentation in an intuitive display, as shown in Fig. 15. Zender *et al.* [4], proposed an integrated system for generating conceptual representations of structured environments. The representation possesses different levels of abstraction, the information needed for each one deriving from different modalities, including a laser sensor, a camera, and a natural language processing system. The system is apt for high level of human-robot communication and conceptual representation. Moreover, the authors in [103] utilized the, so called, *autonomous city explorer* robot for semantic navigation into urban environments. This work employs several attributes such as outdoor localization, traversability assessment, path planning, behavior selection and topological abstraction. Several sensors are utilized such as laser scanners and stereo cameras for the metric mapping and the sensing of the robot's surroundings, respectively. The robot retains an additional human-robot interaction behavioral model according to which a behavior selection module is responsible for choosing the appropriate navigation behavior depending on the current situation. In the European project Viewfinder [104], [105] several methods for indoors and outdoors maps were analyzed and implemented, utilizing various modalities including stereo vision, laser scanners and olfactory sensors. The overall objective of the project was to make use of robotic systems to inspect the ground safety in the aftermath of a fire and to gather data -visual and chemical-, so as to assist first responders. Hence, the constructed 2D and 3D metric maps were annotated with the results of chemical sensors, as well as the output of the victim detection algorithms. The resulting semantic map was made available to the rescue squad via a sophisticated *human machine interface* (HMI) at the base station, as shown in Fig.16.

4.2. Association with Knowledge Representation Formalisms

Knowledge representation formalisms represent information about the world in a way that a robot can use to solve complex tasks [106]. It is therefore straightforward that semantic maps are ideal candidates to combine with. Wyatt *et al.* [107] figured that one of the basic attributes an artificial agent needs, in order to learn in a truly autonomous fashion, is the semantic interpretation of its surroundings. This work included representations of gaps and uncertainty for specific kinds of knowledge, and a goal management and planning system for setting and achieving learning goals.



(a) (b)

Figure 15: View of the mixed reality 3D display: (a) a zoom-in supporting egocentric tasks, e.g. navigation and object identification; (b) a zoom-out supporting exocentric ones, e.g. place recognition and global planning [26]

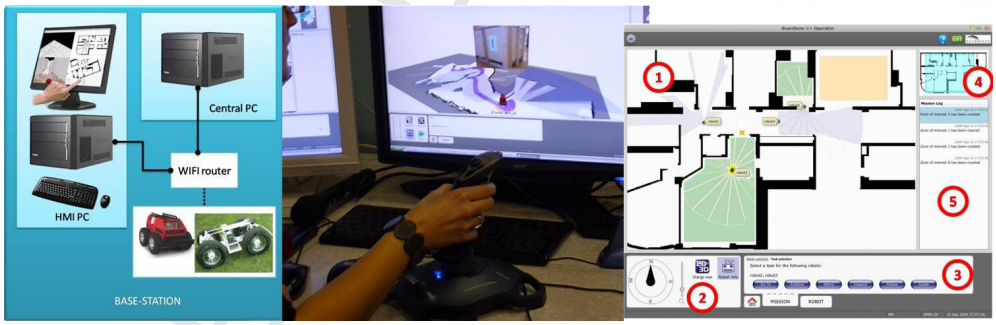


Figure 16: HMI to make available a robot generated semantic map to a firefighters team for search and rescue operations

Lim *et al.* [108] suggested an ontology-based integrated knowledge framework that puts together low-level data and high-level knowledge. It consists of a knowledge description and a knowledge association sections, as shown in Fig. 17(a). Whilst the knowledge description section comprises knowledge embedded in metric and semantic maps, the knowledge association make use of logical inference, Bayesian inference, and heuristics, which permit the robot to solve most queries set. An example of how the ontology schema and the ontology instance layers interweave with the semantic map, for a certain experimental space is shown in Fig. 17(b). Moreover, Bouguerra *et al.* [109] selected *description logics* among the knowledge representation formalisms to represent the semantic domain information. Semantic knowledge concerns objects, their properties and their relations with other objects. Although only limited knowledge was encoded, the paper proves that the semantic knowledge is apt for monitoring processes. Tenorth *et al.* [110] described a way the further enhance the produced semantic maps, by linking the recognized objects to encyclopedic and common-sense knowledge obtained from large, publicly available knowledge bases. The knowledge, organized in an ontological structure, provides the robot with information about what the objects are, what they can be used for, and how to use them. In a different scenario Finucane *et al.* [111] utilized the *linear temporal logic mission planning* toolkit to design high level robot controllers. This toolkit comprises structured English and linear temporal logic suitable for the development of high-level reactive task specifications, which are converted into correct robot controllers that used directly on real robots. Moreover, it is planned to be integrated with ROS framework aiming to interface with a legged robot with a manipulator in the near future. The authors in [112], discussed a solution to the problem of parsing natural language commands to actions and control structures that can be readily implemented in a robot execution system. This approach memorizes a parser based on example pairs of English commands and corresponding control language expressions. To fulfil such an attempt the authors converted the semantic attributes, perceived by the robot's sensory input, as binary features which end up to logical expressions. Liu and von Wichert [60] presented a method to produce a parametric abstract model of the sensed environment. Moreover, task-specific context knowledge was defined as descriptive rules in *Markov logic networks*, relying on predefined abstract terms (e.g. "type", "relation", etc). The corresponding reasoning results were utilized to establish an ex-ante distribution intending to add reasonable constraints to the solution space of semantic maps. Furthermore,

the semantically annotated sensor model exerted allows the interpretation of the context information of the sensor data. Recently, Heintz [113] proposed a practical framework for semantically grounded temporal stream reasoning. Incremental reasoning was attained via systematic progression of *temporal logic* formulae. The semantic reasoning was established by means of an ontology, which sustains reasoning and enables support for semantic mapping from multiple robotic systems.

4.3. Planning and Navigation

Semantic maps were utilized to solve practical problems in mobile robotics, including planning and navigation. In particular, Borkowski *et al.* [114] extracted semantic features from laser scanner raw data and used them to classify objects and to construct semantic maps of building interiors. Moreover, a semantic navigation scheme was introduced based on hypergraphs and shown how semantic information is derived from the digital model of a building, thus facilitating robot navigation. On the other hand, Galindo *et al.* [115] exploited semantic knowledge for robot task planning. A specific semantic map type was introduced, integrating hierarchical spatial information and semantic knowledge. The ability such a map provides to achieve task planning was also investigated in the same work, highlighting the capacity to conclude the current state with non-sensed information, the capability to automatically generate goals to maintain certain conditions, and the means to refine planning. Additionally, the authors in [116] proposed an approach for learning high level concepts from task-based human-robot dialogs. The dialog system encloses a probabilistic model that connects speech to the locations in the physical environment, a system which acquires knowledge that the robot does not know a priori, and a knowledge base which stores the acquired knowledge. The entire system has been evaluated by untrained people which directly interacted with the trained robot aiming to address simple planning and navigation tasks. Liu *et al.* [117] stated that a *network robot system* should integrate physical autonomous robots, environmental sensors and human-robot interactions through network-based cooperation. To this end, the authors introduced a cooperative service framework for networking robots ample to plan tasks compatible with human concepts. In a more futuristic work, the authors in [118] supported that the semantic knowledge can greatly increase the robot's capabilities. The information about how things should be (normal states) has been encoded in such a manner to enable the robot to infer deviations from these normal states and to generate goals to

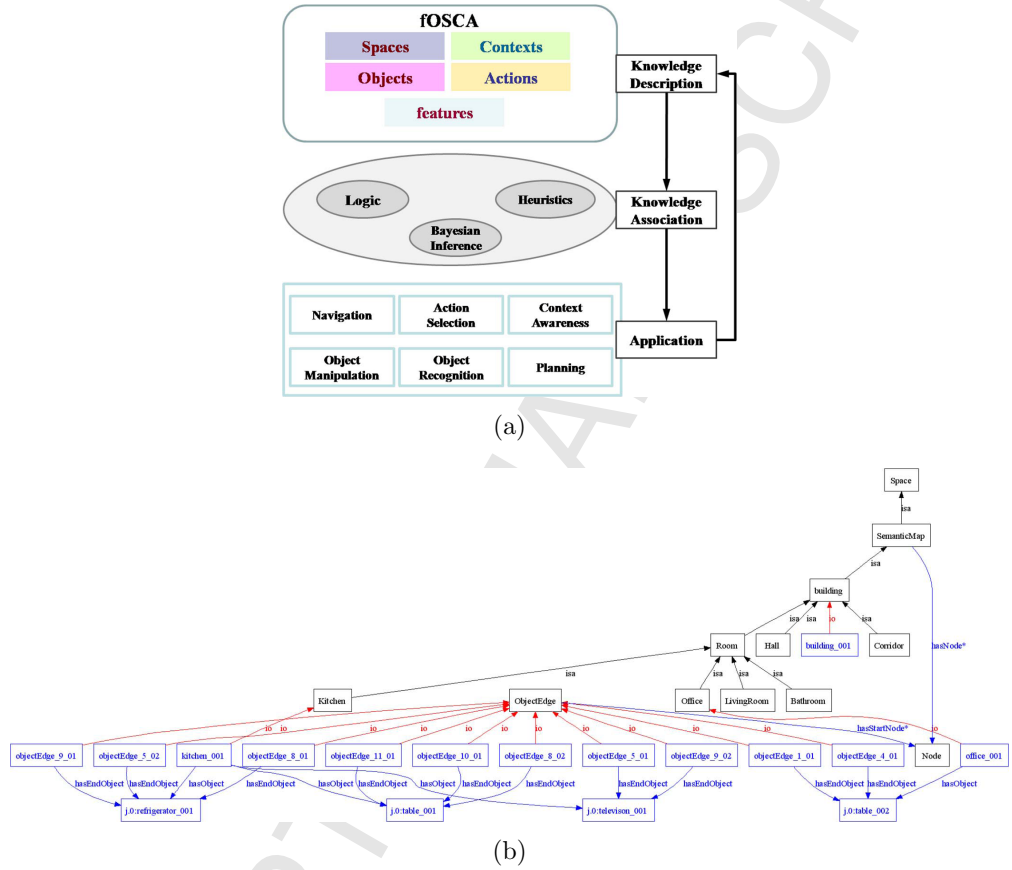


Figure 17: (a) The three parts of the system in [108], namely knowledge description, knowledge association, and application; (b) ontology schema and the instance of the semantic map for a specific place.

correct these deviations. The full-scale system integrates a series of novelties in a single mobile robot, namely a semantic mapping method, a knowledge representation and reasoning practice, a task planner, as well as perception and navigation routines.

5. Benchmarking and Validation Datasets

Any semantic mapping computation set-up constitutes a complex system comprising multiple subordinate modules. Such a system typically retains an increased number of parameters that influence the overall accuracy of the employed methods. That have been said, the demand of various realistic datasets to be utilized both for development and assessment of the produced algorithms is of great importance. Moreover, the establishment of objective metrics is awkward, should we bear in mind that it becomes nearly impossible to fairly compare solutions which are designed, implemented and evaluated in various environments, under different conditions and with dissimilar assumptions. Thus the development of benchmarking real world datasets is an one-way street. Wu *et al.* [119] introduced a consistent dataset known as the VPC, which is suitable for visual place recognition in home environments. In another work [32] a free online data source was introduced, which provides a large and growing amount of human-labeled visual data, much of which contains indoor scenes suitable for place memorization and object recognition. Moreover, in [3] a publicly available dataset was offered [120], captured by a mobile robot in an university environment. It has been acquired by means of an RGB-D sensor mounted on the mobile robot under multiple illumination conditions. Thus, it is suitable for the evaluation of SLAM and place recognition algorithms. Pronobis *et al.* [121] introduced two different datasets, namely the INDECS and the IDOL. INDECS comprises images of the environment from a fixed set of points using a standard camera mounted on a tripod. On the other hand, IDOL consists of image sequences recorded using two mobile robot platforms equipped with perspective cameras and thus is well suited for experimentation and validation of semantic maps. A more completed dataset suitable for the evaluation of robot localization and place recognition algorithms is the COLD one [86]. This particular collection contains three separated parts acquired at three different laboratories, located in three distinct European cities. The camera settings used in all three parts were identical, while both perspective and omnidirectional cameras, mounted together on a portable socket were utilized. The sequences

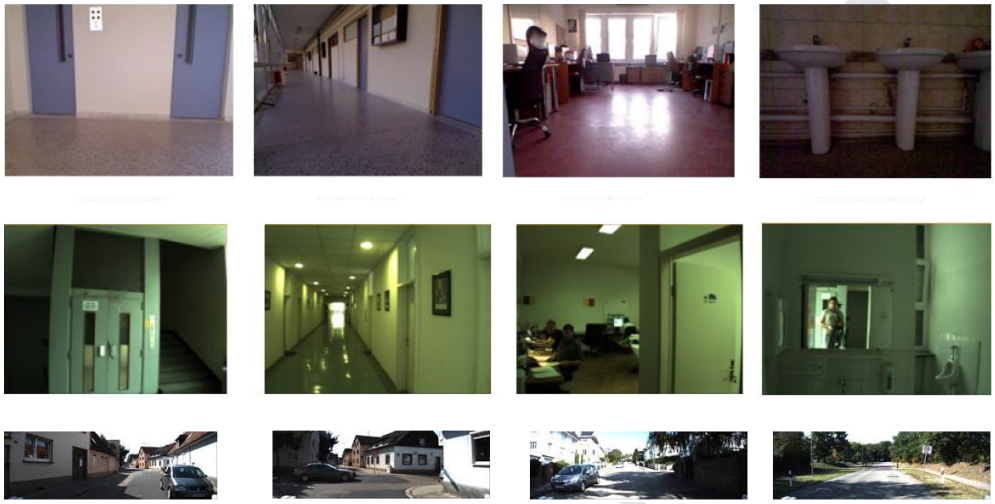


Figure 18: The first row illustrates image samples from the Cognitive Navigation dataset, in the second row instances from the COLD dataset are presented, while in the third row outdoor images from the KITTI dataset are appended.

were acquired under different weather and illumination conditions across a time span of two or three days. This specific dataset is ideal for the assessment of the robustness of visual place recognition algorithms with respect to dynamic, geographic and category changes. This dataset is also free for the community and can be found in [122]. Concerning the outdoors scenarios the authors in [123], [124] and [125] developed a large scale benchmark suite, named KITTI. It is suitable for outdoors semantic mapping evaluation, since it contains data from stereo, optical flow, visual odometry, 3D object detection and 3D tracking. In a smaller scale Sengupta *et al.* [69] offered to the community a hand labeled dataset of 14.8km containing 13 classes of human oriented concepts, suitable for outdoors semantic mapping. An minimum exemplar set of images from the existing datasets is presented in Fig. 18. Last, a plethora of data for the evaluation of the semantic mapping methods can be found in RobotVision@ImageCLEF [126], [127]. This is a great challenge that primary addresses the problem of multimodal place classification. The goal of the challenge is to classify rooms on the basis of image sequences captured by cameras mounted on a mobile robot.

6. Open Issues and Questions

About ten years ago, in a survey on socially interactive robots [128] it was deduced that it is of great moment for the robots to be compatible with the humans requirements, to match the application demands, to be understandable and to allow the interactional comfort the human expects. Although the introduction of semantic mapping in mobile robots have narrowed the distance in all these aspects since then, there is still a long way to go, as well as other open issues and questions in need for a reply, including the following:

What are the minimal criteria for a map to be semantic? According to the study of the existing techniques conducted earlier in this paper, there are several key attributes that a map needs to retain in order to be characterized as a semantic one. Apart from the existence of a metric map, either 2D or 3D, a semantic map should also employ signification cognizing capacities [3]. The abstraction of the metric map in terms of a topological map is primary because it organizes the explored environment as a graph. The vertices of the graph may retain both geometrical and semasiological characteristics, such as labels of the perceived environment, and the edges reveal the physical connection among the detected places [2].

Further to these minimal criteria, some pluses might arise by utilizing a place, a shape or an object recognition algorithm, to assist the robot to apprehend its surroundings in a human compatible manner, providing that these are appropriately integrated within the map . Moreover, the embodiment of the temporal connection can also augment the capacities of the agents to correctly distinguish the explored environment into places [78]. Additionally, the exploitation of multiple modalities and the human assertion may lead to the construction of semantic maps robust enough for the next generation of robots to cope with in any environment [79].

How do we evaluate semantic mapping methods?
Given the fact that a semantic map is a complex system, the existence of a sole evaluation method for the performance of this system is ambiguous. However, each subordinate module could be separately evaluated. For example the accuracy of the respective metric mapping, where available, can be evaluated by having the groundtruth -measured with differential GPS- to compare the trajectory of the visual odometry component similar to the work in [129], when it comes to outdoors scenarios. For indoors scenarios the direct comparison of the produced metric map with the CAD models of the buildings could provide an adequate quantification of the performance.

The place and object recognition algorithm could be evaluated on labeled datasets (see Sec. 5), the performance of which can be assessed by classic precision and recall methods or by ROC curves. However, the evaluation of such methods is also depended on the utilized method for classification. Another common technique to examine the performance of the recognition practices is by means of the confusion matrices which typically exhibit the averaged performance of the examined method with respect to the standard deviation.

What differentiates semantic mapping methods from place recognition ones?

It is important to mention that a semantic mapping system may annex a place recognition one. However, there are place recognition methods that are focused only on recalling places from a queue of randomly placed images. This is much different from the semantic mapping where the various inferences about the visited places are made subject to spatial constraints. For example we may consider the work described in [8] where the authors addressed solely the place recognition problem while, in the work presented in [48] place recognition capabilities integrated to a mobile robot to identify places and recognize semantic categories in an indoor environment. Moreover, in semantic mapping the robot's motion is taken into account to assist the place recognition module. The system should be able to accurately recognize the learned place instances as it moves from one location to another. The latter requires the utilization of robust machine learning techniques competent to deal with any dynamic change of the explored environments. However, a semantic mapping system should hold place categorization capabilities as well, i.e. it should be proficient in categorization and not only in recognition of different places. To this end, the robot should be capable of classifying and producing labels for places about which no prior knowledge is available [3]. Ergo, the system should be able to generalize the knowledge gained by exploring a specific place, so as to infer about the semantic content of any other similar place.

What are the potential areas of application of semantic maps?

It has been apparent throughout this survey that semantic maps were primarily invented to be utilized by domestic robots, such as robot companions or locomotion assistance robots, as a means to facilitate HRI [4]. Yet, the rapid evolvement of this domain and the several implementation solutions presented so far, revealed a number of other application areas, as follows. In the factories of the future the laborers will work in close cooperation with

robots and, therefore, a meaningful representation of the workshop should be available for this collaboration to be efficient. The same principle applies to future storehouses, where *automated guided vehicles* (AGV) need to be equipped with semantic mapping capabilities. When the appropriate sensors are used to signify a map (e.g. olfactory, chemical, CO_2, etc), the result can be suitable for search and rescue operations, in which first responders will be in close cooperation with robots. In outdoors scenarios a semantic representation of the environment is essential in the forthcoming *driverless cars*, to make the passengers' experience smoother (e.g. to pass an order to the car, such as *"at the steak house turn left"*). Agriculture robotics is another application domain, in which the farmer of the future will not need to be a programmer in order to communicate to his/her robot, so as to ask from it for instance to *"pass through the orchard and dig the vineyard"*. Ergo, the previously mentioned examples make clear that any application domain requiring for human robot cooperation will benefit from the semantic mapping technology.

How semantic maps aid knowledge representation and vice versa?
Ontologies and other knowledge representation schemata can yield full description of the robot's surroundings, encoding and revealing attributes even when these are not perceivable [108], [115]. In this sense a semantic map can indeed get support by such a formalism, since the second allows first an information richer representation of the various instances within the map. On the other hand, semantic maps normally recognize objects, places and other classes and they put them into a spatially organized hierarchy. Therefore, an efficient semantic mapping can provide the means to keep the spatial attributes of the individuals up to date. This is of primary importance, since the *ground level objects* that are organized into a knowledge representation formalism are subject to the continuous changes of the environment, which both the robot and the human should be aware of.

What are the research challenges and the future trends?
A challenge for the upcoming research endeavors constitutes the semantic mapping of large scale outdoors scenarios in terms of neighborhoods recognition, or even cities and counties. However, such an attempt demands the existence of a superabundance of data to train all the models. One solution to this problem would be to take advantage of the fact that everybody is connected on the web and, based on cloud technology, to facilitate knowledge exchange between robots and humans located at different spots around the globe. This way artificial agents will be enabled to exchange visual or other

modality memories among themselves and with humans. For this to come into reality, semantic mapping is only a small piece of the puzzle; technologies including cloud robotics, social media, internet of things, social robotics, etc, need to be integrated into a unified mainstream technology.

7. Epilogue

From the plethora of the laborious research that has already been conducted towards this direction, it is revealed that the semantic mapping is an active and ceaselessly growing research area. Although it is a relatively new topic of interest, various aspects of it have emerged through the recent years. This survey intended to summarize and categorize the existing methods. Accordingly, the primary characteristics of the semantic mapping were outlined and clustered corresponding to the scalability of the utilized metric maps, the use of topological maps, the type of the perception and the temporal coherence. Moreover, other characteristics a semantic map might bear were also reviewed and discussed. In conclusion one may say that the majority of the existing methods aspire to solve the problem for indoors environments rather than outdoors. This typically happens due to the fact that the domestic environments are better organized and clearly determined by humans. An important aspect of the semantic mapping for indoors is the employment of place and object recognition, which still is prohibited for outdoors. Nevertheless, many research efforts have been reported for outdoors scenarios being concentrated on the scene analysis in terms of roads, pedestrians, signs, traversability, etc. Another significant attribute, occurring in many works, is the topological map, through which the explored environment is organized in an abstract manner, particularly in terms of graphs. A significant advantage of this strategy is that the computational burden is diminished, owing to the fact that the retrieval and recall of the visual memories is performed hierarchically and not exhaustively. The temporal coherence is also of great importance in semantic mapping, since it makes use of the robot's movement figuring out physically constrained transitions among the explored spots.

References

[1] G. Ryle, Abstractions, Dialogue 1 (01) (1962) 5–16.

[2] A. Pronobis, P. Jensfelt, Large-scale semantic mapping and reasoning with heterogeneous modalities, in: International Conference on Robotics and Automation, IEEE, 2012, pp. 3515–3522.

[3] I. Kostavelis, A. Gasteratos, Learning spatially semantic representations for cognitive robot navigation, Robotics and Autonomous Systems 61 (12) (2013) 1460–1475.

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

[5] Ó. M. Mozos, Semantic labeling of places with mobile robots, Vol. 61, Springer, 2010.

[6] A. Pronobis, Semantic mapping with mobile robots, Ph.D. thesis, KTH Royal Institute of Technology, Stockholm, Sweden (Jun. 2011).
URL `http://www.pronobis.pro/phd`

[7] M.-W. Inc., Merriam-Webster's collegiate dictionary, Merriam-Webster, 2004.

[8] A. Pronobis, B. Caputo, P. Jensfelt, H. I. Christensen, A discriminative approach to robust visual place recognition, in: International Conference on Intelligent Robots and Systems, IEEE, 2006, pp. 3829–3836.

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

[10] G. N. DeSouza, A. C. Kak, Vision for mobile robot navigation: A survey, Transactions on Pattern Analysis and Machine Intelligence, 24 (2) (2002) 237–267.

[11] B. Kuipers, Modeling spatial knowledge, Cognitive science 2 (2) (1978) 129–153.

[12] B. Kuipers, Y.-T. Byun, A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations, Robotics and autonomous systems 8 (1) (1991) 47–63.

[13] B. Kuipers, The spatial semantic hierarchy, Artificial Intelligence 119 (1) (2000) 191–233.

[14] S. Thrun, W. Burgard, D. Fox, et al., Probabilistic robotics, Vol. 1, MIT press Cambridge, 2005.

[15] D. Filliat, J.-A. Meyer, Map-based navigation in mobile robots: I. a review of localization strategies, Cognitive Systems Research 4 (4) (2003) 243–282.

[16] J.-A. Meyer, D. Filliat, Map-based navigation in mobile robots: II. a review of map-learning and path-planning strategies, Cognitive Systems Research 4 (4) (2003) 283–317.

[17] Y.-D. Jian, D. Balcan, I. Panageas, P. Tetali, F. Dellaert, Support-theoretic subgraph preconditioners for large-scale slam, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 9–16.

[18] G. Grisetti, C. Stachniss, W. Burgard, Improved techniques for grid mapping with rao-blackwellized particle filters, IEEE Transactions on Robotics 23 (1) (2007) 34–46.

[19] D. Hahnel, W. Burgard, D. Fox, S. Thrun, An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements, in: International Conference on Intelligent Robots and Systems, Vol. 1, IEEE, 2003, pp. 206–211.

[20] H. Durrant-Whyte, T. Bailey, Simultaneous localization and mapping: part i, Robotics & Automation Magazine, IEEE 13 (2) (2006) 99–110.

[21] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (slam): Part II, Robotics & Automation Magazine 13 (3) (2006) 108–117.

[22] S. Thrun, Learning metric-topological maps for indoor mobile robot navigation, Artificial Intelligence 99 (1) (1998) 21–71.

[23] A. Angeli, S. Doncieux, J.-A. Meyer, D. Filliat, Visual topological slam and global localization, in: International Conference on Robotics and Automation, IEEE, 2009, pp. 4300–4305.

[24] J.-L. Blanco, J.-A. Fernández-Madrigal, J. Gonzalez, A new approach for large-scale localization and mapping: Hybrid metric-topological slam, in: International Conference on Robotics and Automation, IEEE, 2007, pp. 2061–2067.

[25] S. Bazeille, D. Filliat, Incremental topo-metric slam using vision and robot odometry, in: International Conference on Robotics and Automation, IEEE, 2011, pp. 4067–4073.

[26] C. W. Nielsen, B. Ricks, M. A. Goodrich, D. Bruemmer, D. Few, M. Few, Snapshots for semantic maps, in: International Conference on Systems, Man and Cybernetics, Vol. 3, IEEE, 2004, pp. 2853–2858.

[27] I. Kostavelis, A. Gasteratos, E. Boukas, L. Nalpantidis, Learning the terrain and planning a collision-free trajectory for indoor post-disaster environments, in: International Symposium on Safety, Security, and Rescue Robotics, IEEE, 2012, pp. 1–6.

[28] I. Kostavelis, L. Nalpantidis, A. Gasteratos, Collision risk assessment for autonomous robots by offline traversability learning, Robotics and Autonomous Systems 60 (11) (2012) 1367–1376.

[29] R. B. Rusu, B. Gerkey, M. Beetz, Robots in the kitchen: Exploiting ubiquitous sensing and actuation, Robotics and Autonomous Systems 56 (10) (2008) 844–856.

[30] P. Viswanathan, T. Southey, J. J. Little, A. Mackworth, Automated place classification using object detection, in: Canadian Conference on Computer and Robot Vision, IEEE, 2010, pp. 324–330.

[31] P. Viswanathan, T. Southey, J. Little, A. Mackworth, Place classification using visual object categorization and global information, in: Canadian Conference on Computer and Robot Vision, IEEE, 2011, pp. 1–7.

[32] B. C. Russell, A. Torralba, K. P. Murphy, W. T. Freeman, Labelme: a database and web-based tool for image annotation, International journal of computer vision 77 (1-3) (2008) 157–173.

[33] A. Trevor, S. Gedikli, R. Rusu, H. Christensen, Efficient organized point cloud segmentation with connected components, in: International

Conference on Robotics and Automation, 3rd Workshop on Semantic Perception, Mapping, and Exploration, IEEE, 2013, p. In Press.

[34] A. Swadzba, S. Wachsmuth, A detailed analysis of a new 3d spatial feature vector for indoor scene classification, Robotics and Autonomous Systems 62 (5) (2014) 646–662.

[35] O. M. Mozos, H. Mizutani, R. Kurazume, T. Hasegawa, Categorization of indoor places using the kinect sensor, Sensors 12 (5) (2012) 6695–6711.

[36] P. Espinace, T. Kollar, N. Roy, A. Soto, Indoor scene recognition by a mobile robot through adaptive object detection, Robotics and Autonomous Systems 61 (9) (2013) 932–947.

[37] A. Nüchter, O. Wulf, K. Lingemann, J. Hertzberg, B. Wagner, H. Surmann, 3d mapping with semantic knowledge, in: RoboCup 2005: Robot Soccer World Cup IX, Springer, 2006, pp. 335–346.

[38] N. Blodow, L. C. Goron, Z.-C. Marton, D. Pangercic, T. Ruhr, M. Tenorth, M. Beetz, Autonomous semantic mapping for robots performing everyday manipulation tasks in kitchen environments, in: International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 4263–4270.

[39] R. B. Rusu, Z. C. Marton, N. Blodow, A. Holzbach, M. Beetz, Model-based and learned semantic object labeling in 3d point cloud maps of kitchen environments, in: International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 3601–3608.

[40] A. J. Trevor, J. G. Rogers, C. Nieto-Granda, H. I. Christensen, Tables, counters, and shelves: Semantic mapping of surfaces in 3d, in: International Conference on Robotics and Automation, 3rd Workshop on Semantic Perception, Mapping, and Exploration, Georgia Institute of Technology, 2010.

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

[42] A. J. Trevor, J. G. Rogers, C. Nieto-Granda, H. I. Christensen, Feature-based mapping with grounded landmark and place labels, in: Workshop on Grounding Human-Robot Dialog for Spatial Tasks, 2011.

[43] A. Cowley, C. J. Taylor, B. Southall, Rapid multi-robot exploration with topometric maps, in: International Conference on Robotics and Automation, IEEE, 2011, pp. 1044–1049.

[44] M. Gnther, T. Wiemann, S. Albrecht, J. Hertzberg, Building semantic object maps from sparse and noisy 3d data, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 2228–2233.

[45] J. Civera, D. Gálvez-López, L. Riazuelo, J. D. Tardós, J. Montiel, Towards semantic slam using a monocular camera, in: International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 1277–1284.

[46] O. Martinez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, W. Burgard, Supervised semantic labeling of places using information extracted from sensor data, Robotics and Autonomous Systems 55 (5) (2007) 391–402.

[47] M. Michael, N. Roy, S. Thrun, D. Haehnel, C. S. et al., Carmen robot navigation toolkit (2000).
URL `http://carmen.sourceforge.net/`

[48] A. Pronobis, O. M. Mozos, B. Caputo, P. Jensfelt, Multi-modal semantic place classification, The International Journal of Robotics Research 29 (2-3) (2010) 298–320.

[49] S. Ekvall, P. Jensfelt, D. Kragic, Integrating active mobile robot object recognition and slam in natural environments, in: International Conference on Intelligent Robots and Systems, IEEE, 2006, pp. 5792–5797.

[50] Z. Liu, G. von Wichert, Extracting semantic indoor maps from occupancy grids, Robotics and Autonomous Systems 62 (5) (2014) 663–674.

[51] A. K. Krishnan, K. M. Krishna, A visual exploration algorithm using semantic cues that constructs image based hybrid maps, in: International Conference on Intelligent Robots and Systems, IEEE, 2010, pp. 1316–1321.

[52] M. Luperto, A. Q. Li, F. Amigoni, A system for building semantic maps of indoor environments exploiting the concept of building typology, in: RoboCup 2013: Robot World Cup XVII, 2013, pp. 504–515.

[53] A. Pronobis, P. Jensfelt, Understanding the real world: Combining objects, appearance, geometry and topology for semantic mapping.

[54] S. Vasudevan, S. Gächter, V. Nguyen, R. Siegwart, Cognitive maps for mobile robotsan object based approach, Robotics and Autonomous Systems 55 (5) (2007) 359–371.

[55] C. Case, B. Suresh, A. Coates, A. Y. Ng, Autonomous sign reading for semantic mapping, in: International Conference on Robotics and Automation, IEEE, 2011, pp. 3297–3303.

[56] C. Nieto-Granda, J. G. Rogers, A. J. Trevor, H. I. Christensen, Semantic map partitioning in indoor environments using regional analysis, in: International Conference on Intelligent Robots and Systems, IEEE, 2010, pp. 1451–1456.

[57] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, Ros: an open-source robot operating system, in: International Conference on Robotics and Automation, Workshop on Open Source Software, Vol. 3, 2009.

[58] Y. Feng, J. Ren, J. Jiang, M. Halvey, J. M. Jose, Effective venue image retrieval using robust feature extraction and model constrained matching for mobile robot localization, Machine Vision and Applications 23 (5) (2012) 1011–1027.

[59] A. Ranganathan, J. Lim, Visual place categorization in maps, in: International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 3982–3989.

[60] Z. Liu, G. von Wichert, Applying rule-based context knowledge to build abstract semantic maps of indoor environments, in: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, November 3-7, 2013, IEEE, 2013, pp. 5141–5147.

[61] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.-A. Fernandez-Madrigal, J. González, Multi-hierarchical semantic maps for mobile

robotics, in: International Conference on Intelligent Robots and Systems, IEEE, 2005, pp. 2278–2283.

[62] A. Rituerto, A. Murillo, J. Guerrero, Semantic labeling for indoor topological mapping using a wearable catadioptric system, Robotics and Autonomous Systems 62 (5) (2014) 685–695.

[63] J. Fasola, M. J. Matarić, Using semantic fields to model dynamic spatial relations in a robot architecture for natural language instruction of service robots, in: International Conference on Intelligent Robots and Systems, 2013.

[64] M.-L. Wang, H.-Y. Lin, An extended-hct semantic description for visual place recognition, The International Journal of Robotics Research 30 (11) (2011) 1403–1420.

[65] J.-B. Bordes, P. Xu, F. Davoine, H. Zhao, T. Denœux, Information fusion and evidential grammars for object class segmentation, in: International Conference on Intelligent Robots and Systems, 5th Workshop on Planning, Perception and Navigation for Intelligent Vehicles, IEEE, 2013, p. In Press.

[66] S. Sengupta, E. Greveson, A. Shahrokni, P. H. S. Torr, Urban 3d semantic modelling using stereo vision, in: IEEE International Conference on Robotics and Automation, IEEE, 2013, pp. 580–585.

[67] R. Paul, R. Triebel, D. Rus, P. Newman, Semantic categorization of outdoor scenes with uncertainty estimates using multi-class gaussian process classification, in: International Conference on Intelligent Robots and Systems, IEEE, 2012, pp. 2404–2410.

[68] B. Steder, M. Ruhnke, S. Grzonka, W. Burgard, Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation, in: International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 1249–1255.

[69] S. Sengupta, P. Sturgess, P. H. Torr, et al., Automatic dense visual semantic mapping from street-level imagery, in: International Conference on Intelligent Robots and Systems, IEEE, 2012, pp. 857–862.

[70] G. Singh, J. Kosecka, Acquiring semantics induced topology in urban environments, in: International Conference on Robotics and Automation, IEEE, 2012, pp. 3509–3514.

[71] B. L. Saux, M. Sanfourche, Rapid semantic mapping: Learn environment classifiers on the fly, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 3725–3730.

[72] H. Katsura, J. Miura, M. Hild, Y. Shirai, A view-based outdoor navigation using object recognition robust to changes of weather and seasons, in: International Conference on Intelligent Robots and Systems, Vol. 3, IEEE, 2003, pp. 2974–2979.

[73] P. Viswanathan, D. Meger, T. Southey, J. J. Little, A. K. Mackworth, Automated spatial-semantic modeling with applications to place labeling and informed search, in: Canadian Conference on Computer and Robot Vision, IEEE, 2009, pp. 284–291.

[74] S. Vasudevan, R. Siegwart, Bayesian space conceptualization and place classification for semantic maps in mobile robotics, Robotics and Autonomous Systems 56 (6) (2008) 522–537.

[75] A. Pronobis, K. Sjöö, A. Aydemir, A. N. Bishop, P. Jensfelt, Representing spatial knowledge in mobile cognitive systems, in: 11th International Conference on Intelligent Autonomous Systems, 2010.

[76] A. Aydemir, M. Göbelbecker, A. Pronobis, K. Sjöö, P. Jensfelt, Plan-based object search and exploration using semantic spatial knowledge in the real world, in: European Conference on Mobile Robotics, 2011.

[77] D. W. Ko, C. Yi, I. H. Suh, Semantic mapping and navigation: A bayesian approach, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 2630–2636.

[78] O. M. Mozos, W. Burgard, Supervised learning of topological maps using semantic information extracted from range data, in: International Conference on Intelligent Robots and Systems,, IEEE, 2006, pp. 2772–2777.

[79] A. Pronobis, P. Jensfelt, Hierarchical multi-modal place categorization, in: 5th European Conference on Mobile Robots, 2011.

[80] I. Kostavelis, K. Charalampous, A. Gasteratos, Online spatiotemporal-coherent semantic maps for advanced robot navigation, in: International Conference on Intelligent Robots and Systems, 5th Workshop on Planning, Perception and Navigation for Intelligent Vehicles, IEEE, 2013.
URL http://ppniv13.irccyn.ec-nantes.fr/material/session3/paper8.pdf

[81] A. Ranganathan, Pliss: labeling places using online changepoint detection, Autonomous Robots 32 (4) (2012) 351–368.

[82] H. Guillaume, M. Dubois, F. Emmanuelle, P. Tarroux, et al., Temporal bag-of-words-a generative model for visual place recognition using temporal integration, in: VISAPP-International Conference on Computer Vision Theory and Applications, 2011.

[83] C. Cadena, D. Gálvez-López, J. D. Tardós, J. Neira, Robust place recognition with stereo sequences, IEEE Transactions on Robotics 28 (4) (2012) 871–885.

[84] I. Kostavelis, A. Amanatiadis, A. Gasteratos, How do you help a robot to find a place? a supervised learning paradigm to semantically infer about places, in: Hybrid Artificial Intelligent Systems, Springer, 2013, pp. 324–333.

[85] O. Linde, T. Lindeberg, Object recognition using composed receptive field histograms of higher dimensionality, in: 17th International Conference on Pattern Recognition, Vol. 2, IEEE, 2004, pp. 1–6.

[86] M. Ullah, A. Pronobis, B. Caputo, J. Luo, R. Jensfelt, H. Christensen, Towards robust place recognition for robot localization, in: International Conference on Robotics and Automation, IEEE, 2008, pp. 530–537.

[87] A. Pronobis, L. Jie, B. Caputo, The more you learn, the less you store: Memory-controlled incremental svm for visual place recognition, Image and Vision Computing 28 (7) (2010) 1080–1097.

[88] E. Fazl-Ersi, J. K. Tsotsos, Histogram of oriented uniform patterns for robust place recognition and categorization, The International Journal of Robotics Research 31 (4) (2012) 468–483.

[89] U. Weiss, P. Biber, Semantic place classification and mapping for autonomous agricultural robots, in: International Conference on Robotics and Automation, Workshop on Semantic Mapping and Autonomous Knowledge Acquisition, 2010.

[90] C. Stachniss, O. M. Mozos, W. Burgard, Speeding-up multi-robot exploration by considering semantic place information, in: International Conference on Robotics and Automation, IEEE, 2006, pp. 1692–1697.

[91] D. Meger, P.-E. Forssén, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, D. G. Lowe, Curious george: An attentive semantic robot, Robotics and Autonomous Systems 56 (6) (2008) 503–511.

[92] I. Jebari, S. Bazeille, E. Battesti, H. Tekaya, M. Klein, A. Tapus, D. Filliat, C. Meyer, R. Benosman, E. Cizeron, et al., Multi-sensor semantic mapping and exploration of indoor environments, in: Conference on Technologies for Practical Robot Applications, IEEE, 2011, pp. 151–156.

[93] C. Cadena, J. Košecka, Semantic parsing for priming object detection in rgb-d scenes, in: International Conference on Robotics and Automation, 3rd Workshop on Semantic Perception, Mapping, and Exploration, IEEE, 2013, p. In Press.

[94] D. F. Wolf, G. S. Sukhatme, Semantic mapping using mobile robots, Transactions on Robotics 24 (2) (2008) 245–258.

[95] H. He, B. Upcroft, Nonparametric semantic segmentation for 3d street scenes, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 3697–3703.

[96] G. Costante, T. A. Ciarfuglia, P. Valigi, E. Ricci, A transfer learning approach for multi-cue semantic place recognition, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 2122–2129.

[97] A. Anand, H. S. Koppula, T. Joachims, A. Saxena, Contextually guided semantic labeling and search for three-dimensional point clouds, The International Journal of Robotics Research 32 (1) (2013) 19–34.

[98] E. A. Topp, H. I. Christensen, Detecting region transitions for human-augmented mapping, Transactions on Robotics 26 (4) (2010) 715–720.

[99] G.-J. M. Kruijff, H. Zender, P. Jensfelt, H. I. Christensen, Situated dialogue and spatial organization: What, where... and why, International Journal of Advanced Robotic Systems 4 (2) (2007) 125–138.

[100] T. M. Bonanni, A. Pennisi, D. Bloisi, L. Iocchi, D. Nardi, Human-robot collaboration for semantic labeling of the environment, in: International Conference on Robotics and Automation, 3rd Workshop on Semantic Perception, Mapping, and Exploration, IEEE, 2013, p. In Press.

[101] A. J. B. Trevor, A. Cosgun, J. Kumar, H. I. Christensen, Interactive Map Labeling for Service Robots, in: Workshop on Active Semantic Perception in IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2012.

[102] F. Dayoub, T. Duckett, G. e. a. Cielniak, Towards an object-based semantic memory for long-term operation of mobile service robots, in: IEEE International Conference on Robotics and Automation, 3rd Workshop on Semantic Perception, Mapping, and Exploration, Georgia Institute of Technology, 2010.

[103] K. Klasing, G. Lidoris, A. Bauer, F. Rohrmüller, D. Wollherr, M. Buss, The autonomous city explorer: Towards semantic navigation in urban environments, in: 1st International Workshop on Cognition for Technical Systems (CoTeSys), 2008.

[104] J. Penders, Viewfinder:final activity report.

[105] Y. Baudoin, D. Doroftei, G. De Cubber, S. Berrabah, C. Pinzon, F. Warlet, J. Gancet, E. Motard, M. Ilzkovitz, L. Nalpantidis, et al., View-finder: robotics assistance to fire-fighting services and crisis management, in: International Workshop on Safety, Security & Rescue Robotics, IEEE, 2009, pp. 1–6.

[106] E. Prestes, J. L. Carbonera, S. R. Fiorini, V. A. Jorge, M. Abel, R. Madhavan, A. Locoro, P. Goncalves, M. E. Barreto, M. Habib, et al., Towards a core ontology for robotics and automation, Robotics and Autonomous Systems 61 (11) (2013) 1193 – 1204.

[107] J. L. Wyatt, A. Aydemir, M. Brenner, M. Hanheide, N. Hawes, P. Jensfelt, M. Kristan, G. Kruijff, P. Lison, A. Pronobis, et al., Self-understanding and self-extension: A systems and representational approach, Transactions on Autonomous Mental Development 2 (4) (2010) 282–303.

[108] G. H. Lim, I. H. Suh, H. Suh, Ontology-based unified robot knowledge for service robots in indoor environments, Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 41 (3) (2011) 492–509.

[109] A. Bouguerra, L. Karlsson, A. Saffiotti, Semantic knowledge-based execution monitoring for mobile robots, in: International Conference on Robotics and Automation, IEEE, 2007, pp. 3693–3698.

[110] M. Tenorth, L. Kunze, D. Jain, M. Beetz, Knowrob-map-knowledge-linked semantic object maps, in: International Conference on Humanoid Robots (Humanoids), IEEE, 2010, pp. 430–435.

[111] C. Finucane, G. Jing, H. Kress-Gazit, Ltlmop: Experimenting with language, temporal logic and robot control, in: International Conference on Intelligent Robots and Systems, IEEE, 2010, pp. 1988–1993.

[112] C. Matuszek, E. Herbst, L. Zettlemoyer, D. Fox, Learning to parse natural language commands to a robot control system, in: Proc. of the 13th Intl Symposium on Experimental Robotics (ISER), 2012.

[113] F. Heintz, Semantically grounded stream reasoning integrated with ros, in: International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 5935–5942.

[114] A. Borkowski, B. Siemiatkowska, J. Szklarski, Towards semantic navigation in mobile robotics, in: Graph transformations and model-driven engineering, Springer, 2010, pp. 719–748.

[115] C. Galindo, J.-A. Fernández-Madrigal, J. González, A. Saffiotti, Robot task planning using semantic maps, Robotics and Autonomous Systems 56 (11) (2008) 955–966.

[116] T. Kollar, V. Perera, D. Nardi, M. Veloso, Learning environmental knowledge from task-based human-robot dialog, in: International Conference on Robotics and Automation, IEEE, 2013, pp. 4304–4309.

[117] Y. Liu, J. Yang, Z. Wu, Ubiquitous and cooperative network robot system within a service framework, International Journal of Humanoid Robotics 8 (01) (2011) 147–167.

[118] C. Galindo, J. González, J.-A. Fernández-Madrigal, A. Saffiotti, Robots that change their world: Inferring goals from semantic knowledge, in: 5th European Conference on Mobile Robots, 2011, pp. 1–6.

[119] J. Wu, H. I. Christensen, J. M. Rehg, Visual place categorization: Problem, dataset, and algorithm, in: International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 4763–4770.

[120] I. Kostavelis, A. Gasteratos, Cognitive Navigation Dataset, Group of Robotics and Cgnitive Systems, Available at `http://robotics.pme.duth.gr/kostavelis/Dataset.html` (2012).

[121] A. Pronobis, B. Caputo, P. Jensfelt, H. I. Christensen, A realistic benchmark for visual indoor place recognition, Robotics and autonomous systems 58 (1) (2010) 81–96.

[122] A. Pronobis, B. Caputo, COLD: COsy Localization Database, The International Journal of Robotics Research (IJRR) 28 (5) (2009) 588–594. `doi:10.1177/0278364909103912`.
URL `http://www.pronobis.pro/publications/pronobis2009ijrr`

[123] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: Inernational Conference on Computer Vision and Pattern Recognition (CVPR, IEEE, 2012, pp. 3354–3361.

[124] J. Fritsch, T. Kuehnl, A. Geiger, A new performance measure and evaluation benchmark for road detection algorithms, in: International Conference on Intelligent Transportation Systems, 2013.

[125] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The kitti dataset, International Journal of Robotics Research, 2013.

[126] H. Mèuller, P. Clough, T. Deselaers, B. Caputo, ImageCLEF: Experimental Evaluation in Visual Information Retrieval, Vol. 32, Springer, 2010.

[127] J. Martínez-Gómez, I. García-Varea, B. Caputo, Overview of the imageclef 2012 robot vision task., in: CLEF (Online Working Notes/Labs/Workshop), 2012.

[128] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and autonomous systems 42 (3) (2003) 143–166.

[129] I. Kostavelis, E. Boukas, L. Nalpantidis, A. Gasteratos, Visual odometry for autonomous robot navigation through efficient outlier rejection, in: IEEE International Conference on Imaging Systems and Techniques, Proceedings, IEEE, 2013, pp. 45–50.

Antonios Gasteratos is an Associate Professor at the Department of Production and Management Engineering, Democritus University of Thrace (DUTH), Greece. He teaches the courses of Robotics, Automatic Control Systems, Measurements Technology and Electronics. He holds a B.Eng. and a Ph.D. from the Department of Electrical and Computer Engineering, DUTH, Greece. During 1999 - 2000 he was a visiting researcher at the Laboratory of Integrated Advanced Robotics (LIRA-Lab), DIST, University of 34 Genoa, Italy. He has served as a reviewer to numerous Scientific Journals and International Conferences. His research interests are mainly in mechatronics and in robot vision. He has published more than 160 papers in books, journals and conferences. He is a senior member of the IEEE. More details about him are available at http://robotics.pme.duth.gr/

Dr. Ioannis Kostavelis is a Postdoctoral Research Associate at the Democritus University of Thrace, Dept of Production and Management Engineer in the Robotics and Automation Laboratory. He holds a diploma in Production and Management Engineering from the Democritus University of Thrace an M.Sc. in Informatics from the Aristotle University of Thessaloniki. He fulfilled his PhD studies under the hood of the Laboratory of Robotics and Automation, at Department of Production and Management Engineering, Democritus University of Thrace, having as a supervisor Prof. Antonios Gasteratos. His research has been supported from several research projects funded by the European Space Agency, the European Commission and the Greek government. His current research interests include machine vision systems for robotic applications augmented with machine learning strategies, targeting on the construction of semantic maps suitable for high level robot navigation. More details about him are available at http://robotics.pme.duth.gr/kostavelis/

Cover Letter

- HIGHLIGHTS
- Two level navigation
- cognitive navigation
- Spatial semantics