Robotics and Autonomous Systems 93 (2017) 85–104

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Review

Recent trends in social aware robot navigation: A survey

Konstantinos Charalampous *, Ioannis Kostavelis, Antonios Gasteratos

Laboratory of Robotics and Automation, Production and Management Engineering Department, Democritus University of Thrace, Vas. Sophias 12, GR-671 00 Xanthi, Greece[1]

ARTICLE INFO

Article history:
Received 7 May 2016
Accepted 14 February 2017
Available online 7 April 2017

Keywords:
Metric mapping
Semantic mapping
Proxemics
Social mapping

ABSTRACT

With the robots tending to accumulate more and more capabilities beyond the level of acting in a deterministic fashion, the idea of introducing them into our every day lives seems to be closer now. Robotics systems and techniques appeared during the recent years have achieved astonishing potential to perceive and interpret their surrounding not only as low level features but also close to human understandable concepts. Such advances, in conjunction with the aspiration to incorporate robots into domestic or public places, led to the flourishing of fields dealing with their response in human presence. Following this notion, the field of social mapping was recently introduced in order to manage the shared space among robots and individuals in an ordinary fashion. This manuscript aims to systemize the recent literature by describing the required levels of robot perception, focusing on methods related to robot's social awareness, the availability of datasets these methods can be compared with, as well as issues that remain open and need to be confronted when robots operate in close proximity with humans.

Contents

1. Introduction 85
2. Historical overview 87
3. Trends in human–robot symbiosis 87
   3.1. Metric mapping 87
      3.1.1. Optimization criteria 87
      3.1.2. Feature description 88
      3.1.3. Indoor–outdoor environment 89
   3.2. Semantic mapping 89
      3.2.1. Scale based categorization 89
      3.2.2. Topological maps 91
      3.2.3. Temporal coherence 92
   3.3. Social mapping 92
      3.3.1. Indoor–outdoor interpretation 92
      3.3.2. Predictive mechanism 95
      3.3.3. Human body information 96
      3.3.4. Context awareness 97
4. Benchmarking, datasets and norms 97
5. Open issues and questions 99
6. Motivation for future work 100
7. Coda 101
References 101

1. Introduction

During the last years much progress has been done in the field of personal and professional service robots aiming to build robot companions and servants that will operate in assisting living environments with humans. Robots were initially invented to serve as a

* Corresponding author.
E-mail addresses: kchara@pme.duth.gr (K. Charalampous), gkostave@pme.duth.gr (I. Kostavelis), agaster@pme.duth.gr (A. Gasteratos).

[1] URL: http://robotics.pme.duth.gr.

http://dx.doi.org/10.1016/j.robot.2017.03.002

tool for mostly repeatable in nature tasks, with the particular aim to transact them in a precise and rapid fashion. Yet, while robots were primarily inaugurated to fit into industrial tasks intended for humans, soon after the application range was expanded into a big variety of fields and autonomy became condition sine qua non for their further expansion. These led to the definition of several task-dependent robot categories, such as industrial, medical, military, space and household, all of which tended to gradually neglect the necessity of a human operator. Since autonomy is by definition related to the freedom of one's action and a common point in most robot-related tasks is the ability to freely move around, autonomous robotic mobility attained great scientific and technical attention, during the last years. In particular the first autonomous navigation behaviors were simple obstacle avoidance techniques and a random navigation [1–3]. Then the community focused on the formation of a metric map, within which the robot should have been able to get self-localized using the so-called SLAM[2] techniques [4–6]. These allowed the creation of a 3D map of an unknown area, by exploiting a variety of sensors [7] and successively update the robot's attitude and the respective map. Although there is a variety regarding SLAM methods [8,9] capable of deriving consistent maps as well as accurately computing the robot's egomotion, their representation is distant from the human natural perception. Technically speaking, metric maps enable autonomous cruising of the robot from one geometrical position to another, implying that it is essential for the starting point to be known with respect to the global frame of the map, prior to computing the goal position.

The next big step in mapping technology was the ability to produce maps directly understandable by humans, which would bridge the gap between the robot's geometrical interpretation and high level concepts, thus facilitating the integration of robots into human populated environments. This issue is confronted by semantic mapping [10] that administers an interface between humans and robots. Semantic mapping provides an embellish representation of the surrounding environment containing high level components. Such components are meaningful to humans and they are often accompanied by appropriate geometrical ones, which are essential for an accurate navigation. The extraction of such components is achieved by enhancing the robot's perception with respect to its surroundings, i.e. the identification of the place, the objects that exist in the respective place and their positioning. Moreover, it is imperative to distinct different places and organize each one with the respective characteristics that accompany them. The way such attributes are organized shapes the interface with the user as well as the robot's capabilities to navigate by receiving instructions close to human concepts. Navigation premises the connection of high level attributes with geometrical information and a metric map, in order for the robot to be able to interpret high level commands to concepts suitable for this task. Therefore the majority of semantic mapping methodologies append human understandable features on top of metric maps, enabling their localization based on a geometrical representation, while communicating using human understandable concepts. Semantic mapping constitutes an important step towards HRI[3] [11], along with communication interfaces like natural language processing [12,13] or sign gestures [14,15].

The pioneering work of C. L. Breazeal primarily introduced the notion of social robots and contributed to the foundation of HRI. The scope of Breazeal's work extends to several areas within those scientific fields, aiming to social interaction and learning. This research includes (but not limited to) the endowment of robots with the proficiency of social cognition, imitation and tutelage [16]. It is worth however noting that although the field of social robots has broad extends, this manuscript focuses on the respective mapping techniques. Consequently, it is only natural that the subsequent steps in HRI – after the establishment of semantic mapping – would involve the manner a robots navigates in a human inhabited environment, which introduced the field of social mapping [17–19] and their behavior towards individuals which led to robot social behavior [20]. Social mapping deals with the problem of human-aware robot navigation. While the scope of autonomous navigation is constrained to obstacle avoidance and reaching the goal, the social one aims to navigate by additionally considering factors such as human comfort, naturalness and sociability [19]. The term human comfort refers to navigation manner that gives an individual the feeling of safety [21]. Safety can be attained from autonomous navigation by simply avoiding humans, however the derived trajectory in this case is probably a "rough" one and does not attach to the user the feeling of safety which leads to comfort. The term naturalness concerns the derivation of paths that are similar to the ones produced by humans. Most methods attempting to extract such trajectories focus on obtaining smoothness between successive point of the path or adjusting the speed of the robot. Last, sociability refers to abstract decisions about robot's maneuvering tied to regional and ethical notions; such an example would be a robot choosing to pass by a group of people being in a conversation and not through them. As previously mentioned, socially acceptable navigation relies on several factors where the robot's distance from individuals is the most extensively studied. The latter may be due to the work of Hall [22] that studied the spatial relation which holds among people's interaction and the introduction of the term "proxemics". The latter presents a conceptual personal space that people obey upon their interactions. Moreover, Hall's work quantified for the first time those conventions and found to be dependent upon the type of connection between individuals – for example friends or strangers – as well as the objective of their interaction. Yet, this study aimed at interaction among humans, by changing the counterpart to a robot, social mapping needs to consider more aspects and differences that arise.

By endowing robots with high level capacities in terms of perception and navigation, the idea of a robot being present in a human environment in a daily basis becomes more tangible. Under this prism, several recent works attempt to attach to robots roles which are currently performed by humans. Such roles require social skills in order to apprehend the user and to respond accordingly as well. Contemporary works explore the application of robots as support tools for children with autism [23]. These studies aim at designing robots that can stimulate social responses to children and act as a third party in a social interaction between children and adults. Moreover, robots can also be used in therapy sessions in order to advance the interplay of children with other individuals [24]. Social aware robots in terms of behavior are also developed for assisting elderly people [25]. Such robots besides of performing tasks essential for elderly people they should also be in place of carrying out task promoting the psychological wellbeing of the elders. The majority of relative studies reveal the positive effects of such robotic systems in terms of several psychological parameters such as stress, mood, communication and so on [26].

The scope of this paper is to provide a systematic overview of the most recent works related to robot perception and planning for an unimpeded human–robot symbiosis. This work aims to cover all related aspects initiating from the construction of metric maps which are essential for autonomous navigation and expanding to concepts that regard the human factor, namely semantic and social mapping. For each of the previously mentioned topics a qualitative taxonomy is provided, according to the most significant attributes that characterizes each one. Concerning the metric mapping and more particularly SLAM techniques, there are several survey works

[2] SLAM stands for Simultaneous Localization and Mapping.

[3] HRI stands for Human–Robot Interaction.

that provide in depth description. Here, we attempt to report the most recent works which are not included in the previously published surveys. The examined traits that affect the formation of metric maps are also taken into consideration here as well as the method used for prediction, the sensors, the type of features and the environment (viz. indoors or outdoors). Regarding the semantic mapping, the paper follows the taxonomy introduced in [10], thus examining whether those maps rely on single or multiple cues, the temporal component and the existence of a topological map. With respect to the social mapping techniques, the derived taxonomy concerns modalities and methods assessed to infer about the map. According to the attributes taken into consideration, the resulting maps may be particularly different, e.g. the pose of an individual might limit the space available for the robot to traverse. Another factor which may be used in cope with pose is the person's velocity. Several works, exploit prediction models in order to attain long term social maps and, thus, be able to draw long range planning. Another attribute that may be taken into account is the consideration of a single individual or multiple ones, which is also related to whether the technique draws maps for indoor or outdoor environments. Additionally, a subset of methods are context aware, meaning that the derived maps are affected for example from the activity performed by the human. Besides the categorization of these works, the paper at hand also discusses about the availability of datasets as well as norms and benchmarking strategies. Additionally, we have included the motivation for future work suggesting the development of joined technologies, which will effectively bridge the gap of perception and social mapping. Last, the paper concludes posing currently unresolved issues, questions and summarizing the respective trends.

2. Historical overview

Metric map formation is tightly related to SLAM methods, especially when they are constructed in an incremental fashion [27]. Such maps retain all geometrical details and are suitable for providing low level commands to the robot. The pioneering works in SLAM by Newman [28] and Davison and Murray [29] have inspired other researchers to grip with this scientific area. Several review papers exist regarding these methodologies [30–33] each of which examine the subject under a different prism. This work cites the most recent advances due to the fact that metric mapping constitutes the basis for human–robot symbiosis and is the stepping stone for semantic and social mapping. The formation of detailed metric maps and subsequently accurate robot localization paved the way for recognizing autonomous robots as reliable and capable of undertaking critical tasks, such as in search and rescue missions [34,35]. Yet, due to limitations of metric maps to provide semasiological concepts of the places and in conjunction with the need for a more user friendly robots semantic mapping emerged as a need. A primary target was the development of robots capable of receiving high level ''GO-TO'' commands. This aim would constitute robots easy to be used in a variety of real life applications, for example search and rescue missions are simplified when the operator is in place to instruct in a high level, naming a place that the robot should head to, instead of providing coordinates in a metric map [36]. Moreover, semantic mapping enables the utilization of robotic systems in human environments, undertaking every day tasks from consumers that are not familiar with operating such a complicated system. Kuiper's original work [37,38] provided a structured manner for enhancing robot's perception, endowing it to apprehend and abstract geometrical information to higher level concepts. This field, while established over two decades ago, flourished after the development of robust SLAM methods and real-time vision based techniques (see e.g. the early work in [39,40]).

Nevertheless, the acceptance of robots with semantic mapping capabilities on a daily basis requires a sense of familiarity. Robots could not fall under the category of domestic appliances; trades including their appearance and their ability to act and communicate at some level with people strengthen this opinion. Thus, in order to be admitted in a human environment they need to behave accordingly. This is not only due to aesthetic reasons but primarily due to uneasiness or distrust that could be risen by individuals. Robotics research addressed to social studies with the aim to decode and embed them in the respective resulting systems. The studies of Hall [22] set the groundwork by introducing the proxemics. However, those studies examined solely the social distances among humans. Subsequent ones, reveal that there exist more attributes that provide in robot socially agreeable cruising, such as the speed [41] and the relative angle of approach [42]. Early works that took under consideration such social constrains have been presented in [43,44], dealing with particular cases like standing in a queue [43] or allowing natural passage to people when meeting in hallways [44].

The establishment of methods for appropriate robot deployment in human residential, working or resting environments led to the emergence of the scientific area of robot social behavior. The latter deals with the development of robot systems that perform a certain task while presenting a socially acceptable behavior. The majority of such techniques need to be studied in a long-term fashion, to induct about their effectiveness [45]. Such preliminary studies [46,47] investigate the gradually alternation of how individuals respond towards robots when they are accepted as part of their every day routines. The inaugural work of Giusti et al. [48] was one of the first that examined the social part of assistance robotics. In that work, elderly patients with cognitive and behavioral disturbances participated in rehabilitation sessions where the robot was exploited as an interaction mediator and catalyst of social activity.

3. Trends in human–robot symbiosis

3.1. Metric mapping

Although the study of metric maps marginally falls into the scope of this paper as described in the previous sections, for the seek of completeness we will examine here the recent trends in this facet of perceiving the common environment of humans and robots. Moreover, the reader should constantly bear in mind that the borderline between a metric and a semantic or a social map are frequently not clearly distinguished and many semantic or social mapping techniques rely on a well estimated geometric map. This statement is further reinforced if we consider the recent work in social navigation [49]. Efficient robot navigation – in socially constrained environments – is tightly related to the quality of the metric mapping. Therefore, in Table 1 a summary of the taxonomy for the studied metric mapping techniques is exhibited.

3.1.1. Optimization criteria

In order for a robot to retain a socially acceptable ''behavior'', during its operation in human populated environments, and to apply safe manipulations, it should be endowed with accurate position estimations. However, since the robot navigation is applied on metric mapping, which in turn is constructed using noisy sensorial measurements, additional optimizations and filtering is required. Therefore, safe human aware navigation strongly depends on the post-processing of the modeled environment and the continuous update about the robot's poses. Additionally, the SLAM methodologies comprise the cornerstone of metric mapping strategies and the belief distribution of the robot's poses during its perambulation in the constructed map are typically filtered using Kalman Filter (KF) and Expectation Minimization (EM) [8,9].

Fig. 1. An example of illumination invariance application between two places at different hours. The images on the right are the illumination invariant transformation of the left images. It can be seen how the work in [56] reduces the effects produced by sunlight and shadows.

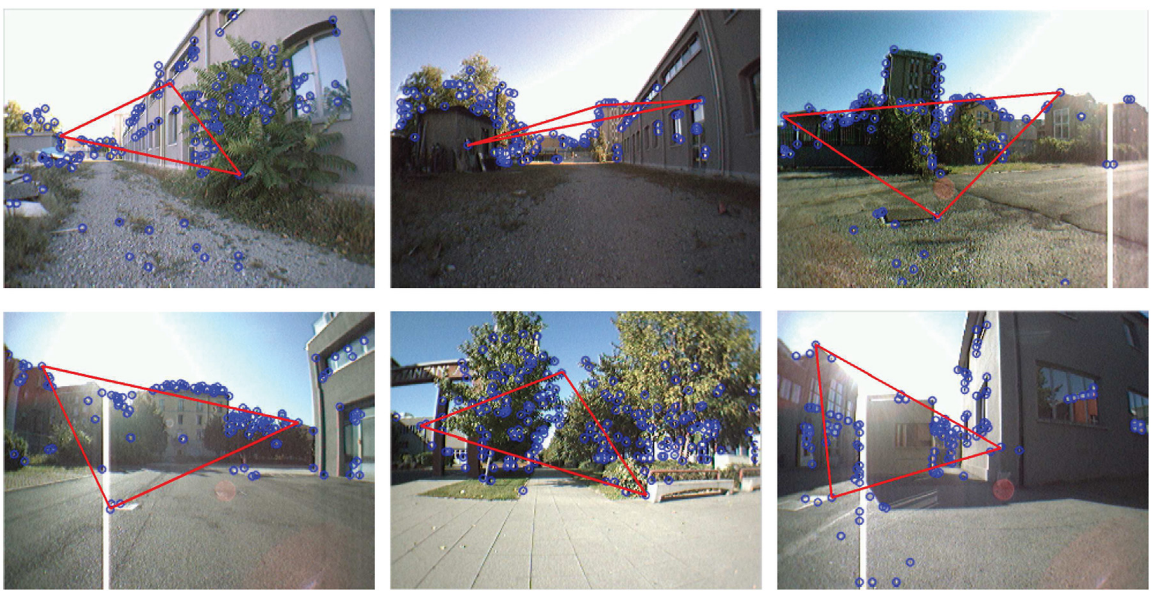


Fig. 2. Example frames from a sequence. The detected OOMC features are plotted in blue circles and the selected triplets are plotted in red triangles [57].

Pascoe et al. [50] utilized the Normalized Information Distance (NID) criterion to maximize the similarity between a virtual camera and the actual data w.r.t. the robot's pose. The maximization is achieved via the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, requiring the analytical expression of the respective Jacobian on every frame. Due to the computational burden, the authors proposed a paralleled-based solution. A graph-based formulation is given in [51] using the chordal distance as a metric. The authors minimized the angular error and expressed the problem as a quadratic one with quadratic equality constraints. The dual form of the problem enables a faster computation of the lower bound, while the methodology is in place to disambiguate global from local minima. In [52], a graph-based solution is also considered, where the graph is represented in a factorized form and is solved using least-squares optimization. The respective graph is a bipartite one representing measurements as well as poses and planes, where the two components are connected through edges. The least-squares formulation attains a linearized form w.r.t. the current estimation, thus, the solution may be computed either by Cholesky or QR factorization. Leung et al. [53] reformulated the SLAM problem by introducing Random Finite Sets (RFS) in contrast to the majority of works that use random vectors. In [53], the authors investigated the benefits of the proposed approach, proving that is a generalized schema of random vectors. Moreover, they uncover the conditions under which RFS-SLAM reduces to the random vector one when it is assumed ideal conditions, for example when the probability of detecting a landmark equals to one. The paper in [54] proposes an information-based approach for landmark detection and poses in order to perform SLAM. This method exploits a subset of possible landmarks while retaining the accuracy intact, thus, reducing the computational cost. Moreover, the authors administer an incremental version of this method, suitable for online applications.

3.1.2. Feature description

Feature description comprises an abstract representation of the perceived environment and is useful in reducing the great dimensionality of the sensorial data. On the one hand, the detection and description of prominent features can be utilized as landmarks for the construction of the metric map and the optimization of the robot's pose while, on the other hand, such characteristic representations can be utilized to code the human's presence within the environment, such as the Microsoft Kinect skeleton data. Therefore, this work would be incomplete unless study of the most recent feature description strategies employed in the social robot navigation is mentioned.

Towards this direction, and considering the abundance of the vision sensors, the data type each sensor produces needs to combine with the appropriate description. Such a description is essential to characterize the current input as well as the previous ones and using this common base, to compare them in order to infer about the robot's and human's attitude within the operational environment. The authors in [55] suggested a framework for long range SLAM that relies on CURB features. Those features are suitable for road extraction while the Kernel Fisher Discriminant Analysis is used as a final step to classify the CURB candidates. Arroyo et al. [56] proposed a description of RGB images as binary codes, thus reducing the computational cost as well as the memory requirements. The Local Difference Binary descriptor is used for this purpose, which are applied in a sequence of images. Then, the derived binary vectors are concatenated to form the overall descriptor. Their binary nature allows efficient matching utilizing the Hamming distance instead of the Euclidean one, an illustrative example of this system being given in Fig. 1. Another work that relies on RGB sensor is the one in [57] where the detected feature points are given as an input to the Optimal Observability and Minimal Cardinality (OOMC) method to infer about triplets of points that maximize the corresponding criterion, as depicted in Fig. 2. Afterwards, they are used as input to a KF. A different approach was followed in [58], inspired by the human attention model, where the points of interest were estimated through saliency maps.

3.1.3. Indoor–outdoor environment

The environment in which the robot operates has a particular impact on the robot setup as well as on the methodology followed. Factors such as illumination, structure of the surrounding area, dimensions of space and duration of the SLAM procedure force methodologies to adapt, thus, separating in indoors or outdoors.

Indoors Environment. Jalobeanu et al. [59] proposed a system relying on Kinect sensor to map indoor environments. This system is capable of combining a Truncated Signed Distance Function grid map with a particle filter in real-time, taking advantage of noisy depth readings and increases the FoV[4] by generating virtual views from local maps. The authors in [60] suggested a system for long term map formation by introducing the term summary map. In this map only the essential landmarks are kept, based on a scoring policy. This algorithm allows the improvement of the map as well as keeping it up-to-date when the robot revisits the same place. Kaess [52] proposed a SLAM method that incorporates information from the detected planes, such as stairs, ceiling and walls. By including planes into the problem's formulation the author derived a minimal representation that improves the convergence.

Outdoors Environment. The authors in [56] proposed a long-term localization system for outdoors scenarios invariant to factors such as illumination, weather or seasons. The latter is attained by transforming RGB images to an illuminance invariant color space and subsequently coding the derived information into binary representation. The work in [55] exploits CURB features to navigate in semi-structured environments where roads exist. Wolcott and Eustice [61] suggested a framework that also deals with weather conditions using a lidar scanner in conjunction with a branch-and-bound, multi-resolution schema. This framework exploits Gaussian mixture maps to represent the surrounding environment while operating in real-time. A demonstration of this work is illustrated in Fig. 3.

[4] FoV stands for Field of View.

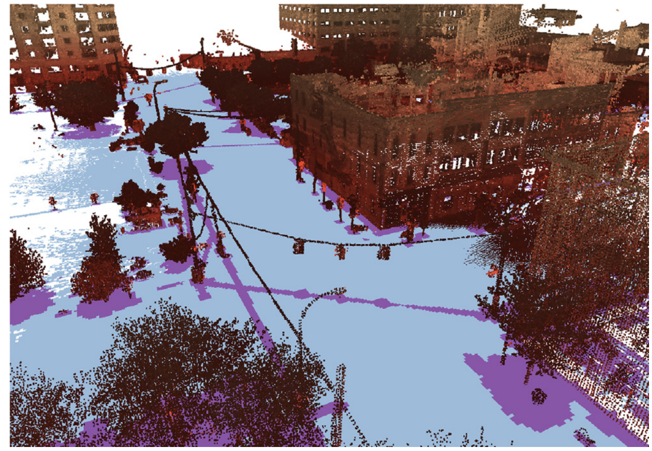


Fig. 3. Illustrated demonstration regarding the efficacy of the method in [61] w.r.t. the formation of the point-cloud.

3.2. Semantic mapping

The semantic mapping bridges the gap among robots and humans, concerning the apprehension of the space, by providing the means to represent the explored environment in a human compatible manner [62]. More specifically, it relates recognized objects and places with their spatial arrangement extracted from the environment's metric map. When the semantic map is augmented with human modeled actions it becomes the social aware mapping, where the robot can apprehend concepts such as "*humans typically walk in a corridor*". Therefore, since the semantic and social mapping act complementary [49], a brief discussion of the contemporary semantic mapping methodologies herein would better highlight the objective of this work.

3.2.1. Scale based categorization

As previously mentioned, metric maps comprise the basis for the majority of semantic or social mapping schemes. As a result, a straightforward distinction would be between indoor and outdoor systems. The latter is justified due to the fact that most of the indoor methods operate in a local fashion, ascribing high level characteristics to the robot's current view, while the outdoors solutions tend to incrementally augment the metric map with semantic attributes, thus appointing geometrical information to places or object under a common coordinate system. Moreover, Table 2 summarizes the taxonomy of the examined semantic mapping techniques.

Indoors single scene interpretation. Muller and Behnke [63] proposed a framework for semantic annotation of RGB-D[5] data. More precisely, an efficient GPU[6] implementation of Random Forests to cluster the scene and the derived spatial relation is modeled through a superpixel-based module. Then, the pixels are transformed into La*b* color space creating a distribution which is the input to a Conditional Random Field (CRF) model, thus providing unary features. Last, pair-wise depth-based featured are formed by taking into consideration the vertical alignment, the depth difference and the orientation of the normals. Another approach also relying on CRF and Random Forest is given in [64], where each RGB-D scene gets over-segmented; for each segment a feature vector is created considering the compactness, the planarity, the angle with the ground plane and the color valued in the La*b*

[5] RGB-D stands for Red, Green, Blue and Depth.

[6] GPU stands for Graphics Processing Unit.

Table 1
Survey summary of metric mapping methodologies.

Attribute/method	Optimization criterion	Feature description	Indoors	Outdoors
[50]	BFGS	NID		✓
[51]	Quadratic	–	✓	✓
[52]	iSAM	Normals	✓	
[53]	RFS	–		✓
[54]	Non-linear LS	Information theoretic		✓
[55]	Extended KF	CURB		✓
[56]	–	ABLE-M		✓
[57]	Extended KF	OOMC	✓	✓
[58]	Extended KF	iNVT		✓
[59]	Particle filter	–	✓	
[60]	–	–	✓	
[61]	Extended KF	LIDAR measurements		✓

space. Then, each computed feature vector is classified according to a Random Forest schema while the CRF is used to improve the labeling procedure. Mason et al. [65] suggest a method for discovering and categorizing objects in RGB-D images in an unsupervised manner. The work in [66] proposed a method that considers a point-cloud as a graph which is segmented in two different ways, by considering the respective triangular mesh and the projected image upon the 3D cloud. Carrillo et al. [67] proposed a method for place categorization from a single frame utilizing l_1-norm minimization. In this work, a dictionary is constructed from previously seen places and, by following the Lasso formulation, the current frame is reconstructed. Last, by calculating the Hadamard product the method infers about the place.

Indoors large scale interpretation. Hermans et al. [68] suggested a method for semantic segmentation of 3D scenes into distinct places. Qualitative results from this work are illustrated in Fig. 4. First, a 3D reconstruction process occurs, appending the new scene to the previously acquired ones. Afterwards, a soft classification step labels each point and comprises the initialization step of a 3D refinement procedure. In order to achieve the globally optimal labeling, the neighbor points are affected according to the distance, the color and the orientation of the normals. Ranganathan [69] proposed the place labeling through image sequence segmentation. This is an online method that relies on change point detection in order to infer about the parameter in a Bayesian model. The model is capable of segmenting RGB streams and labeling them. Moreover, new places can be categorized through statistical testing hypothesis. The authors in [70] suggested a method for place recognition and loop closure that relies on binary feature description that outperforms in terms of computation cost algorithms employing feature description mechanisms. Lu et al. [71] introduced a framework comprising layered costmaps, each of which has a distinctive semantic meaning. Low level ones deal with static obstacles such as walls, while higher ones infer about objects and humans. The work in [72] combines a metric map along with the object space. The method considers several attributes regarding objects – such as label and pose–and attempts to accurately localize them by marginalizing the posterior distribution of objects w.r.t. the occupancy grid. Zhao and Chen [73] suggested a method that combines the labeling of objects using the RGB data with SLAM. The classification is achieved using a CRF model which considers geometrical as well as semantic information, where the labeled objects are fused with the derived 3D map, yielding results as depicted in Fig. 5. Likewise, the work in [74] proposes a system that detects and classifies object in 3D scenes, then the objects along with their geometrical attributes are used as landmarks in order to refine the localization output of the SLAM procedure. Last, detected object are stored in a database which can be utilized either for SLAM or service robot procedures. Oliveira et al. [75] presented a perceptual memory system, which deals with computational issues and addresses them using a lightweight database and a multiplexing scheme, thus enabling the parallel processing for object categorization and perception. SEMAP [76] is a framework developed under ROS[7] to efficiently store and manipulate objects in a semantic map. Cleveland et al. [77] proposed a method for online object detection and map segmentation suitable for robots that perform inventory tasks.

Outdoors interpretation. The work in [78] described a system for semantic segmentation in outdoor scenes using a stereo camera. The stereo pair of RGB images is used to compute the depth while the visual odometry module provides the respective transformation which is utilized to append the scene with the already acquired map. Also a feature description step occurs on the RGB pair which are used in conjunction with volumetric mean-field inference approach for object labeling. The authors in [79] introduced the concept of semantic octree using stereo image pairs. The pairs are used for depth estimation and computation of object cues, which comprise the input in the octree to provide occupancy hypothesis. Afterwards, for every voxel of the octree a CRF is defined and an energy minimization procedure infers about the object labels. An illustrative example from this work is depicted in Fig. 6. Zhang et al. [80] presented a multi-modal framework for scene annotation that performs segmentation and feature extraction in multiple scales. Then, a classification step occurs in each modal individually. The fusion of modalities and a refinement step takes place using a CRF. The authors in [81] exploited man-made structures to enhance feature-based place recognition. By discarding features related to buildings the method improves the accuracy of various classification strategies. The work in [82] is also a multi-modal approach that exploits the formulation of CRF to merge sensors with different FoVs. This problem is treated as a graph, where the 3D cloud and the image are nodes connected by edges that attain weights through pairwise feature functions. Stumm et al. [83] have also regarded the problem as a graph-based one, where matching procedures take place to compare locations. This method utilizes loose structural information, which is sufficient to handle the problem of perceptual aliasing that most of the appearance-based methods suffer. Paul et al. [84] used Gaussian Processes (GP) to perform unsupervised semantic classification of 3D scenes, where an example is depicted in Fig. 7. The point-cloud is used to compute a triangular mesh in order to compute normal vectors and apply a segmentation module. Then, for each mesh a feature vector is formed by considering shape factors and relations among normal vectors. Last, those vectors are used as input to the GP classifier. Steder et al. [85] proposed an algorithm for place recognition and simultaneous pose refinement using 3D range data. The current input is used as a query in order for the method to return potential matches along with a similarity score. The score values are attained following the BoW[8] notion while for each candidate the respective pose transformation is computed. The pose being

[7] ROS stands for Robot Operating System.

[8] BoW stands for Bag of Words.

Fig. 4. Qualitative results of 3D reconstructions from method in [68]. (top row) Semantic reconstructions (bottom row) Corresponding RGB reconstructions.

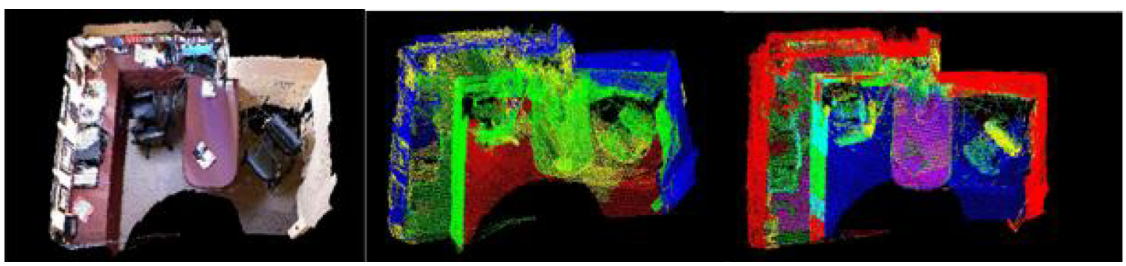


Fig. 5. An example of the semantic map generated from the method in [73]. The left is the global map generated from SLAM. The middle is a semantic map which the structural class labels are drawn. The right is a semantic map which object category labels are drawn.

closer to the current one is considered to be the winner. The BoW approach is also followed in [86] where the authors use local descriptors in order to generate 3D global ones. The BoW rationale permits the formulation of descriptors with fixed-dimensionality regardless of the keypoint detector or the combination of local descriptors. Another method that uses image query to solve the place recognition problem is proposed in [87]. This methodology presents a graph kernel approach in order to compute similarity between co-occurrent images. The resulted similarity criterion is used to form a hierarchical place recognition module that further enhances the methods discrimination capabilities. Semantic mapping methods have also been developed for navigation. Drouilly et al. [88] provide an object-based representation, where the objects form a semantic graph and the weights are computed through a similarity function. The derived graphs are exploited for high level navigation according to those objects. The authors in [89] merge metric and semantic information to deploy an automated parking system. The metric one is responsible for providing the geometry of the surrounding and accurate localization while the semantic map provides information about the road network and dynamic changes.

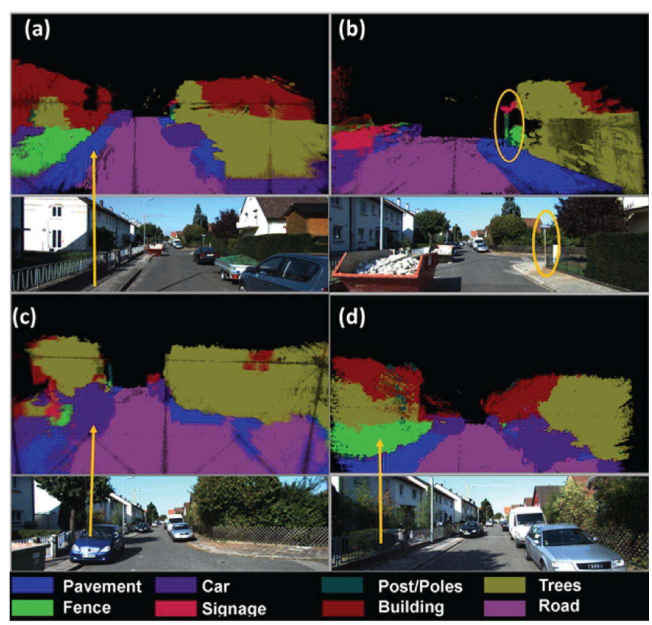


Fig. 6. Semantic octree qualitative results along with class labels from the method in [79]. The leaf level occupied voxels are shown while the arrows relate the position in the model and the image of: (a) the pavement and the road; (b) the post and signage; (c) the car and (d) the fence.

3.2.2. Topological maps

In the most of semantic mapping methodologies space is organized in an abstract manner, often referred as topological map [10]. The intuition for such an arrangement of space relies on the assumption that the places typically are characterized by humans using labels. These annotated representations are also widely used in the social mapping methodologies [49]. More precisely, the topological maps are graph representations of the explored environment. Each of its nodes corresponds to a different place and bears related characteristics, such as geometrical. The edges connecting different nodes attain weights according to the respective method, therefore, it may represent a short of metric distance or the probability of visiting a certain place while being in the current one. The authors in [90] proposed a supervised method for creating topological maps. The method uses range data to compute geometrical invariant features about the robot's pose. Then, they

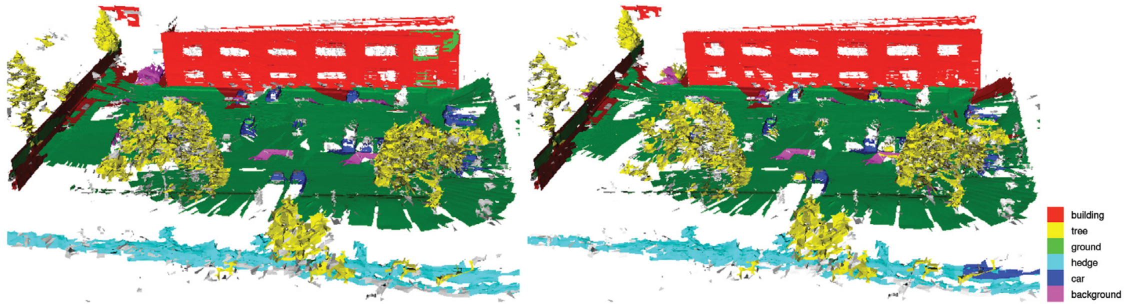


Fig. 7. Classification result of 9 point clouds from the work in [84]. Left: Ground truth. Right: Classification result. Only minor errors are visible.

become the input to an AdaBoost classifier to infer about each free cell of the occupancy grid. A probabilistic relaxation step occurs to smooth the classifier's decisions, which relies on labels from neighbor cells. Last, a topological graph is extracted according to heuristic rules about the area of a room and the doorways. Akdeniz and Bozma [91] introduced an online topological map methodology that operates on a local and a global level. Considering the local one, the robot decides the direction to move and appends an edge from the current node to a new defined one. The route between the two nodes is evaluated w.r.t. visibility and the detected obstacles according to a utility function and labels the respective edge. Regarding the global one, the algorithm infers about the long-term goal which is to maximize the coverage of the unexplored area. Long-term topological learning is also the goal in [92], where a place-memory system is introduced, named Bubble Descriptor Semantic Tree. The latter preserves information in a structured fashion from the learned places. Moreover, the structure is capable of updating itself in an unsupervised fashion using a hierarchical clustering method. Particular exploitation of graphs occurs in [83] being used not only to define new places as nodes but also to retrieve the current location as query to the already existing ones. Additionally, to enhance the decision about a place, a graph matching procedure takes place providing the likelihood for the candidates. The authors in [93] introduce a method that forms topological maps based on semantic image information for outdoor purposes.

Several works deal with hybrid solutions by combining metric and topological maps. Drouilly et al. [88] proposed a topological map where each node comprises a local metric sub-map and the edge weights are computed based on pairwise pose estimations. In [94] a three-level architecture is presented, where the lowest one consists of RGB spherical data, connect according to visual odometry information. The second level retains labeled images w.r.t. the objects and is directly related to the previously mentioned level. The last level is an abstract one where a graph structure connects different regions through nodes and edges are weighted according the number of common objects.

3.2.3. Temporal coherence

Semantic maps can be augmented by exploiting the inherent temporal proximity of successive video frames. This in turn enables the description of the physical constraints within the explored environment. More precisely, the temporal component attempts to capture various relations among different places, such as the transitional probability between two places. This dimension is common in the social mapping methodologies where, the robots typically monitor the humans actions for a time window where transitions among one place to another are observed, regulating thus their navigational decisions respectively. The authors in [95] perform semantic robot localization via spatio-temporal classification. First, the spatial classification occurs, where the input is divided into a grid and each cell grid attains a label, namely asphalt, cobblestones, grass, or gravel. Then, regarding the temporal one, the method uses visual odometry to merge the derived maps. The labeled maps are projected on the grid and a probabilistic criterion refines the grid labels by considering the neighbor cells. Hemachandra et al. [96] proposed a semantic framework that models the environment from natural language descriptions and scene classifications. This approach introduces the semantic graphs, which comprises metric, topological, and semantic maps. The topology of the latter graph has nodes declaring the robots trajectory and edges indicating the connectivity among the nodes. The temporal component participates on the updating procedure of the graph topology by considering previous metric exteroceptive sensor data, scene appearance observations and natural language descriptions. The authors in [97] exploit the temporal component, introducing a time-evolving navigation graph that provides a semantic topology of the explored area and the connectivity among the detected places in terms of the inter-place transition probability.

3.3. Social mapping

3.3.1. Indoor–outdoor interpretation

As previously mentioned, an apparent aggregation of mapping methodologies is the one regarding the type of the environment, i.e. indoors or outdoors. The reason for such a separation lies on the fact that most of these works ground their functionality on a metric map. The majority of indoor approaches deal with less number of individuals when comparing to the outdoor ones, yet due to confined space they need to provide sophisticated paths.

Indoor Interpretation. Sisbot et al. [98] introduced a human aware motion planner where the grid cell upon which the robot operates consist of the metric map as well as the social one. There are two criteria under consideration in order for the social map to be formed, namely safety and visibility. The first one regulates the distance between the robot and the individual. The method uses a Gaussian distribution to emulate the cost of approach the current position of a person, decaying when the robot is getting distanced from him/her. Additionally, the shape of the cost function is also dependent to the user's state, i.e. sitting, standing etc. The visibility criterion shapes the social map according to the effort needed in order for the user to see the robot. Ideally, the robot should be always within the FoV but not in the center. The latter criterion insures that the robot is always in the user's line of sight, thus eliminating the element of surprise. The work in [99] regards more criteria such as environment structure, unknown objects, social conventions, proximity constraints etc., to derive the respective social map. The method utilizes Voronoi diagrams to compute the map and starts by imprinting the environment structure on the Voronoi cells as obstacles. Then, the cost of the remaining free cells is examined by considering three set of rules. The first one refers to general social practices like traversing on the right half of a hallway or passing by an individual from his/her left. The second set applies

Table 2
Survey summary of semantic mapping methodologies.

Attribute/ method	Indoors single scene	Indoors large scale	Outdoors	Topological	Temporal coherence
[63]	✓Random forests partitioning and super-pixel creation in La*b* space to be used in CRF				
[64]	✓Feature vectors of RGB-D segmentation used Random forest classification and CRF label refining				
[65]	✓Unsupervised discover and categorization of objects in RGB-D images				
[66]	✓Graph segmentation of point-clouds				
[67]	✓l_1-norm minimization for place recognition based on Lasso simulation				
[68]		✓Point-wise soft classification of each consecutive scene			
[69]		✓Detection of viewpoint changes and Bayesian inference of parameters for place labeling			
[70]		✓Binary feature description for place recognition	✓Interpretation of outdoor scenes		✓Loop closure detection
[71]		✓Layers of costmaps with distinct semantic meaning			
[72]		✓Posterior distribution in object space combined with metric data			
[73]		✓CRF model based on object geometrical and semantic information fused with metric maps			
[74]		✓Object based landmarks for localization refinement			
[75]		✓Parallel processing for detecting and categorizing objects based on custom lightweight database			
[76]		✓ROS based storing and manipulation of objects in semantic maps			
[77]		✓Online object detection and metric segmentation			
[78]			✓Stereo based visual odometry and volumetric mean-field inference for object labeling		
[79]			✓Stereo based octree occupancy hypothesis and CRF object labeling		
[80]			✓Multi-scale feature extraction in multi-modal scene segmentation and CRF refining		

(continued on next page)

Table 2 (*continued*)

Attribute/ method	Indoors single scene	Indoors large scale	Outdoors	Topological	Temporal coherence
[81]			✓Outdoor classification based on man-made structures		
[82]			✓Graph based representation and CRF classification using various sensors		
[83]			✓Local structure information handle the problems of perceptual aliasing	✓Graph based location representation	
[84]			✓Unsupervised classification of outdoor scenes using Gaussian Processes		
[85]			✓Combined place recognition with pose refinement using 3D range data using BoW		
[86]		✓BoW for to generate 3D global descriptors from local ones			
[87]			✓Hierarchical place recognition module based on graph-kernel similarity between images		
[88]			✓Semantic map with object based representation	✓Sub-map construction based on topological map	
[89]			✓Combined metric and semantic information for road network modelling		
[90]	✓Adaboost classification of geometrical invariant features computed from lidar data			✓Rule based topological representation about the explored environment	
[91]	✓Utility function for labeling the transitioning among two places			✓Local and global topological representation maximizing the coverage area w.r.t navigation affordances	
[92]		✓Bubble Descriptor Semantic tree		✓Long term topological representation using hierarchical clustering	
[93]			✓Consecutive inference based on vision data	✓Large scale topological structures based for autonomous driving	
[94]		✓Multi-level structure; the first level comprises the RGB spherical the second level comprises the object labeling step		✓The third level is an abstract graph representation weighted w.r.t common labeled objects	
[95]			✓Semantic labeling of three class occupancy grid cells		✓Map partitioning and assembly based on coherent visual odometry data
[96]		✓Natural language description and scene classification			✓Time depended graph update considering previous exteroceptive sensor data
[97]		✓BoW spatial representation for place labeling		✓Abstract labeled graph of nodes for place partitioning	✓Time evolving navigation graph

the proxemic rules, i.e. the four social zones, which are expressed as elliptical regions around a person. The last set is an adaptive one that performs check to whether there is sufficient space for an individual to pass by considering static obstacles and human comfort. The authors in [100] introduced a perceptual model that considers the relative pose between an individual and the robot, the speech volume level and the manner the human performs a gesture. The relationship among those modes are modeled in a Bayesian network, thus inferring about the social map. The authors in [101] introduced a framework that detects a human in a scene and forms a social map by replacing his/hes bounding box according to a weight function. Rios-Martinez et al. [102] suggested a stochastic technique to compute a social map by means of stochastic and adaptive optimization. In particular, an objective function is proposed that incorporates a discomfort model and a movement prediction one. The authors in [103] suggested a social mapping algorithm that address the problem as a density estimation one. Towards this end, the method employs Kernel PCA[9] in order to provide ellipsoid solution in the 2D space. The respective social zones are provided in terms of iso-contours which result from kernel-based regression. An extension of this work in [104] utilizes skew-normal probability density functions to better describe the social zones while providing smoother transitions and adaptive spacing between successive time instances. An example is illustrated in Fig. 8.

Outdoor Interpretation. Several social mapping method rely on the Social Force Model (SFM) to predict human motion, such as the one in [105]. SFM steps on the concept that alternations in a persons trajectory can be explained in terms of social fields and forces. This work enhances this idea by relating an individual's trajectory with other nearby people or surrounding obstacles as repulsive effects. Regarding the prediction step, it is treated as sequential data classification one, thus employing a Bayesian predictor. The pedestrians trajectories along with the social force analysis shape the social map. Ferrer and Sanfeliu [106] expanded the SFM to represent interactions of three different types, namely person–person, object–person and robot–person. Shiomi et al. [107] used a SFM suitable for low-density situations like those occurring in a shopping mall. The authors in [108] use the SFM in conjunction with a multi-hypothesis motion prediction schema. Each hypothesis comprises a set of combinations regarding humans intentions and the hypothesis testing is examined via the computation of the joint probability of the intentions from all people in the scene. In [109], the Extended-SFM is introduced, allowing to compute a local path that minimizes the disturbances to other pedestrians in an online fashion. Moreover, this method exhibits proactive characteristics, thus changes in trajectories occur without being the reaction to events. Chung and Huang [110] present a methodology that combines the Spatial Behavior Cognition Model (SBCM) and feature-based Specific Spatial Effects (SSE). SBCM operates under the assumption that spatial responses of a person is affected by the environment, other humans or the place they are. Thus, SBCM models a cost function that describes the relations among the previously mentioned factors using location-based SSE. In this work, the pedestrian model depends on feature-based SSE which relates objects or features of the surrounding environment with human movements. Subsequently, the cost function computes a map with weights on which the robot navigates. The after-effect of each step of this work is depicted in Fig. 9. O' Callaghan et al. [111] incorporate social assumptions into the planning procedure by observing motion patterns of individuals. The latter is attained through GP, thus deriving a continuous probabilistic function that decides about the robot's trajectory. The work in [112] also learns

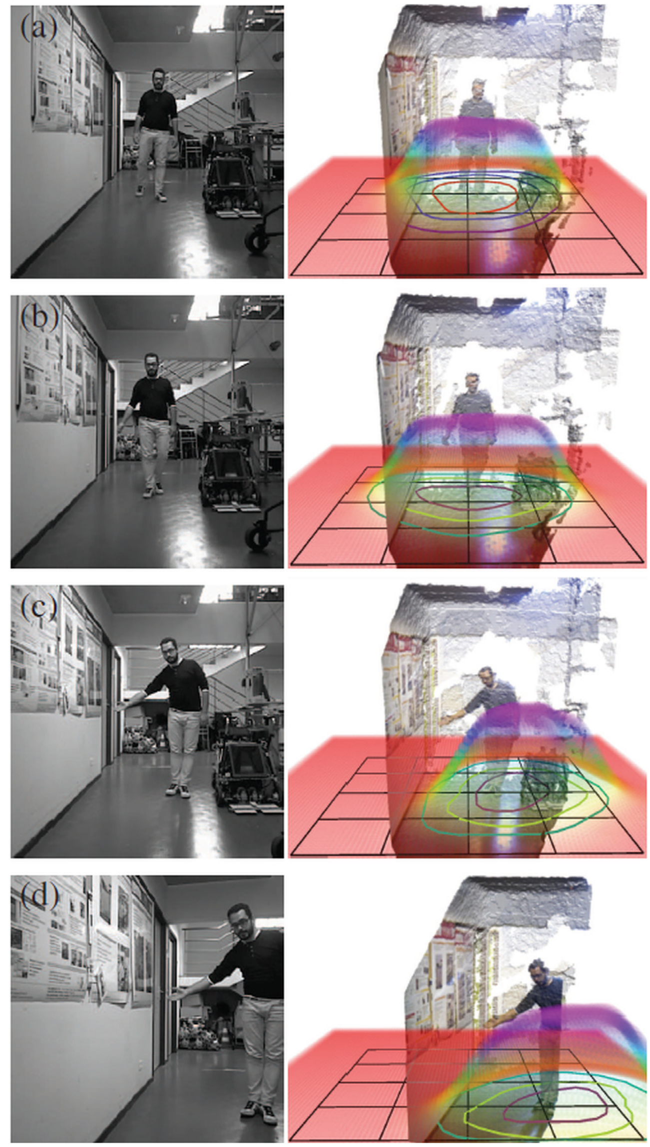


Fig. 8. Frames (a)–(d) show the concept of adaptive spacing by considering certainty and social cues in [104].

human motion patterns using an unsupervised approach which combines Dynamic Time Warping (DTW) with Bayes Information Criterion. Luber and Arras [113] proposed a framework that detects and learns socio-spatial relation among individuals. The framework is capable of inferring about the social groups and predict their motion. The authors in [114] exploit the approach of Inverse Reinforcement Learning (IRL) to model social interactions. The authors systematically evaluate different features which are given as input to the IRL and report the respective effects.

3.3.2. Predictive mechanism

The majority of social mapping techniques embed a kind of prediction mechanism regarding the trajectory of individuals and the reason is two-fold: first, social maps are dynamic in nature and achieving real-time update without any a-priori knowledge in conjunction with the current computing power is not feasible. Second, the prediction step allows to draw smooth trajectories, thus avoiding spasmodic alterations while cruising. Several methods follow a probabilistic approach, usually Bayesian or KF. The work in [113] uses a KF where the update step considers social

[9] PCA stands for Principal Component Analysis.

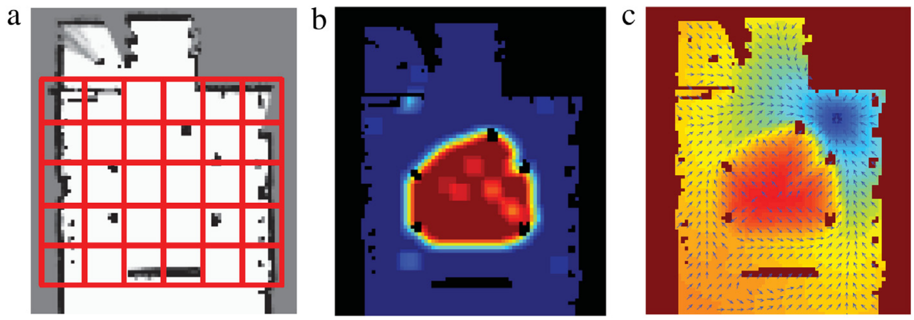


Fig. 9. (a) The occupancy grid map of environment: the red grid represents the coarse discrete cost function of SSE; (b) learned cost function of SSE and (c) the learned pedestrian policy of the method in [110]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

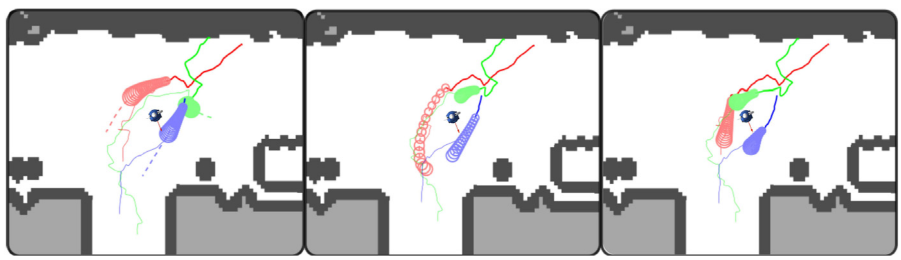


Fig. 10. Prediction example as presented in [108]. Each of the targets is depicted in the same color on each picture. On the left picture, the algorithm predicts that target green may stop, due to the interaction of the blue target and the robot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

relations among individuals as well as geometric ones. In order for the update step to occur, a probability distribution is estimated in which the samples are extracted through a Monte Carlo approach. Then, the weights used in KF are computed based on the factorization of the distribution and the drawn samples. The KF is also exploited in [115] for short-term and long-term prediction. The KF for short-term prediction considers a predefined individual motion model, while in the other one the motion model is a combination of individual long-term pedestrian models with different weights on each one. The authors in [108] define a set of behaviors which corresponds to the interaction between two persons along with a set of parameters. The proposed solution is a Hidden Markov Model that infers about the hidden behavior state and it is solved via an EM algorithm. The work in [111] learns human motion patterns using GP to estimate a nonparametric probabilistic model. Once the training is completed it becomes possible to predict the motion pattern. Xiao et al. [116] predict human motion using an SVM[10] to separate motion patterns into homogeneous classes. Afterwards, a clustering scheme occurs where a representative sample is selected from each cluster. Those representatives are exploited to predict a person's motion by matching the current pattern with a defined one and using the remaining as prediction (see Fig. 10).

3.3.3. Human body information

Social mapping methods implicitly demand a human detection scheme that supplies with the coordinates of an individual either in a global or a local frame. Yet, several methods make use of further information regarding humans such as body orientation or velocity to improve the mechanisms regarding the formation of social zones. A human model that comprises position, body orientation and velocity for each axis individually is used in [103] and it is updated in every time instance. This low-level information is exploited as follows: in the case the individual is static, a Gaussian distribution is placed with its center coinciding the person's coordinate position. The distribution is asymmetrically scaled and rotated according to the front direction of the torso. Considering the dynamic case, the magnitude of scaling is in proportion with the velocity. The work in [103] is expanded in [104], by means of a bivariate Skew-Normal distribution. Due to the parameters of the distribution, the latter may attain several different shapes namely egg-shape, ellipse and dominant-side which are all prominent according to the respective theoretical literature. This approach is considered a flexible one that can be used in various different cases, e.g. when approaching another person or obstacles. Ferrer et al. [105] used speed, direction and mass to model a persons trajectory, as illustrated in Fig. 11. However, the trajectory is also influenced from the models of the remaining people, objects in the surrounding area and the robot's position as well. Their subsequent work treats people as moving particles and the corresponding model takes into consideration the velocity and the acceleration in addition to the previously mentioned attributes [109]. Sisbot et al. [117] use the position and the body orientation to define a

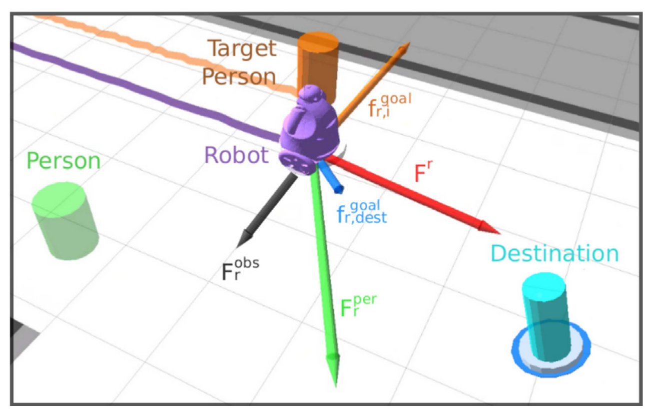


Fig. 11. Forces applied to the robot while accompanies a person as presented from the work in [105].

[10] SVM stands for Support Vector Machines.

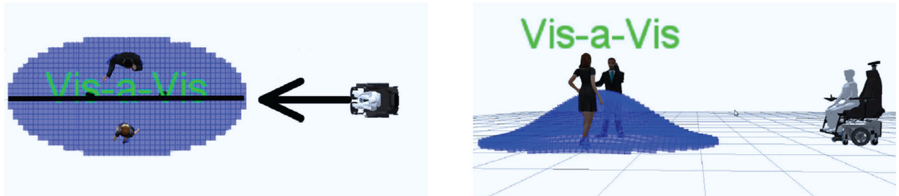


Fig. 12. O-space calculated by the Social Filter Module [120] for a Vis-a-Vis formation. The maximum risk of disturbance is located at o-space center, in the picture the disturbance is represented by the Gaussian function height.

semicircle that represents the FoV and defines two areas: the first one has a small radius and it is prohibited to the robot to trespass it; the second one has larger radius but its span is less than 2π and, so, interactions occur. Gómez et al. [118] used position, orientation and velocity to divide people into two sets one comprising individuals which are engaged in a social interaction while the second one comprising those who are not. The personal spaces are formed accordingly, a single person attains its own space while people belonging in a social group contribute to the group social area. The latter is an expansion of their work in [119]. In [102] position and orientation availed the definition of a discomfort model and a function defining a personal space, which can be either the area in front of a person or two people interacting.

3.3.4. Context awareness

Context awareness refers to those methodologies that seek relations, groups or interaction among individuals in order to shape the respective social map. The latter implies that such methods do not treat each person independently but seek associations among them and imprint those on the respective social map. The authors in [119] detect groups and configures the social map under the assumptions that the participants form a circle-shaped group. The subsequent work in [118] detects groups of people interacting and according to the robot's task shapes the social maps. The first one regards following the group, the second one approaching the group with the intention to interact and the last one the interaction itself. Cabello et al. [120] used a social filter to detect interactions. According to the type of interaction (e.g. vis-a-vis) and the geometrical attributes (e.g. the in-between distances of people) they compute the cost map of the surrounding area w.r.t. social conventions. An illustrated example is depicted in Fig. 12. Ball et al. [121] investigated approaching methods conforming to the way participants of a group are seated. The work regards three different setups, namely opposite to each other, perpendicular to one-another and side-by-side. The authors in [103,104] confront the case of multiple individuals by introducing the idea of interpersonal spaces as a combination of distributions referring to atomic social spaces. The latter leads to an overall density function that is used as an input to a one-class density estimation problem concluding to a map that strengthens the density within groups of individuals and ensures that the density is inflated on a person's position. The introduced taxonomy regarding the social mapping methodologies is summarized in Table 3 while the respective illustrative form is depicted in Fig. 13.

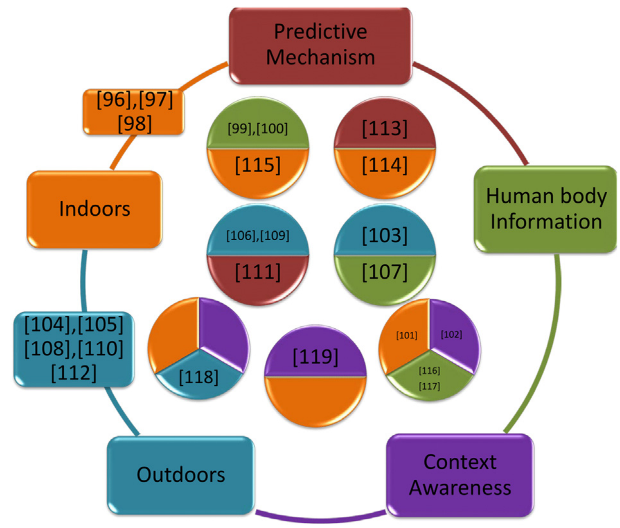


Fig. 13. Illustrative survey summary of social mapping methodologies. Cited works lying on the circumference of the outer circle suggest that they present only one of the examined attributes, i.e. the referred color. Colors in circular disks indicate methodologies that endow multiple attributes, corresponding to those colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Benchmarking, datasets and norms

The strict definition of metric mapping and SLAM techniques facilitates the derivation of metrics regarding the evaluation of such methodologies. The dominant one considers average localization error for every point, once the mapping of the entire map is concluded. Of course, this approach demands the existence of a ground truth. The authors in [122] introduced a framework for evaluating such methods, depending on relations between poses and not on the global frame. Moreover, this approach enables the comparison of methods with different estimation modules. There are several datasets available for such purposes allowing the evaluation of such method in a common ground. Additionally, they offer variability in terms of sensors and environment. The authors in [123] demonstrated a collection of RGB-D sequences using a Kinect sensor in indoor environments suitable for SLAM benchmarking. There are in total 39 sequences ranging from an office environment to an industrial one. Another dataset with the same characteristics, i.e. captured in indoors environment with Kinect, was presented in [124]. Considering the outdoor cases, the authors in [125] recorded a dataset of length equal to 36.8 km, utilizing a stereo camera and five laser scanners mounted on a vehicle. The dataset includes urban areas and it is convenient for testing algorithms under realistic cases. Likewise, the dataset in [126] comprises 22 sequences stereo sequences making appropriate for comparing monocular or stereo visual odometry, laser-based SLAM or fusion-based techniques.

The definition of the error may be straightforward in metric mapping, yet the latter is not true for the rest of the presented fields. Considering semantic mapping, relative methodologies may consist of various modules, relying on different cues in order to build the respective map. Moreover, the lack of a quantitative metric imposes the utilization of publicly available datasets, in order to compare such approaches. An in depth description of analogous datasets can be found in [10]. A recent proposition regarding the evaluation of such frameworks may be found in [127], which relies on the assumption a that a ground truth for semantic maps exists and provide a framework considering the formation of such

Table 3
Survey summary of social mapping methodologies.

Attribute/ method	Indoors	Outdoors	Predictive mechanism	Human body information	Context awareness
[98]	✓Human aware motion planner operating on a grid cell emulating Gaussian distribution to regulated the distance from human				
[99]	✓Voronoi space partitioning and rule based robot behavior				
[100]	✓Bayessian modeling of the relative pose of the human and robot incorporating also speech volume and human gestures				
[101]	✓RGB-D based navigation and grid cell space abstraction			✓Human detection and costmap modeling	
[102]	✓Social mapping based on adaptive optimization incorporating a discomfort model			✓Human movement prediction	
[103]	✓Kernel PCA for 2D space modeling			✓Body pose and orientation estimation	✓Regression to model human presence in high level
[104]	✓Modeling in 3D space			✓Skew-normal probability density functions, Interpersonal spaces modeling	✓Social zones approximation using smoother transitioning, Interpersonal spaces modeling
[105]		✓SFM to predict human motion		✓Direction and mass to model person's trajectory	
[106]		✓SFM to represent interactions among objects, humans and robot			
[107]		✓SFM suitable for low-density situations like those occurring in a shopping mall			
[108]		✓SFM with multi-hypothesis motion prediction schema	✓HMM to infer about the hidden behavior state, solved via EM algorithm		
[109]		✓Incorporation of objects and robot positioning		✓Human velocity and acceleration incorporation	
[110]		✓Spatial Behavior Cognition Model combined with feature based Specific Spatial Effects			
[111]		✓Social assumption integrated into planning procedure using motion patterns	✓Human motion learning patterns using GP		
[112]		✓Dynamic Time Warping with Bayes to learn human motion patterns			
[113]		✓Detection and learning of socio-spatial relation among individuals	✓KF to consider social relations among individuals and geometric features		

(continued on next page)

Table 3 (*continued*)

Attribute/ method	Indoors	Outdoors	Predictive mechanism	Human body information	Context awareness
[114]		✓Modeling social interactions through Inverse Reinforcement Learning			
[115]	✓Short and long term pedestrian modeling		✓KF for short-term motion prediction using predefined individual motion model		
[116]	✓Human motion modeling in open spaces		✓Human motion prediction using SVM		
[117]	✓Use of human models			✓Modeling human body and defining semicircles representing the FoV	
[118]	✓Task oriented social modeling based on groups of people			✓Group partitioning using person's position orientation and velocity	✓Task related creation of social map
[119]	✓Indoors social aware modeling			✓Modeling human presence using pose and velocity distribution	✓Group partitioning assuming that people form circle-shaped group
[120]	✓Operation in indoors and outdoors	✓Operation in indoors and outdoors			✓Social filtering for interactions detection and modeling on costmap
[121]	✓RGB-D based human detection				✓Modeling human approaching methods based on the way participants of a group are seated

datasets. The latter allows the fair comparison between different semantic mapping techniques.

Similar to semantic mapping, the social one suffers from the difficulty of defining a broadly accepted metric for evaluation purposes. Although the proxemics constitute a quantified approach regarding the personal spaces, one may intuitively argue that retaining those distances is not sufficient, whilst anthropologists consider them cultural dependent. Additional factors such as the velocity and the angle of approximation an individual are also crucial. To this extend several studies attempt to provide an answer/indication regarding the formula that evaluates such social norms. Takayama and Pantofaru [128] investigated issues concerning human personal space and conducted experiments in a controlled environment. Some of findings include that the sex affects the radius of the personal area, while the radius remains the same when a robot approaches an individual and vise-versa. The authors in [129] conduct a study examining how the distance between a human and a robot affects an individual's experience as well as the robot's perception. Henkel et al. [130] explore a scaling function to model human–robot distancing which was evaluated from a set of participants by completing a questionnaire. With respect to publicly available datasets, Carnegie Mellon University offers a collection of 55 sequences [131] referring to social interactions, ranging from walking and conversing to dancing. This dataset of motion capture data is recorded using 12 infrared cameras at the refresh rate of 120 Hz and the resolution of the images is equal to 4 megapixels. The available space for capturing the mentioned interaction is a 3 m $\times$ 8 m area and all participants wear jumpsuit along with 41 markers. Potential users can exploit either the position of the markers or the respective skeleton movement. Regarding the outdoor cases, a well-known dataset is the Edinburgh Informatics Forum Pedestrian Database (EIPD) [132]. The latter is recorded using a fixed camera at 23 m height, looking towards a hallway of the respective campus, covering an area of 15.8 m $\times$ 11.8 m for a few months period. The sensor provides RGB data at 9 Hz rate, while the resolution is at 640 $\times$ 480. The dataset comprises over 92,000 observed trajectories accompanied with the corresponding bounding boxes.

5. Open issues and questions

Enabling robots to co-exist in a human-populated environment requires tremendous research in the previously mentioned fields. A lot of research work has been conducted aiming to bridge the gap in HRI by enhancing robots with capabilities that resemble humans, or at least obey social norms deriving from humanity. Yet, those fields rose in the last decade and this relatively short-term occupation restated known questions or formulated new ones. Confronting issues such as the following ones will decrease the distance from the aim.

What is the need for social maps? Should not dynamic obstacle avoidance be sufficient?

Social maps cope with the issue of human-aware robot coexistence and principally emphasize in robot navigation within human environments. While autonomous navigation is limited to obstacle avoidance and reaching the goal, social navigation should additionally consider elements including human comfort, naturalness and sociability. Human comfort is relevant to the navigation manner providing the feeling of safety to an individual. Although safety can be achieved by simply avoiding humans in an autonomous navigation manner, the trajectory might be spiking with harsh stops and starts. In such a case the human user losses the feeling of safety due to the accruing discomfort. Towards the same direction naturalness involves the extraction of robot paths similar to the ones extracted by humans, which usually is obtained by adjusting the profile of the acceleration to be continuous, leading to a smooth velocity profile and a polynomial function for the trajectory. Last, sociability concerns abstract decisions about robot's action according to social and ethical notions (e.g. a social robot should not interrupt a conversation by passing though a group of people, but by them). Towards this end, contemporary robots should be attached with roles, currently being performed by humans. Such roles require the robot to be endowed with social skills and appropriate behavioral models, allowing to apprehend the user and respond appropriately.

How do we evaluate social mapping techniques?

There are two different notions that can be followed regarding this issue. The first one comprises the annotation of datasets where an individual traverses in a human populated environment and his/her pose is labeled along with the respective time-stamp. Then, the robot's trajectory can be compared with the respective ground truth allowing the usage of metrics similar to the ones used in metric maps. The drawback of this case is owing to the fact that human's trajectory is not always unique, but there might be another socially acceptable one, but different. The second notion considers the evaluation of the robot's behavior from the participants themselves, by means of an appropriate questionnaire. This option enables the evaluation of terms that up to now are only intuitively described, such as comfort, naturalness and sociability [18]. However, the nature of those experiments do not provide repeatability which is also an important aspect of evaluation.

Is social mapping methods adequate for crowded applications?

Although there exist methodologies for indoor and outdoor cases, each one having its own characteristics, an analogy holds between them. Indoor cases regard smaller number of individuals but in confined spaces, in contrast to the outdoor one. If the ratio between the number of people and the available area increases, conjecture and difficulties to robot's cruising will turn out. Considering the difficulties a person confronts in such cases, the case of the robot is more complicated due to limited perception of the environment, prediction mechanism and degrees of freedom in movement. Such occasions may be dealt by imposing specific policies upon their detection that will attach a kind of neutrality, waiting for the necessary availability to return to the default social mapping module.

Is there a need for combining outdoors and indoors schemata?

Outdoors approaches tend to learn motion patterns from pedestrians; either to imitate them or to predict their movement [111]. Moreover, in order to handle the people's flow and for computational purposes, many approaches seek groups in the population in order to simplify the problem at hand [113]. Such methods provide socially acceptable mapping suitable for open areas with several pedestrians. Yet, indoor ones concentrate on smaller scale problem where there is limited space for cruising and study in depth the positioning of individuals. Due to the restrictions imposed in such environments instead of providing with solutions for generic cases they reach for precise results and fine-tuned maps. However, a robot will need both strategies in order to cope with real life cases, thus, there is a need for blending those two notions to enhance robot's capabilities and act appropriately in more cases.

How far are we today from using socially adequate robots in domestic and working environments?

An affirmative answer to such a question implies that the robots are be able to: (i) understand, interpret and represent their environment in a human compatible manner; (ii) apprehend the human occurrence and its activities so as to act accordingly in human inhabited environments and (iii) retain behavioral models that facilitate seamless cooperation among the two aforementioned principals.

A response to the first note is the semantic mapping, which has been proved a reliable human-oriented modeling of the explored environment, thus bridging the gap between robots and humans concerning the apprehension of the space. As mentioned in the respective section, it relates recognized objects and places with their spatial arrangement extracted from the environment's metric map. The produced map is also augmented with the inherent temporal proximity during the robot's movement enabling the description of the physical constraints within the explored places. By retaining such a map the robot could perform high-level tasks, such as navigation and fetching ones. However, the co-existence of humans and robots in the same environment entails the risk of violating the Asimov's first law. The solution to this barrier, which is also the response to the second principal and may be sought in the social mapping. In order for a successful human–robot co-existence, social mapping is among the main features being handy when integrating human social intelligence into robotic perception and action planning. Social mapping endows domestic robots with the ability to identify the humans' existence in a place and to recognize their current state, in order to safely run any task. This can be accomplished by embodying the modeled social behaviors, the humans' posture and action detection into the robots' perception mechanism. The third principal, which is the link between the semantic and social mapping, comprises the development of robot behavioral models, according to which the agent should be able to perform its regular assignments without being interrupted by the normal human activity, the environment constraints and its attributes. Taking into account the evolvement of the social and semantic mapping, the ground has now become fertile to establish robot behavioral models. Such a model would equip the artificial agent with the framework to perform a daily schedule involving household tasks, human serving, elder people assisting etc. The behavioral model that describes this schedule is essential to be adaptive, in order to comply with the human activities, which are mainly dynamic and entail high variability. The ability of the robot to adapt and differentiate its regular tasks according to the existence and deployment of humans within its operating space may define its behavioral model.

How do we assess the role of social aware navigation in shaping intelligent robotic agents in the near future?

Familiarity is an aspect of importance in working environments and even more in domestic ones. Social aware navigation aims towards this direction and it is expected in the near future methodologies that provide comfort, naturalness, sociability and high level capacities. The endowment of robots with such social capabilities facilitates their acceptance from people, even from sceptic ones and pave the way for co-existence. Gradually, it is expected from humans to weigh feelings of mistrust and doubt with respect to robot agents due to the fact that such agents understand and respect social norms. This sort of acceptance is anticipated to expand the already established market of domestic robots and consequently social aware navigation will draw further attention from both the research community as well as from the industry. Such sort of navigation has already become part of current robotic research projects and will be an indispensable one in the proximal future.

6. Motivation for future work

From the analysis conducted herein regarding the social navigation and mapping field, it is revealed that there are several limitations constraining existing methodologies to operate under controlled conditions or at small scale. This is mainly due to the fact that the development of a complete social navigation method demands the existence of fundamental perception layers, such as the human and environment monitoring and modeling, the metric and semantic mapping for abstract place representation and the analysis of the human role within the scene. These limitations can be outperformed by incorporating several combinatorial techniques to all the layers of the hierarchy of social mapping and navigation, beginning from the environment perception and extending to the modeling of the robot's navigation rules utilizing at the same time the observed human actions.

Feature description: The detection and description of features comprises the fundamental basis of the robot perception of the environment; it compensates noise presence in any vision-based measurement. Although this strategy performs well during developing stand-alone modules such as metric mapping or human action recognition, when it comes to the implementation of intricate systems, such as social mapping ones, the utilization of various feature description methodologies complicates the architecture in a certain degree and renders the realization of large scale applications difficult. A key solution to this problem could be the introduction of features able to be reused during the developing of the social mapping framework. Such features should allow modeling both the environment and the human activity. An alternative would be the features description provided directly by RGB-D sensors, such as human skeleton detection, which can model the human presence within specific regions. Furthermore, human interaction with objects, obtained from the metric map, can label the regions in topological maps in context-wise manner adjusting respectively the robot navigation rules. Additionally, features embedding temporal proximity such as patterns of human trajectories can also be used to predict congestive regions along the robot coarse leading to more efficient social-aware navigation.

Metric mapping and optimization: It is an evident that a lot of laborious research has already been conducted in the field of robot mapping and navigation during the last decades that led to outstanding progress in the respective field, involving the ability to construct large scale metric maps with competent accuracy. Various optimization modalities and sensor fusion techniques pushed towards this direction, by their integration into SLAM methods. However, the robots should be designed to operate in dynamic environments and, therefore, continuous update of the reconstructed maps is required. The term lifelong mapping initially introduced in [133] should be extended to cover large changes in the indoor environment (e.g. replace a sofa with an arm chair) rather than considering only moving elements e.g. human motion, that can be typically handled as outliers during the robot's motion estimation. Such a workaround attempted from the most recent work presented in [134] where Octomapping supported dynamic changes and representations of the explored environment. However, this methodology can operate sufficiently only on with environmental changes and the computational effort increases when the changes cover large areas of the 3D occupancy grid. A future solution, that would support lifelong metric mapping would be the application of various optimization solutions depending on the context of the environmental change, while drastic map update would be emerged solely on large changes of the environment topology.

Topological and temporal coherence: can augment a thorough representation of the environment. On the one hand, the organization of the metric map into topological regions with hierarchical dependencies on the metric representations, i.e. octrees, combined with semantic information about the nodes of the topological maps, could directly connect the metric with the semantic mapping. On the other hand, the inherent temporal proximity of the sensors' measurements during the robot motion could lead to the utilization of less optimization strategies required for the metric mapping. Additionally, temporal coherence could further augment the information stored to the edges of the topological map indicating transition from semantically reach to semantically poor regions of the environment.

7. Coda

The scope of this review paper is to present state-of-the-art methodologies that advance human–robot symbiosis. Several characteristics were drawn and the respective methodologies were clustered accordingly. This work demonstrated the fact that for each characteristic there exist a variability of methods w.r.t. the problem they attempt to solve. While each section has a different level of maturity, they all possess common environmental constrains. The definition of an evaluation metric in cases like semantic mapping, social mapping and behavior is a complex assignment which further perplexes the comparison and selection of a methodology. However, due to this difficulty a pluralism of algorithms with different starting points emerged, which rapidly enriches the respective literature. Even though important steps have been made, the effectiveness of the those works need to be further improved and standardize in order for robots to be accepted and co-exist in human populated environments in a daily fashion; while there are important aspects, as presented in Section 5 that is essential to be confronted. Such constrains need to be exceeded in order to initiate discussions about a common framework that comprises seamlessly operating modules from all the sections mentioned in this paper.

References

[1] I.R. Nourbakhsh, D. Andre, C. Tomasi, M.R. Genesereth, Mobile robot obstacle avoidance via depth from focus, Robot. Auton. Syst. (ISSN: 0921-8890) 22 (2) (1997) 151– 158. http://dx.doi.org/10.1016/S0921-8890(97)00051-1.

[2] I. Ulrich, J. Borenstein, VFH+: reliable obstacle avoidance for fast mobile robots, in: Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on, vol. 2, (ISSN: 1050-4729), 1998, pp. 1572–1577. http://dx.doi.org/10.1109/ROBOT.1998.677362.

[3] L. Nalpantidis, A. Gasteratos, Stereovision-based fuzzy obstacle avoidance method, Int. J. Humanoid Robot. 08 (01) (2011) 169–183. http://dx.doi.org/10.1142/S0219843611002381.

[4] I.J. Cox, Autonomous Robot Vehicles, ISBN: 978-1-4613-8997-2, 1990, pp. 221–228. http://dx.doi.org/10.1007/978-1-4613-8997-2_17.

[5] T. Lozano-Perez, I.J. Cox, G.T. Wilfong, Autonomous Robot Vehicles, Springer Science & Business Media, 2012.

[6] L. Nalpantidis, G.C. Sirakoulis, A. Gasteratos, Non-probabilistic cellular automata-enhanced stereo vision simultaneous localization and mapping, Meas. Sci. Technol. 22 (11) (2011) 114027. http://dx.doi.org/10.1088/0957-0233/22/11/114027.

[7] A. Adheen, V. D, A survey on sensing methods and feature extraction algorithms for SLAM problem, CoRR abs/1303.3605, 2013. URL http://arxiv.org/abs/1303.3605.

[8] H. Durrant-Whyte, T. Bailey, Simultaneous localization and mapping: part I, IEEE Robot. Autom. Mag. (ISSN: 1070-9932) 13 (2) (2006) 99–110. http://dx.doi.org/10.1109/MRA.2006.1638022.

[9] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (SLAM): part II, IEEE Robot. Autom. Mag. (ISSN: 1070-9932) 13 (3) (2006) 108–117. http://dx.doi.org/10.1109/MRA.2006.1678144.

[10] I. Kostavelis, A. Gasteratos, Semantic mapping for mobile robotics tasks: A survey, Robot. Auton. Syst. (ISSN: 0921-8890) 66 (2015) 86–103. http://dx.doi.org/10.1016/j.robot.2014.12.006.

[11] M. Goodrich, A. Schultz, Human-robot interaction: a survey, Found. Trends Hum.-Comput. Interact. 1 (3) (2007) 203–275.

[12] S. Tellex, T. Kollar, S. Dickerson, M.R. Walter, A.G. Banerjee, S. Teller, N. Roy, Understanding natural language commands for robotic navigation and mobile manipulation, in: Proceedings of the National Conference on Artificial Intelligence, AAAI, San Francisco, CA, 2011, pp. 1507–1514.

[13] D.L. Chen, R.J. Mooney, Learning to interpret natural Language navigation instructions from observations, in: Proceedings of the 25th AAAI Conference on Artificial Intelligence, AAAI, 2011, pp. 859–865.

[14] K. Nickel, R. Stiefelhagen, Visual recognition of pointing gestures for human–robot interaction, Image Vis. Comput. (ISSN: 0262-8856) 25 (12) (2007) 1875–1884. http://dx.doi.org/10.1016/j.imavis.2005.12.020.

[15] M. Pateraki, H. Baltzakis, P. Trahanias, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation, Comput. Vis. Image Underst. (ISSN: 1077-3142) 120 (2014) 1–13. http://dx.doi.org/10.1016/j.cviu.2013.12.006.

[16] C.L. Breazeal, Designing Sociable Robots, MIT press, 2004.

[17] R.S. Pol, M. Murugan, A review on indoor human aware autonomous mobile robot navigation through a dynamic environment survey of different path planning algorithm and methods, in: Industrial Instrumentation and Control, ICIC, International Conference on, 2015, pp. 1339–1344. http://dx.doi.org/10.1109/IIC.2015.7150956.

[18] J. Rios-Martinez, A. Spalanzani, C. Laugier, From proxemics theory to socially-aware navigation: A survey, Int. J. Soc. Robot. 7 (2) (2015) 137–153.

[19] T. Kruse, A. Pandey, R. Alami, A. Kirsch, Human-aware robot navigation: A survey, Robot. Auton. Syst. 61 (12) (2013) 1726–1743.

[20] T.W. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots: Concepts, design, and applications (CMU-RI-TR-02-29), 2002.

[21] E. Mitka, A. Gasteratos, N. Kyriakoulis, S. Mouroutsos, Safety certification requirements for domestic robots, Saf. Sci. (ISSN: 0925-7535) 50 (9) (2012) 1888–1897. http://dx.doi.org/10.1016/j.ssci.2012.05.009.

[22] E.T. Hall, The Hidden Dimension, Doubleday & Co, 1966.

[23] E. Kim, L.D. Berkovits, E. Bernier, D. Leyzberg, F. Shic, R. Paul, B. Scassellati, Social robots as embedded reinforcers of social behavior in children with autism, J. Autism Dev. Disord. 43 (5) (2013) 1038–1049.

[24] H. Kozima, M.P. Michalowski, C. Nakagawa, Keepon: a playful robot for research, therapy, and entertainment, Int. J. Soc. Robot. (ISSN: 1875-4805) 1 (1) (2008) 3–18. http://dx.doi.org/10.1007/s12369-008-0009-8.

[25] J.G. Prez, M. Lohse, V. Evers, Position paper: robots as companions and therapists in elderly care, in: Taking Care of Each Other, ICSR Workshop: Synchronisation and Reciprocity for Social Companion Robots, Vienna University of Technology, Vienna, Austria, 2013, pp. 11–14.

[26] R. Bemelmans, G.J. Gelderblom, P. Jonker, L. De Witte, Socially assistive robots in elderly care: A systematic review into effects and effectiveness, J. Amer. Med. Dir. Assoc. 13 (2) (2012) 114–120.

[27] E. Garcia-Fidalgo, A. Ortiz, Vision-based topological mapping and localization methods: A survey, Robot. Auton. Syst. 64 (2015) 1–20.

[28] P. Newman, On the Structure and Solution of the Simultaneous Localisation and Map Building Problem (Doctoral diss.), University of Sydney, 1999.

[29] A.J. Davison, D.W. Murray, Simultaneous localization and map-building using active vision, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 865–880.

[30] S. Thrun, et al., Robotic mapping: A survey, Exploring artificial intelligence in the new millennium (2002) 1–35.

[31] J. Aulinas, Y. Petillot, J. Salvi, X. Lladó, The SLAM problem: A survey, in: Proceedings of the Conference on Artificial Intelligence Research and Development, IOS Press, ISBN: 978-1-58603-925-7, 2008, pp. 363–371.

[32] G. Dissanayake, S. Huang, Z. Wang, R. Ranasinghe, A review of recent developments in simultaneous localization and mapping, in: Industrial and Information Systems, ICIIS, 6th IEEE International Conference on, IEEE, 2011, pp. 477–482.

[33] J. Fuentes-Pacheco, J. Ruiz-Ascencio, J.M. Rendón-Mancha, Visual simultaneous localization and mapping: a survey, Artif. Intell. Rev. 43 (1) (2015) 55–81.

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

[35] Y. Baudoin, D. Doroftei, G. De Cubber, S.A. 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: Safety, Security & Rescue Robotics, SSRR, IEEE International Workshop on, IEEE, 2009, pp. 1–6.

[36] J. Penders, Viewfinder: Final Activity Report, Sheffield Hallam University, Materials and Engineering Research Institute, 2010. URL http://shura.shu.ac.uk/217/.

[37] B. Kuipers, R. Froom, W.-Y. Lee, D. Pierce, The semantic hierarchy in robot learning, in: J.H. Connell, S. Mahadevan (Eds.), Robot Learning, Springer, US, Boston, MA, ISBN: 978-1-4615-3184-5, 1993, pp. 141–170. http://dx.doi.org/10.1007/978-1-4615-3184-5_6.

[38] B. Kuipers, The spatial semantic hierarchy, Artificial Intelligence (ISSN: 0004-3702) 119 (12) (2000) 191–233. http://dx.doi.org/10.1016/S0004-3702(00)00017-5.

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

[40] A. Pronobis, B. Caputo, Confidence-based cue integration for visual place recognition, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2007, pp. 2394–2401. http://dx.doi.org/10.1109/IROS.2007.4399493.

[41] E. Pacchierotti, H. Christensen, P. Jensfelt, et al., Human-robot embodied interaction in hallway settings: a pilot user study, in: Robot and Human Interactive Communication, IEEE International Workshop on, IEEE, 2005, pp. 164–171.

[42] K. Dautenhahn, M. Walters, S. Woods, K.L. Koay, C.L. Nehaniv, A. Sisbot, R. Alami, T. Siméon, How may I serve you?: a robot companion approaching a seated person in a helping context, in: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, ACM, 2006, pp. 172–179.

[43] Y. Nakauchi, R. Simmons, A social robot that stands in line, Auton. Robots 12 (3) (2002) 313–324.

[44] E. Pacchierotti, H.I. Christensen, P. Jensfelt, Embodied social interaction for service robots in hallway environments, in: Field and Service Robotics, Springer, 2006, pp. 293–304.

[45] I. Leite, C. Martinho, A. Paiva, Social robots for long-term interaction: a survey, Int. J. Soc. Robot. 5 (2) (2013) 291–308.

[46] T. Kanda, T. Hirano, D. Eaton, H. Ishiguro, Interactive robots as social partners and peer tutors for children: A field trial, Hum.-Comput. Interact. 19 (1) (2004) 61–84.

[47] T. Salter, K. Dautenhahn, R. Bockhorst, Robots moving out of the laboratory-detecting interaction levels and human contact in noisy school environments, in: Robot and Human Interactive Communication, 13th IEEE International Workshop on, 2004, pp. 563–568.

[48] L. Giusti, P. Marti, Interpretative dynamics in human robot interaction, in: Robot and Human Interactive Communication, 15th IEEE International Symposium on, 2006, pp. 111–116.

[49] K. Charalampous, I. Kostavelis, A. Gasteratos, Robot navigation in large-scale social maps: An action recognition approach, Expert Syst. Appl. 66 (2016) 261–273. http://dx.doi.org/10.1016/j.eswa.2016.09.026.

[50] G. Pascoe, W. Maddern, A.D. Stewart, P. Newman, FARLAP: Fast robust localisation using appearance priors, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 6366–6373. http://dx.doi.org/10.1109/ICRA.2015.7140093.

[51] L. Carlone, F. Dellaert, Duality-based verification techniques for 2D SLAM, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4589–4596. http://dx.doi.org/10.1109/ICRA.2015.7139835.

[52] M. Kaess, Simultaneous localization and mapping with infinite planes, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4605–4611. http://dx.doi.org/10.1109/ICRA.2015.7139837.

[53] K.Y.K. Leung, F. Inostroza, M. Adams, Generalizing random-vector SLAM with random finite sets, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4583–4588. http://dx.doi.org/10.1109/ICRA.2015.7139834.

[54] S. Choudhary, V. Indelman, H.I. Christensen, F. Dellaert, Information-based reduced landmark SLAM, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4620–4627. http://dx.doi.org/10.1109/ICRA.2015.7139839.

[55] H. Lee, J. Park, W. Chung, Curb feature based localization of a mobile robot in urban road environments, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 2794–2799. http://dx.doi.org/10.1109/ICRA.2015.7139579.

[56] R. Arroyo, P.F. Alcantarilla, L.M. Bergasa, E. Romera, Towards life-long visual localization using an efficient matching of binary sequences from images, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 6328–6335. http://dx.doi.org/10.1109/ICRA.2015.7140088.

[57] G. Zhang, P.A. Vela, Optimally observable and minimal cardinality monocular SLAM, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 5211–5218. http://dx.doi.org/10.1109/ICRA.2015.7139925.

[58] L.J. Lyrio, T. Oliveira-Santos, C. Badue, A.F.D. Souza, Image-based mapping, global localization and position tracking using vg-ram weightless neural networks, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 3603–3610. http://dx.doi.org/10.1109/ICRA.2015.7139699.

[59] M. Jalobeanu, G. Shirakyan, G. Parent, H. Kikkeri, B. Peasley, A. Feniello, Reliable kinect-based navigation in large indoor environments, in: IEEE International Conference on Robotics and Automation, ICRA, 2015 URL http://research.microsoft.com/apps/pubs/default.aspx?id=249742.

[60] M. Dymczyk, S. Lynen, T. Cieslewski, M. Bosse, R. Siegwart, P. Furgale, The gist of maps - summarizing experience for lifelong localization, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 2767–2773. http://dx.doi.org/10.1109/ICRA.2015.7139575.

[61] R.W. Wolcott, R.M. Eustice, Fast LIDAR localization using multiresolution Gaussian mixture maps, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 2814–2821. http://dx.doi.org/10.1109/ICRA.2015.7139582.

[62] I. Kostavelis, A. Gasteratos, Semantic maps from multiple visual cues, Expert Syst. Appl. 68 (2017) 45–57. http://dx.doi.org/10.1016/j.eswa.2016.10.014.

[63] A.C. Müller, S. Behnke, Learning depth-sensitive conditional random fields for semantic segmentation of RGB-D images, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 6232–6237. http://dx.doi.org/10.1109/ICRA.2014.6907778.

[64] D. Wolf, J. Prankl, M. Vincze, Fast semantic segmentation of 3D point clouds using a dense CRF with learned parameters, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4867–4873. http://dx.doi.org/10.1109/ICRA.2015.7139875.

[65] J. Mason, B. Marthi, R. Parr, Unsupervised discovery of object classes with a mobile robot, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 3074–3081. http://dx.doi.org/10.1109/ICRA.2014.6907301.

[66] D. Gutirrez-Gmez, W. Mayol-Cuevas, J.J. Guerrero, What should I landmark? Entropy of normals in depth juts for place recognition in changing environments using RGB-D data, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 5468–5474. http://dx.doi.org/10.1109/ICRA.2015.7139963.

[67] H. Carrillo, Y. Latif, J. Neira, J.A. Castellanos, Place categorization using sparse and redundant representations, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 4950–4957. http://dx.doi.org/10.1109/IROS.2014.6943266.

[68] A. Hermans, G. Floros, B. Leibe, Dense 3D semantic mapping of indoor scenes from RGB-D images, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 2631–2638. http://dx.doi.org/10.1109/ICRA.2014.6907236.

[69] A. Ranganathan, PLISS: labeling places using online changepoint detection, Auton. Robots (ISSN: 0929-5593) 32 (4) (2012) 351–368. http://dx.doi.org/10.1007/s10514-012-9273-4.

[70] D. Galvez-Lopez, J.D. Tardos, Bags of binary words for fast place recognition in image sequences, IEEE Trans. Robot. (ISSN: 1552-3098) 28 (5) (2012) 1188–1197. http://dx.doi.org/10.1109/TRO.2012.2197158.

[71] D.V. Lu, D. Hershberger, W.D. Smart, Layered costmaps for context-sensitive navigation, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 709–715. http://dx.doi.org/10.1109/IROS.2014.6942636.

[72] L.L.S. Wong, L.P. Kaelbling, T. Lozano-Perez, Not seeing is also believing: Combining object and metric spatial information, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 1253–1260. http://dx.doi.org/10.1109/ICRA.2014.6907014.

[73] Z. Zhao, X. Chen, Semantic mapping for object category and structural class, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 724–729. http://dx.doi.org/10.1109/IROS.2014.6942638.

[74] S. Choudhary, A.J.B. Trevor, H.I. Christensen, F. Dellaert, SLAM with object discovery, modeling and mapping, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 1018–1025. http://dx.doi.org/10.1109/IROS.2014.6942683.

[75] M. Oliveira, G.H. Lim, L.S. Lopes, S.H. Kasaei, A.M. Tom, A. Chauhan, A perceptual memory system for grounding semantic representations in intelligent service robots, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 2216–2223. http://dx.doi.org/10.1109/IROS.2014.6942861.

[76] H. Deeken, T. Wiemann, K. Lingemann, J. Hertzberg, SEMAP - a semantic environment mapping framework, in: Mobile Robots, ECMR, European Conference on, 2015, pp. 1–6. http://dx.doi.org/10.1109/ECMR.2015.7324176.

[77] J. Cleveland, D. Thakur, P. Dames, C. Phillips, T. Kientz, K. Daniilidis, J. Bergstrom, V. Kumar, An automated system for semantic object labeling with soft object recognition and dynamic programming segmentation, in: Automation Science and Engineering, CASE, IEEE International Conference on, 2015, pp. 683–690. http://dx.doi.org/10.1109/CoASE.2015.7294159.

[78] V. Vineet, O. Miksik, M. Lidegaard, M. Nießer, S. Golodetz, V.A. Prisacariu, O. Kähler, D.W. Murray, S. Izadi, P. Pérez, P.H.S. Torr, Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 75–82. http://dx.doi.org/10.1109/ICRA.2015.7138983.

[79] S. Sengupta, P. Sturgess, Semantic octree: Unifying recognition, reconstruction and representation via an octree constrained higher order MRF, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 1874–1879. http://dx.doi.org/10.1109/ICRA.2015.7139442.

[80] R. Zhang, S.A. Candra, K. Vetter, A. Zakhor, Sensor fusion for semantic segmentation of urban scenes, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 1850–1857. http://dx.doi.org/10.1109/ICRA.2015.7139439.

[81] A. Mousavian, J. KoÅeck, J.-M. Lien, Semantically guided location recognition for outdoors scenes, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 4882–4889. http://dx.doi.org/10.1109/ICRA.2015.7139877.

[82] C. Cadena, J. KoÅeck, Semantic segmentation with heterogeneous sensor coverages, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 2639–2645. http://dx.doi.org/10.1109/ICRA.2014.6907237.

[83] E. Stumm, C. Mei, S. Lacroix, M. Chli, Location graphs for visual place recognition, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 5475–5480. http://dx.doi.org/10.1109/ICRA.2015.7139964.

[84] R. Paul, R. Triebel, D. Rus, P. Newman, Semantic categorization of outdoor scenes with uncertainty estimates using multi-class gaussian process classification, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2012, pp. 2404–2410. http://dx.doi.org/10.1109/IROS.2012.6386073.

[85] 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: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2011, pp. 1249–1255. http://dx.doi.org/10.1109/IROS.2011.6094638.

[86] J. Martínez-Gómez, V. Morell, M. Cazorla, I. García-Varea, Semantic localization in the PCL library, Robot. Auton. Syst. 75 (2016) 641–648.

[87] M. Mohan, D. Gálvez-López, C. Monteleoni, G. Sibley, Environment selection and hierarchical place recognition, in: Robotics and Automation (ICRA), IEEE International Conference on, 2015, pp. 5487–5494. http://dx.doi.org/10.1109/ICRA.2015.7139966.

[88] R. Drouilly, P. Rives, B. Morisset, Semantic representation for navigation in large-scale environments, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 1106–1111. http://dx.doi.org/10.1109/ICRA.2015.7139314.

[89] H. Grimmett, M. Buerki, L. Paz, P. Pinies, P. Furgale, I. Posner, P. Newman, Integrating metric and semantic maps for vision-only automated parking, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 2159–2166. http://dx.doi.org/10.1109/ICRA.2015.7139484.

[90] O.M. Mozos, W. Burgard, Supervised learning of topological maps using semantic information extracted from range data, in: Intelligent Robots and Systems, IEEE/RSJ International Conference on, 2006, pp. 2772–2777. http://dx.doi.org/10.1109/IROS.2006.282058.

[91] B.C. Akdeniz, H.I. Bozma, Exploration and topological map building in unknown environments, in: Robotics and Automation, ICRA, IEEE Social Conference on, 2015, pp. 1079–1084. http://dx.doi.org/10.1109/ICRA.2015.7139310.

[92] O. Erkent, H.I. Bozma, Long-term topological place learning, in: Robotics and Automation, ICRA, IEEE Social Conference on, 2015, pp. 5462–5467. http://dx.doi.org/10.1109/ICRA.2015.7139962.

[93] F. Bernuy, J. Ruiz del Solar, Semantic mapping of large-scale outdoor scenes for autonomous off-road driving, in: IEEE Social Conference on Computer Vision Workshop, ICCVW, 2015, pp. 124–130. http://dx.doi.org/10.1109/ICCVW.2015.26.

[94] R. Drouilly, P. Rives, B. Morisset, Fast hybrid relocation in large scale metric-topologic-semantic map, in: Intelligent Robots and Systems, IROS, IEEE/RSJ Social Conference on, 2014, pp. 1839–1845. http://dx.doi.org/10.1109/IROS.2014.6942804.

[95] S. Laible, A. Zell, Building local terrain maps using spatio-temporal classification for semantic robot localization, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, IEEE, Chicago, IL, 2014, pp. 4591–4597. http://dx.doi.org/10.1109/IROS.2014.6943213.

[96] S. Hemachandra, M.R. Walter, S. Tellex, S. Teller, Learning spatial-semantic representations from natural language descriptions and scene classifications, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 2623–2630. http://dx.doi.org/10.1109/ICRA.2014.6907235.

[97] I. Kostavelis, K. Charalampous, A. Gasteratos, J.K. Tsotsos, Robot navigation via spatial and temporal coherent semantic maps, Eng. Appl. Artif. Intell. (ISSN: 0952-1976) 48 (2016) 173–187. http://dx.doi.org/10.1016/j.engappai.2015.11.004.

[98] E.A. Sisbot, L.F. Marin-Urias, R. Alami, T. Simeon, A human aware mobile robot motion planner, IEEE Trans. Robot. (ISSN: 1552-3098) 23 (5) (2007) 874–883. http://dx.doi.org/10.1109/TRO.2007.904911.

[99] A.K. Pandey, R. Alami, A framework towards a socially aware mobile robot motion in human-centered dynamic environment, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2010, pp. 5855–5860. http://dx.doi.org/10.1109/IROS.2010.5649688.

[100] R. Mead, M.J. Matarić, Perceptual models of human-robot proxemics, in: The 14th International Symposium on Experimental Robotics, ISBN: 978-3-319-23778-7, 2016, pp. 261–276. http://dx.doi.org/10.1007/978-3-319-23778-7_18.

[101] K. Charalampous, C. Emmanouilidis, A. Gasteratos, Social mapping on RGB-D scenes, in: Imaging Systems and Techniques, IST, IEEE International Conference on, 2014, pp. 398–403. http://dx.doi.org/10.1109/IST.2014.6958512.

[102] J. Rios-Martinez, A. Renzaglia, A. Spalanzani, A. Martinelli, C. Laugier, Navigating between people: A stochastic optimization approach, in: Robotics and Automation, ICRA, IEEE International Conference on, 2012, pp. 2880–2885. http://dx.doi.org/10.1109/ICRA.2012.6224934.

[103] P. Papadakis, A. Spalanzani, C. Laugier, Social mapping of human-populated environments by implicit function learning, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2013, pp. 1701–1706. http://dx.doi.org/10.1109/IROS.2013.6696578.

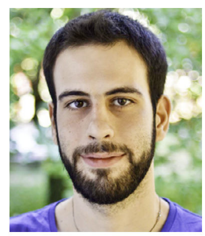
[104] P. Papadakis, P. Rives, A. Spalanzani, Adaptive spacing in human-robot interactions, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 2627–2632. http://dx.doi.org/10.1109/IROS.2014.6942921.

[105] G. Ferrer, A. Garrell, A. Sanfeliu, Robot companion: A social-force based approach with human awareness-navigation in crowded environments, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2013, pp. 1688–1694. http://dx.doi.org/10.1109/IROS.2013.6696576.

[106] G. Ferrer, A. Garrell, A. Sanfeliu, Social-aware robot navigation in urban environments, in: Mobile Robots, ECMR, European Conference on, 2013, pp. 331–336. http://dx.doi.org/10.1109/ECMR.2013.6698863.

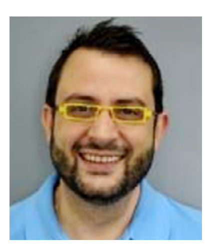
[107] M. Shiomi, F. Zanlungo, K. Hayashi, T. Kanda, Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a

pedestrian model, Int. J. Soc. Robot. (ISSN: 1875-4805) 6 (3) (2014) 443–455. http://dx.doi.org/10.1007/s12369-014-0238-y.
[108] G. Ferrer, A. Sanfeliu, Behavior estimation for a complete framework for human motion prediction in crowded environments, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 5940–5945. http://dx.doi.org/10.1109/ICRA.2014.6907734.
[109] G. Ferrer, A. Sanfeliu, Proactive kinodynamic planning using the extended social force model and human motion prediction in urban environments, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 1730–1735. http://dx.doi.org/10.1109/IROS.2014.6942788.
[110] S.-Y. Chung, H.-P. Huang, A mobile robot that understands pedestrian spatial behaviors, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2010, pp. 5861–5866. http://dx.doi.org/10.1109/IROS.2010.5649718.
[111] S.T. O'Callaghan, S.P.N. Singh, A. Alempijevic, F.T. Ramos, Learning navigational maps by observing human motion patterns, in: Robotics and Automation, ICRA, IEEE International Conference on, 2011, pp. 4333–4340. http://dx.doi.org/10.1109/ICRA.2011.5980478.
[112] M. Luber, L. Spinello, J. Silva, K.O. Arras, Socially-aware robot navigation: A learning approach, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2012, pp. 902–907. http://dx.doi.org/10.1109/IROS.2012.6385716.
[113] M. Luber, K.O. Arras, Multi-hypothesis social grouping and tracking for mobile robots, in: Robotics: Science and Systems, RSS, Berlin, Germany, 2013.
[114] D. Vasquez, B. Okal, K.O. Arras, Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 1341–1346. http://dx.doi.org/10.1109/IROS.2014.6942731.
[115] S. Chung, H. Huang, Predictive navigation by understanding human motion patterns, Int. J. Adv. Robot. Syst. 8 (1) (2011) 52–64.
[116] S. Xiao, Z. Wang, J. Folkesson, Unsupervised robot learning to predict person motion, in: Robotics and Automation, ICRA, IEEE International Conference on, 2015, pp. 691–696. http://dx.doi.org/10.1109/ICRA.2015.7139254.
[117] E.A. Sisbot, L.F. Marin-Urias, X. Broquère, D. Sidobre, R. Alami, Synthesizing robot motions adapted to human presence, Int. J. Soc. Robot. (ISSN: 1875-4805) 2 (3) (2010) 329–343. http://dx.doi.org/10.1007/s12369-010-0059-6.
[118] J.V. Gomez, N. Mavridis, S. Garrido, Fast marching solution for the social path planning problem, in: Robotics and Automation, ICRA, IEEE International Conference on, 2014, pp. 1871–1876. http://dx.doi.org/10.1109/ICRA.2014.6907105.
[119] J.V. Gómez, N. Mavridis, S. Garrido, Social Path Planning: Generic Human-Robot Interaction Framework for Robotic Navigation Tasks, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, Tokio, 2013, pp. 1–10.
[120] A. Escobedo, A. Spalanzani, C. Laugier, Using social cues to estimate possible destinations when driving a robotic wheelchair, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2014, pp. 3299–3304. http://dx.doi.org/10.1109/IROS.2014.6943021.
[121] A. Ball, D. Silvera-Tawil, D. Rye, M. Velonaki, Group comfortability when a robot approaches, in: Social Robotics, ICSR, 6th International Conference, Springer International Publishing, ISBN: 978-3-319-11973-1, 2014, pp. 44–53. http://dx.doi.org/10.1007/978-3-319-11973-1_5.
[122] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, A. Kleiner, On measuring the accuracy of SLAM algorithms, Auton. Robots (ISSN: 1573-7527) 27 (4) (2009) 387–407. http://dx.doi.org/10.1007/s10514-009-9155-6.
[123] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, A benchmark for the evaluation of RGB-D SLAM systems, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2012, pp. 573–580. http://dx.doi.org/10.1109/IROS.2012.6385773.
[124] A. Schmidt, M. Fularz, M. Kraft, A. Kasiński, M. Nowicki, An indoor RGB-D dataset for the evaluation of robot navigation algorithms, in: Advanced Concepts for Intelligent Vision Systems, ACIVS, 15th International Conference, Springer International Publishing, ISBN: 978-3-319-02895-8, 2013, pp. 321–329. http://dx.doi.org/10.1007/978-3-319-02895-8_29.
[125] J.-L. Blanco, F.-A. Moreno, J. González-Jiménez, The Málaga urban dataset: High-rate Stereo and Lidars in a realistic urban scenario, Int. J. Robot. Res. 33 (2) (2014) 207–214. http://dx.doi.org/10.1177/0278364913507326.
[126] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: Computer Vision and Pattern Recognition, CVPR, IEEE Conference on, 2012, pp. 3354–3361. http://dx.doi.org/10.1109/CVPR.2012.6248074.
[127] R. Capobianco, J. Serafin, J. Dichtl, G. Grisetti, L. Iocchi, D. Nardi, A proposal for semantic map representation and evaluation, in: Mobile Robots, ECMR, European Conference on, 2015, pp. 1–6. http://dx.doi.org/10.1109/ECMR.2015.7324198.
[128] L. Takayama, C. Pantofaru, Influences on proxemic behaviors in human-robot interaction, in: Intelligent Robots and Systems, IROS, IEEE/RSJ International Conference on, 2009, pp. 5495–5502. http://dx.doi.org/10.1109/IROS.2009.5354145.
[129] Y. Kim, B. Mutlu, How social distance shapes humanrobot interaction, Int. J. Hum.-Comput. Stud. (ISSN: 1071-5819) 72 (12) (2014) 783–795. http://dx.doi.org/10.1016/j.ijhcs.2014.05.005.
[130] Z. Henkel, C.L. Bethel, R.R. Murphy, V. Srinivasan, Evaluation of proxemic scaling functions for social robotics, IEEE Trans. Hum.-Mach. Syst. (ISSN: 2168-2291) 44 (3) (2014) 374–385. http://dx.doi.org/10.1109/THMS.2014.2304075.
[131] CMU, CMU Graphics Lab Motion Capture Database, 2003. URL http://mocap.cs.cmu.edu/.
[132] B. Majecka, Statistical Models of Pedestrian Behaviour in the Forum (Master's thesis), School of Informatics, University of Edinburgh, 2009.
[133] K. Konolige, J. Bowman, Towards lifelong visual maps, in: 2009 International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 1156–1163. http://dx.doi.org/10.1109/IROS.2009.5354121.
[134] A. Hornung, K.M. Wurm, M. Bennewitz, C. Stachniss, W. Burgard, OctoMap: An efficient probabilistic 3D mapping framework based on octrees, Auton. Robots 34 (3) (2013) 189–206. http://dx.doi.org/10.1007/s10514-012-9321-0.



Konstantinos Charalampous was born in Thessaloniki, Greece, in 1986. He received the Bachelor's and M.Sc. (with Honors) degrees both from the School of Informatics, Aristotle University of Thessaloniki, Greece, in 2009 and 2011, respectively. He is currently a Ph.D. candidate with the Laboratory of Robotics and Automation, in the Department of Production and Management Engineering, Democritus University of Thrace. He has also participated in research projects funded by the European Commission and the Greek state. His areas of interest include machine learning and human–robot interaction.

Dr. Ioannis Kostavelis is a Postdoctoral Research Associate at the Democritus University of Thrace, Department 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 and a M.Sc. in Informatics from the Aristotle University of Thessaloniki. He fulfilled his Ph.D. studies under the hood of the Laboratory of Robotics and Automation, at the Department of Production and Management Engineering, Democritus University of Thrace, under the supervision of Prof. Antonios Gasteratos. His research has been supported through 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/.



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 (LIRALab), DIST, University of Genoa, Italy. He has served as a reviewer for numerous Scientific Journals and International Conferences. His research interests are mainly in mechatronics and 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/.