Binding human spatial interactions with mapping for enhanced mobility in dynamic environments

Panagiotis Papadakis, Patrick Rives

Binding Human Spatial Interactions with Mapping for Enhanced Mobility in Dynamic Environments

Panagiotis Papadakis · Patrick Rives

Abstract For mobile robots to operate in compliance with human presence, interpreting the impact of human activities and responding constructively is a challenging goal. In this paper, we propose a generative approach for enhancing robot mapping and mobility in the presence of humans through a joint, probabilistic treatment of static and dynamic characteristics of indoor environments. Human spatial activity is explicitly exploited for the purpose of passage detection and space occupancy prediction while effectively discarding false positive human detections using prior map information. In turn, this allows the execution of plan trajectories within unexplored areas by using human presence for resolving the uncertainty or ambiguity that is due to dynamic events. A series of experiments with an indoor robot navigating in close human proximity within a multi-floor building demonstrate the effectiveness of our approach in realistic conditions.

Keywords socially-compliant robots · human spatial interaction · augmented mapping · human action · non-stationary pdf

1 Introduction

Mapping an environment allows robots to be deployed in diverse workspaces, marking this skill as a primary element in the integration of robots into human-populated environments. The merits of this integration are largely expected through the development of human-compliant robot skills that would allow robots to exploit and respond constructively to human activity rather than proactively avoiding the interruption of human activities.

team LAGADIC, INRIA Méditerranée
Sophia-Antipolis, France
E-mail: {panagiotis.papadakis, patrick.rives}@inria.fr

This is a very challenging problem, as it requires the use of raw sensor readings in conjunction with implicit information in order to interpret the effects of human interaction into the process of robot mapping as well as navigation.

In the domain of assistive robotics in particular, human-robot interaction (HRI) is essential for the acceptance of the respective technology by the target user group. Regardless of the form that HRI can take on in such settings (i.e. explicit or implicit interaction), human presence and motion need to be explicitly modelled in the development of various robot skills that range from environment mapping to spatially situated communication in the form of socially compliant path-planning or high-level dialogue.

Typically, the problem of environment mapping is treated through a decomposition into layers that each represents a desired abstraction level, ranging from the low-level 2D/3D metric and topological [6] to the semantic [25], object-based and social activity level [24] (we refer the interested reader to [27] and [17] for the most recent survey articles). The latter layer (i.e. social mapping) has a strictly local spatio-temporal character in contrast to all the remaining that focus on the static, long-term properties of the observed environment [14]. Ideally, however, the different mapping layers should be able to interact and complement each other during robot operation in the event of insufficient/corrupted sensing in a particular layer, likewise in the case of abundant information within another. In fact, we should expect such imbalance to increase as robots enter human-populated environments wherein increased frequency of dynamic events results in reduced spatial-temporal consistency.

Indicatively, consider the scenario shown in Fig. 1 where conventional mapping would fail to resolve am-
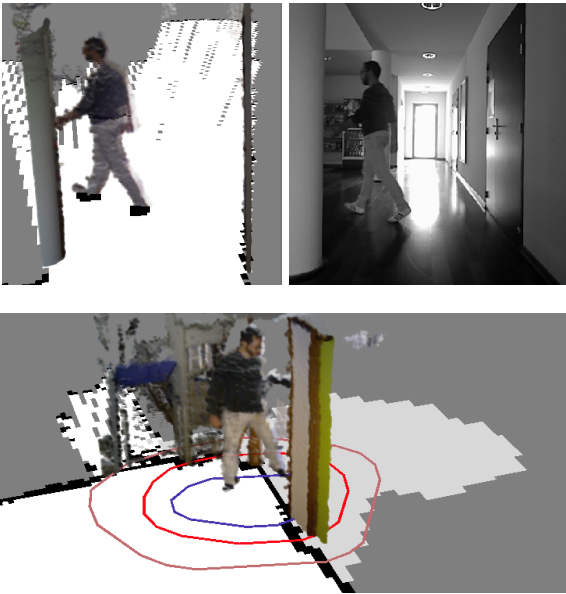
Fig. 1 Human presence or actions such as door opening may interfere in robot operation if they are not properly taken into account during mapping and navigation.

biguities that could typically arise during robot operation. A human or a door may be treated as obstacles and inserted into the metric map while on the contrary, both imply the presence of unoccupied space. Instead of discarding such conflicting information, a robot could employ a probabilistic interpretation of such events making its operation more compliant to human activity.

By explicitly accounting for human presence and motion in combination to static characteristics, the work that we present in this paper advances the domain of robotics with contributions on the level of perception and navigation in dynamic environments. Our work is part of the broader initiative referred to as **PAL** (Personally Assisted Living) project whose aim is to develop technologies and services for improving the quality of life of the elderly and fragile people [1]. Towards these objectives, we propose a mapping framework that considers human spatial interaction cues combined with raw mapping measurements, in order to augment map information and in turn, extend the horizon of feasible path-plans. On this basis, the presence of humans not only contributes to the process of mapping but it further allows path planning within areas that would otherwise be considered as inaccessible. This is achieved through a probabilistic integration of our recent work on social-mapping [24, 23] with conventional occupancy mapping in order to: (i) infer the presence of passages such as doors and staircases, (ii) filter the process of human tracking by discarding false positives and (iii) predict the presence of unoccupied space. We evaluate our approach on an indoor robot operating in the presence

of humans and show the qualitative and quantitative advantage of our hybrid scheme, in realistic scenarios.

We organize the remainder of the paper as follows. First, in Section 2, we provide a concise literature review in the domain of mapping using human input, in Section 3 we unfold the details of our contribution and finally we evaluate our approach and summarize our findings in Sections 4 and 5 respectively.

2 Background

Across the broad range of robotic applications, we are here concerned about mobile robots designed to operate within indoor environments that can build consistent representations of the real-world either implicitly through the extraction of human action semantics [36] or explicitly though intuitive HRI. These skills can be subsequently exploited in diverse modalities from robot companions to automated wheelchairs. Under this perspective, while notable progress could be acknowledged for methods that map separate environment layers, interactive mapping schemes where the human is explicitly taken into account are starting to emerge relatively recently and with constrained application scope.

In a baseline approach, Zender et al. [34] presented a robotic system equipped with multi-modal perception modalities, using auditory human feedback to classify areas and objects in a supervised manner. The same principle wherein explicit human feedback is given to the robot is also present within several other works [35, 3, 7, 18, 30, 16], however, their strong dependence on human supervision constrains their operation alongside trained users and simplified scenarios.

Within more generic settings, Grzonka et al. [12] demonstrated the ability to build approximate environment maps by using door closure/opening events as candidate loop closure positions and room delimiters. A conditioned dilation operation on the 2D image of a human path was employed to emerge the coarse structure of the environment. Similarly, Li et al. [20] cast the problem of semantic mapping as the problem of human activity recognition, using wearable sensors to recognize human pose sequences that have been previously associated to furniture classes and annotate the 2D metric map with the furniture label that corresponds to the detected activity. Sundaram et al. [29] present a wearable human activity recognition system where scene motion is first subtracted, bringing forth spatio-temporal activity blocks that are described using a bag-of-features approach. Thereafter, recognized activity classes are associated to the corresponding visual landmarks within a sparse group of sub-maps. While all the aforementioned works propose mapping schemes by interpreting

human interaction cues, their practical utility is limited as they are all based on wearable sensors, while only a single human can be taken into account.

The recent work from Wong et al. [32] formalizes the probabilistic dependencies between metric measurements and static object poses, as a means to predict object presence attributes from space occupancy and vice-versa. By showing the added value in binding metric mapping with static object recognition in constrained examples, that work is incentive to our contribution that advances towards linking human action cues with metric mapping in realistic conditions.

Other relevant approaches are focused on exploiting observed human actions as a means for developing socially-intelligent path planning behaviours. Using Gaussian process-based regression, [21] proposed the use of *navigational maps* to encode human tendencies when navigating towards a goal, while [2] refined the probability of space occupancy using the human pose traces as training points for fitting the likelihood function. On the same track, the work of Wada et al. [31] extracted distinctive walking patterns across different areas by aggregating leg traces while they recently integrated human articulation cues [22] as a means to map the dependence of human motion classes in specific areas. All those works rely on off-line data collection through human observations to build constant map representations, while our higher objective is to develop human interaction-based mapping schemes that improve the local spatio-temporal consistency. This objective is pursued in recent works [24, 23] where the concept of *social-mapping* and *adaptive spacing* were introduced. Spatially-situated human interactions are taken into account in real-time on a probabilistic basis, allowing the analysis of various features for the purpose of socially-compliant navigation.

Finally, it is worth referring to the equivalent research stream in the community of computer vision, i.e. 3D scene understanding by human action observation as outlined in the recent work of Fouhey et al. [9]. Despite sharing the same high-level goal, the problem becomes much more challenging in robotics due to constrained resources thus inhibiting the application of computationally demanding algorithms.

In the present work, we advance the state-of-the-art by upgrading social-mapping and linking it with semantic and occupancy mapping. This enhances the situation awareness of a robot in the presence of humans and augments its state space by resolving perceptual ambiguities and predicting the existence of free space within unexplored areas. Notably, our approach is evaluated on a mobile robot during its operation that interprets human spatial interaction cues without the need for dedicated human-robot interfaces.

3 Methodology

3.1 Atomic space function

We begin our description by briefly recalling the notion of *atomic space* [23] that we will subsequently exploit in the integration of semantic and occupancy mapping.

Each detected human (see Section 4 for the human detection method adopted in this work) is appointed with a parametrized atomic space function $IS_{\mathbf{p}}(.)$ that conveys a selected group of social behaviour cues, that are common in spatially-situated communication. In detail, we consider: (i) the $2D$ position $\mathbf{t} = (t_x, t_y)^T \in \mathbb{R}^2$, (ii) orientation $\theta \in [0, 2\pi)$ and (iii) dominant side $d \in \{-1, +1\}$ of the human where $-1, +1$ correspond to the left-hand and right-hand side respectively. These cues are collected from a module that performs human detection and tracking of the articulated human motion (see subsection 3.3). Human body position and orientation are both obtained from the "torso" skeletal joint. The dominant side is inferred as the minimum lateral distance of the human from occupied space from the 2D metric map, based on the hypothesis that people maintain a smaller distance from their dominant side and vice versa [13]. In this way, a person's parameter vector is denoted as $\mathbf{p} = (\mathbf{t}^T, \theta, d)^T$.

To express these characteristics within a representative function we denote the atomic space function as $IS_{\mathbf{p}}(.)$ and appoint a bivariate Skew-Normal (SN) probability function [4,5] denoted as $\tilde{\mathcal{N}}$ (see examples in Fig. 2). The motivation in choosing the SN probability is attributed to the fact that it fulfills all conditions as expressed in [23] (cf. Section II.A), its notational simplicity as well as because it generalizes all previous partial models that were using the normal distribution. $IS_{\mathbf{p}}(.)$ is a probability density function distributed as $\tilde{\mathcal{N}}(\mathbf{t}, \Omega, \alpha)$ defined as:

$$IS_{\mathbf{p}}(\mathbf{u}) = 2\phi(\mathbf{u})\Phi(\alpha^T \mathbf{u}) \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^2$, $\phi(.)$ denotes the normal probability density function $\mathcal{N}(\mathbf{0}, \Omega)$ with covariance matrix Ω,
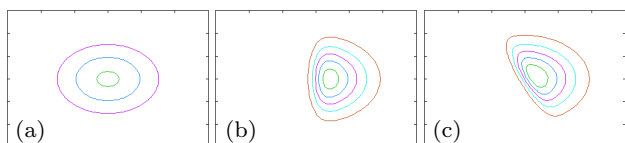
Fig. 2 Instances of SN distributions; (a) Normal, (b) lateral skewness, (c) lateral & vertical skewness.

$\Phi(.)$ is the cumulative distribution function of ϕ and $\alpha = (\alpha_1, \alpha_2)^T$ is the parameter vector that controls the skewness. Without loss of generality, we take the position of the human $\mathbf{t}$ to be always centred at the coordinates' origin $\mathbf{0}$ and with body direction $\theta = \pi/2$ aligned to the vertical axis. Finally, a scaling transformation is applied so that *proxemics*-based distances [15] are assigned to the iso-contours of the function.

According to proxemics theory [15], humans are implicitly circumscribed by zones of varying radius that convey the level of social sensitivity that is aroused when another person is located within the corresponding area. These zones are referenced as the **intimate**, **personal**, **social** and **public** zones that we initially set for the base-line model of concentric circles (see Fig. 2 (a)) in order to correspond to the experimentally established distances. The iso-contours are then automatically adjusted for all remaining models (cf. [19], Fig. 9) following the variation of the parameter vector α.

3.2 Passage Detection

Understanding the functional characteristics of certain environment structures is useful for a robot especially regarding its mobility within a domestic environment. Here, we are particularly interested for the detection of passages in the form of doors and staircases following an approach that is entirely invariant to their static visual or geometric appearance and instead depends only on the interaction of a human with the environment. Our motivation is that the advance of robots integration in human-populated environments in only feasible by improving the level of human-activity understanding because static environment data become less informative and consistent as a consequence of human presence in the foreground of the robot.

The detection of a passage via human observation is modelled as the joint probability of three statistically independent events A_i, $i = 1, 2, 3$ defined in the sequel.

Upon the detection of a human, event A_1 is defined as the binary random variable that determines whether the human currently, at time t, resides on unknown space as derived from his position within the 2D map. The discrete probability of A_1 is formally defined as:

$$P(A_1) = \begin{cases} 1, & \forall \mathbf{u}(IS_{\mathbf{p},t}(\mathbf{u}) > p_{th} \wedge M(\mathbf{u}) = -1) \\ 0, & otherwise \end{cases} \quad (2)$$

where $M(.) : \mathbb{R}^2 \to \{-1, 0, 1\}$ is the function which gives the occupancy value of the corresponding cell within the 2D metric map and $\{-1, 0, 1\}$ correspond to *unknown*, *occupied* and *free* space respectively (see Section
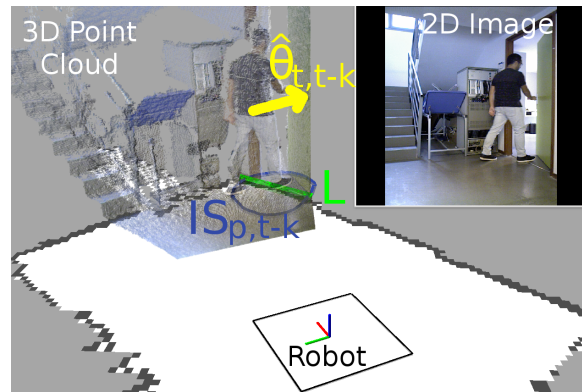
Fig. 3 Door detection example with annotated variables. The iso-contour $IS_{\mathbf{p},t-k}(\mathbf{u}) = p_{th}$ is depicted in blue, the 2D fitted line L in green and orientation $\hat{\theta}_{t,t-k}$ in yellow.

4 for the mapping method employed in this work). The p_{th} value is a constant threshold and controls the total area under his atomic space function that is considered.

Upon successful detection of a human who is currently residing in unknown space, we make a backward, temporal assessment of whether the same human passed over occupied space during the previous short-term path at time $t - k$. We define event A_2 as the corresponding binary random variable and define:

$$P(A_2) = \begin{cases} 1, & \exists \mathbf{u}(IS_{\mathbf{p},t-k}(\mathbf{u}) > p_{th} \wedge M(\mathbf{u}) = 0) \\ 0, & otherwise \end{cases} \quad (3)$$

where k defines the backward temporal range where we check the human's previous position. This parameter is set to a value which ensures that whenever a human passes from occupied space to unknown space then this event will not be missed (in practice, we set $k = 1.5$ sec for all humans assuming a moderate walking speed when traversing a passage, or alternatively, obtain k as a function of the specific speed of the respective human at time t).

If $P(A_2) = 1$, then we proceed by fitting the 2D line L that is defined by the occupied cells that lie within the area around the human's atomic space function, i.e. cells $U_{pass} = \{u \in M^{-1}(\mathbf{u}) = 0 | IS_{\mathbf{p},t-k}(\mathbf{u}) > p_{th}\}$, in order to determine the dominant orientation of the passage (door or staircase). This is required for defining event A_3 as the binary random variable that determines whether the direction of human motion within the time interval k and the orientation of the candidate passage are mostly orthogonal to one another. This is based on the observation that humans conventionally tend to go through passages while facing the passage which is a condition that is useful in discarding false positives that could appear if only events A_1 and A_2 would be taken into account. Therefore, we define $P(A_3)$ as:

$$P(A_3) = \begin{cases} 1, & L \perp \hat{\theta}_{t,t-k} \\ 0, & otherwise \end{cases} \quad (4)$$

where $\hat{\theta}_{t,t-k}$ is the average body direction within the time interval k. The final probability for the detection of a passage is then taken as:

$$P(D) = \prod_{i=1}^{3} P(A_i), D \in \{0, 1\} \quad (5)$$

In Fig. 3 we provide an instructive example of passage detection for reference of the used notations.

At this point, it is worth elaborating on certain aspects of the proposed passage detection process. The first aspect concerns the case of doors which our approach will detect in the case where they were initially closed during 2D metric mapping and subsequently opened and traversed by a human. Our approach serves this objective and resolves the ambiguity which arises when a robot initially treats the closed door as occupied space while after door opening and human traversal becomes free to access (example shown in Fig. 4), by explicitly setting the value of the corresponding 2D cells to 0 (i.e. *free*).

In turn, this implies that we are not specifically concerned at treating doors that are already opened during mapping as those do not pose mobility problems for a robot neither do they create ambiguities in the process of mapping. As mentioned in the beginning of the article, our purpose is to analyse the interaction of the human within the environment in order to derive consistent representations of the real-world that is otherwise infeasible by solely treating static sensory data. For scenarios where static sensory data of doors are not cluttered-obstructed due to human presence, the interested reader could refer to our relevant recent work in [8], where door detection is performed through feature extraction on RGB-D (colour and depth) data.

Finally, in order to discriminate between door passages and staircase passages, we measure the height difference of the human (as given by the *torso* skeletal
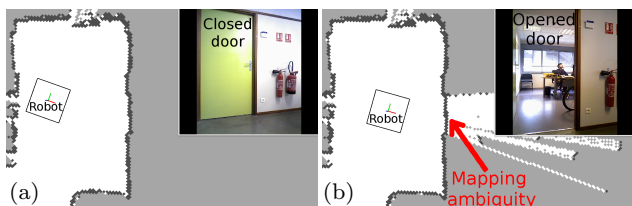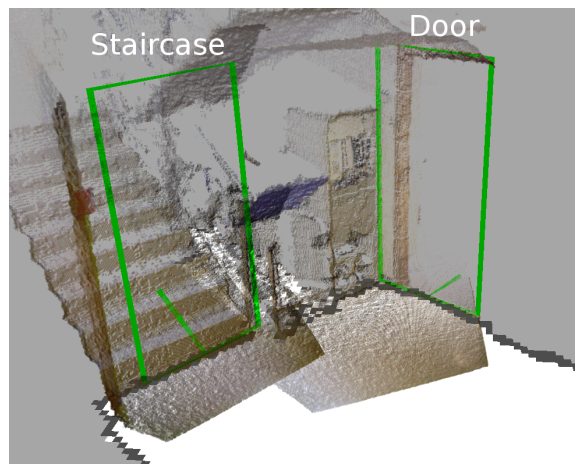
Fig. 5 Staircase and door detection example. The location/orientation of the passages is denoted through the respective green rectangular frames.
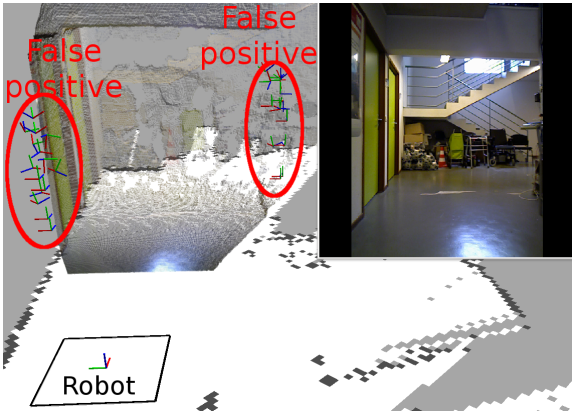
joint) between the two poses at times $t-k$ and t, which is also used to derive the staircase inclination (the occurring height difference when the human traverses a staircase is far more pronounced than the height deviation due to noise). Fig. 5 shows an example of staircase and door detection that are located in close proximity.

3.3 Human Detection Filtering

For robots that operate alongside with humans, detection and tracking of their articulated motion is a primary objective. In addressing this goal, there exist multiple challenges that arise either due to the environment, the co-located humans or the robot itself.

The work of Shotton et al. [28] which forms the basis of the well-known human detection and tracking module used in RGB-D sensors such as Kinect or AsusXtion, represents the state-of-the-art approach in marker-less tracking of the articulated human motion. Unfortunately, its use in robotic applications is not as straightforward due to the fact that the underlying algorithm is based on the assumption that the sensor is fixed on a particular position and the background environment remains static.

In practice, when this functionality is tested on a mobile robot it results in the detection of multiple *phantom* human detection candidates which do not correspond to real humans and are essentially false positives (example shown in Fig. 6). Due to the motion of the robot, static structures of the environment that may resemble certain human poses (such as chairs, parts of the wall, etc) are mistaken for real humans which in turn has a severe negative impact on any subsequent analysis of human behaviour, such as the detection of passages

Fig. 4 Ambiguity in space occupancy which arises as a result of door opening: (a) Occupancy map while door is closed, (b) inconsistent map update after door opening.

Fig. 6 False positive detections of humans as a consequence of robot motion when using [28].

(Section 3.2), space occupancy prediction (Section 3.4) and of course, in socially compliant path-planning.

To address this problem, we introduce a probabilistic filter $Fil(\mathbf{p})$ for each detected person $\mathbf{p}$ that operates on the coordinate frame of the global 2D occupancy map and that allows us to effectively discard these false positives, by defining:

$$Fil(\mathbf{p}) = P(A_2|k = 0) \wedge (\|\mathbf{v_p}\| \simeq 0) \qquad (6)$$

where $\mathbf{v_p}$ is the current velocity of the human.

In essence, eq. (6) states that a human detection is discarded ($Fil(\mathbf{p}) = 1$) whenever the atomic space function currently contains occupied cells and there is no human motion. Otherwise, $Fil(\mathbf{p}) = 0$ and the respective human detection candidate is retained as a true positive. Note that we make use of the event A_2 as defined in the previous section, but since we do not need to look into the past, we set $k = 0$ in order to evaluate the respective probability at the present time.

Through the application of this filter, the robot can discard the false positives that arise due to its motion since those are attributed onto structures of the static environment whose corresponding 2D cells have normally been designated as occupied within the 2D metric map and in turn exhibit no motion.

3.4 Space Occupancy Prediction

So far we have exploited the interaction of the human within the 2D occupancy map in order to detect passages as well as filter the process of human detection and tracking. In this section, we use human motion as a cue for the prediction of free space in the occupancy map within areas that are beyond the spatio-temporal field of view of the 2D mapping process.

To introduce the effect of human presence into the process of 2D space occupancy mapping, our starting point will be the space occupancy posterior introduced by [32]. In their work, they jointly take into account 2D laser and static object pose measurements by:

$$\tilde{M}(\mathbf{u}) = M(\mathbf{u}) \cdot (1 - p_{overlap}) \qquad (7)$$

where $p_{overlap}$ is the probability that the location of a detected object overlaps with position $\mathbf{u}$ and $\tilde{M}(.)$ is the updated occupancy value after taking into account object presence within the 2D map.

To extend this formulation to account for human presence we introduce the time dimension to account for human motion and for the purpose of predicting free space, rather than occupied space. More particularly, our extension is based on the argument that every space that is traversable by the human should be equivalently accessible by the robot. For example, if a human is observed to turn around a corner the robot should infer the presence of free space at that direction, even though the space beyond the corner is not currently visible. Similar conditions may appear when a human enters a room whose interior has not yet been mapped by the robot or for a corridor.

Let us define as $\tilde{M}(.)$ the occupancy value of a cell after considering the dynamic human presence in the environment. We can then obtain the updated occupancy value of a cell at position $\mathbf{u}$ by:

$$\tilde{M}(\mathbf{u}) = P_{traverse} + M(\mathbf{u}) \cdot (1 - P_{traverse}) \qquad (8)$$

where $P_{traverse}$ is the discrete probability that the human has traversed the respective area but is not currently at time t residing in the corresponding area, i.e.:

$$P_{traverse} = (IS_{\mathbf{p},\bar{t}}(\mathbf{u}) > p_{th}) \wedge (IS_{\mathbf{p},t}(\mathbf{u}) < p_{th}) \qquad (9)$$

where $\bar{t}$ is the residual time period after excluding the current time t.

The time window along which human traversal is considered is always centralized at the current time t and extends backward and forward in time. In this way, we can account for the entire human motion history and a future, short-term displacement in the case where the human is moving. The future, short-term prediction is obtained linearly and applied on the human's parameter vector $\mathbf{p}_{t+\Delta t} = \mathbf{p}_t + \dot{\mathbf{p}}_t \Delta t$ and in turn, on the attributed atomic social space function $IS_{\mathbf{p},t+\Delta t}(.)$. Qualitative examples of space occupancy prediction are presented in the following section.
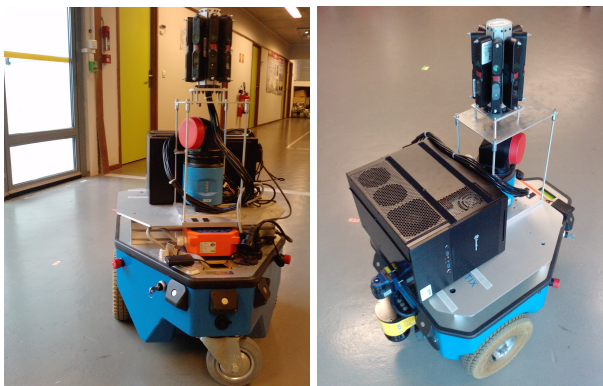
Fig. 7 View of the robotic platform that was used for the indoor experiments.

4 Experiments

In this section we provide a thorough evaluation of the proposed methods, staring by the description of the scenarios and followed by the attained performance.

We have performed our experiments using an indoor robot Neobotix MP-500 which employs a two-wheel differential drive mobility system (see Fig. 7), using the Robot Operating System (ROS) [26]. The robot is equipped with 2D SICK laser S300 that is used by the 2D mapping process [11] for building a global 2D metric map of the environment and by path-planning [10] where the 2D map constitutes the state-space of the robot.

Robot navigation was performed in a semi-supervised mode with human and passage detection entirely unsupervised. To detect the articulated human motion we used the Asus XtionPRO Live camera which captures synchronized RGB-D images in conjunction with the OpenNI (Open Natural Interface) framework [28]. An array of 8 cameras in total are mounted sideways on top the robotic platform, by regular 45° increments that allow a panoramic sensing of the environment.

The experiments were performed in the interior of 3 floor building at the facilities of INRIA Sophia-Antipolis in France. The first 2 floors are comprised of 17 office rooms (and 17 corresponding door passages) in total

Table 1 Basic characteristics of experiments

Task type	Mapping in human proximity
Robot morphology	Functional
Human roles	Bystander
Physical proximity	Passing
Time	Synchronous
Space	Collocated
Human-Robot ratio	1
Interaction	Robot & Human
Autonomy/Intervention	75%/0%

and are connected via 2 distinct staircases. The third floor contains a reception hall, 3 rooms (and 3 corresponding door passages) and 1 staircase while all floors contain rest-rooms. Here, we note once again that any specific visual or geometric characteristics of that or any other environment, is transparent to the application of our framework which only depends on the effects of human interaction with the corresponding map. On this basis, the application of the proposed methodologies in an assistive setting is straightforward provided that a robot is capable of performing 2D SLAM and human detection and tracking. Table 1 provides a description of the experiments on the basis of the HRI Taxonomy [33]. Note that even though the robot and the human do not explicitly interact, the fact that the human and the robot are collocated denotes an implicit interaction. The robot purposefully analyses human activity and occasionally follows the traced human paths.

Among the total number of passages (doors and staircases) present in the environment the robot encountered the traversal of all 3 different staircases and 9 different doors, performed by 3 distinct humans at random order. Human robot encounters were all deliberately staged for the purpose of the experiments. On the other hand, participants were free to follow their casual way of walking or traversing a door/staircase. We deemed this as a reasonable balance between collecting a sufficient volume of experimental data and testing the proposed methods with a realistic variance among individuals.

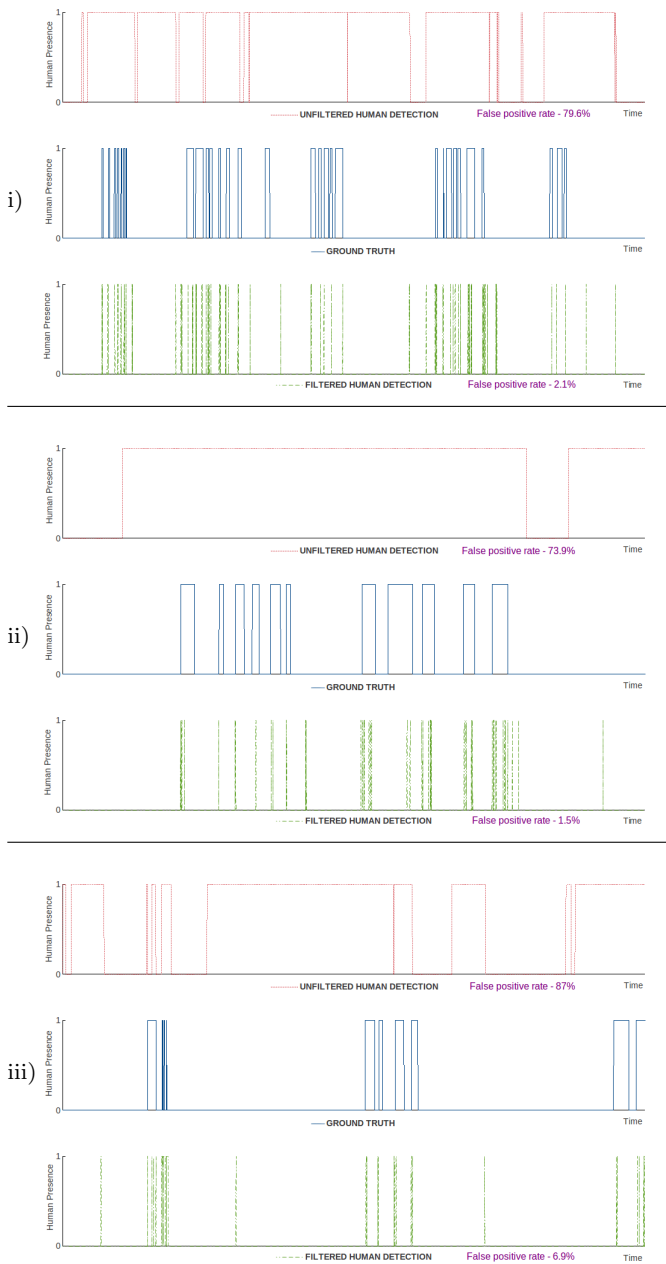### 4.1 False positives filtering results

Filtering the output of human detection and tracking is an essential element within the proposed framework, since human-like shapes may be misinterpreted as real humans and mislead the robot. As explained in section 3.3, the false positives appear mainly due to the motion of the robot and the resulting 3D *flow* of the environment while impoverished by the perceptual limits and noise in the RGB-D sensor.

Fig. 8 provides the evaluation of the proposed filtering stage (see section 3.3), decomposed into the three robot operation trials for each of the three building floors respectively. The results demonstrate that relying on the original, unfiltered human detection module is clearly unrealistic since the false positive rate is too high and humans are detected almost all along the experiment trials (see top plot of each trial). On the contrary, the ground truth (middle plot of each trial) shows that real humans are encountered by the robot for only a short overall period of time for each trial compared to

Fig. 8 Performance in filtering false positives of human detections, during three trials (i), (ii) and (iii) of robot exploration corresponding to the three floors of the test building.

the total duration, either bypassing the robot or when passing through a passage.

The bottom plot of each trial shows the detection rate after employing the proposed filtering stage where we can observe that it largely matches the respective ground truth, with minimal false positive rate. On average, the false positive rate before filtering is 80.5% while after filtering it is reduced to 3.2%, a considerable gain that allows us to nearly match the optimal performance.

Table 2 Passage detection performance by human action observation

|	Passage id	Recall
	door #1	100%
	door #2	100%
Floor i	door #3	60%
	staircase #1	66.7%
	staircase #2	50%
Floor ii	door #4	66.7%
	door #5	50%
	door #6	80%
Floor iii	door #7	100%
	door #8	100%
	door #9	100%
	staircase #3	66.7%
Door average		84.1%
Staircase average		61.1%
Overall		78.3%

We can claim that the attained 3.2% rate of false positives is in fact an overestimated score since the human detection module retains, by its design, a short-memory for humans that had been most recently observed, even if they are not currently present. This also implies that false positive human detections have a minimal time duration until the corresponding memory expires, at which time the *phantom* humans get discarded. In contrast, the ground truth information attests human presence on a firm frame-by-frame basis. However, this is useful in setting a lower bound in the performance of our approach while still using the original configuration of the human detection module.

4.2 Passage detection results

In Table 2 we provide the detection score that was attained for each encountered passage and on average in terms of the true positive (TR) rate, namely, the *Recall* score.

These performances should be evaluated on the basis that passage detection using human interaction cues is entirely dependent on the performance of human detection and tracking. Simply said, if a human actually traverses a passage but is not detected by the respective module then it immediately follows that passage detection will also fail. This effect is inevitable and constitutes the only reason for which we could not attain the highest possible passage detection performance. Whenever a human traversed a passage and was consistently detected the robot would always detect the corresponding passage.

This dependence largely explains the reason why staircase detection performance is lower than that of doors. The difference is due to the fact that the detec-
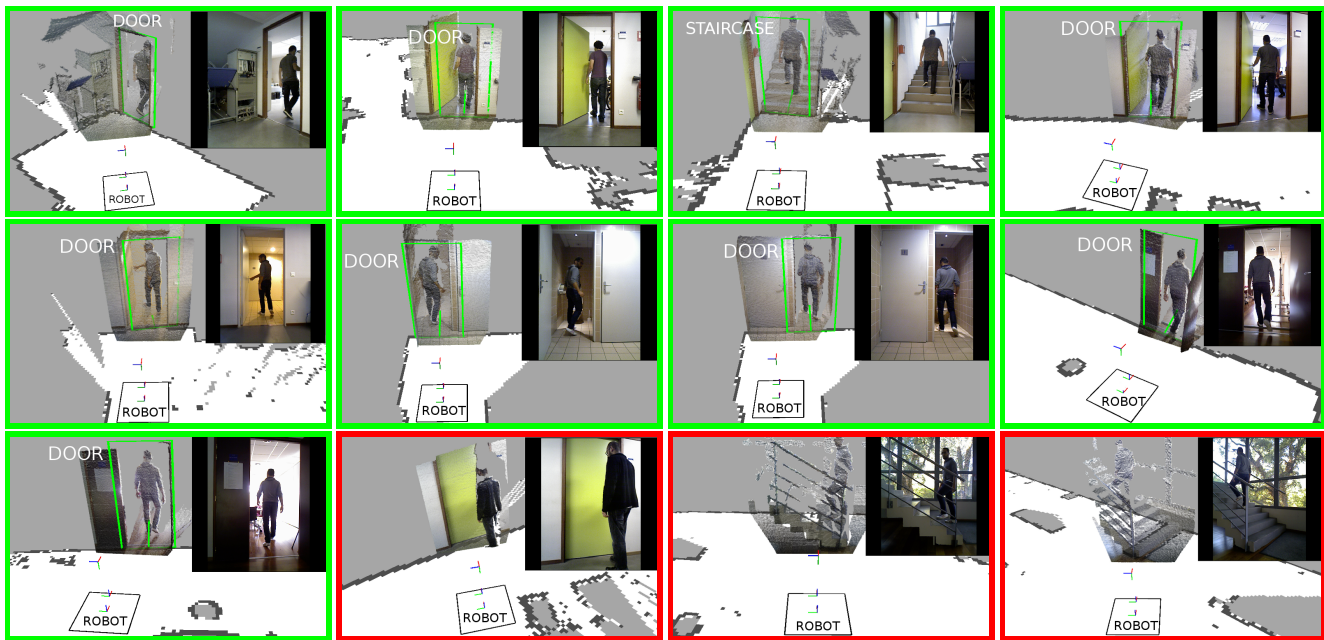
Fig. 9 Qualitative performance in the detection of passages. Examples framed in green denote true positives. False negatives (framed in red) may arise due to failure of the human detection module in cases where the human figure is not clearly visible.

tion of humans traversing staircases was occasionally obstructed by the sidebars of the staircases, while in other cases the ascending human figures were exceeding the narrow field of view of the robot. In certain cases it also happened that the corresponding human traversed a passage too rapidly for the robot to acknowledge the human presence. Irrespective of the reason for failing to detect a human, the results of Table 2 are informative of the level of performance that can be expected in realistic conditions. To elaborate on these results, we provide in Fig. 9 a number of instructive examples from the total duration of the trials, accompanied with a video demonstration through the following link https://www.youtube.com/watch?v=8SSDdBm5sRo.

Interestingly, the difference in the detection performance of doors and staircases through the analysis of human activity appears to be reversed in comparison to the use of purely static photo-geometric characteristics as in [8]. In short, 3D plane estimation is applied for the extraction of walls followed by 2D line and salient point detection within the respective patches for finding door axis and handle. For staircases, 3D plane estimation and grouping of planes based on inclination proved an effective baseline detection approach.

For reference, Table 3 gives the respective average performance that was attained for a representative part of the same building.

By comparing the respective performances, it appears that a hybrid approach that would combine those static and dynamic features would increase the total

Table 3 Passage detection performance by static features

Passage type	Recall
Door average	67%
Staircase average	95%

performance at any condition. Door detection would be mostly reliable by observing humans and conversely for the staircases that are more salient as static structures. Such a synthetic detection approach could be implemented by attributing a certainty score onto the human detection output, e.g. a distance-based function in the same spirit as in [23] (cf. Fig. 3, function $\rho(.)$).

4.3 Space occupancy prediction

Human presence is an explicit indicator of free space along the observed trajectory. This is particularly useful for a robot as it allows the mapping of areas that were previously unknown or resolving ambiguities that arise from dynamic environment aspects (e.g. as in Fig. 4).

In Table 4, we provide the performance in the prediction of free space after applying the posterior rule of eq. (9). We use the *precision* score which is defined as the ratio of number of true positives to the total number of positives (true plus the negative).

The quantitative results strongly indicate that a robot could almost completely rely on the analysis of human motion for resolving uncertainties in mapping

Table 4 Free space prediction performance

| Building floor	Precision
(i)	96%
(ii)	97%
(iii)	98%

space occupancy, namely, free space. Similarly to the case of passage detection, this stage is also directly dependent on the precision of human detection and tracking. On the other hand, false positives in free space prediction do not necessarily pose a threat for a robot as these areas can always be bypassed *on-the-fly* from the local path planner. Since the proportion of the false positive areas is very small compared to true positives, there is a considerable benefit in the navigation of the robot that can exploit the correctly predicted areas in order to visit spaces that could not be previously mapped.

In Fig. 10 we show the qualitative performance of free space prediction for the entire test environment, decomposed into the three distinct floors of the building. The figure is also useful in showing the proportion of the areas where humans were encountered by the robot compared to the total explored environment. It shows the final constructed maps at the end of the trials.

To highlight the immediate positive effect that free space prediction induced for the robot, we present a number of consecutive instances of the robot operation in Fig. 11 together with the corresponding path plan. It is worth noting that using our approach, the robot is capable of finding paths that would normally cross over occupied space within the original 2D map despite the fact that they are truly feasible in the real environment. This is exactly the case in the examples shown in Fig. 11, however the robot bypasses this limitation by inferring the presence of free space by observing the human and extrapolating space occupancy. Consequently, path plans that would not be feasible in the original map $M(u)$ may now be computed and safely executed within the updated map $\tilde{M}(\mathbf{u})$ (see eq. (8)).

5 Conclusions

We have presented an effective approach for using human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages, inferring space occupancy and allowing navigation within unexplored areas. We have adopted a probabilistic treatment which uses elements from the domain of spatially situated human behaviour to address concrete problems that are

(i)

(ii)

(iii)
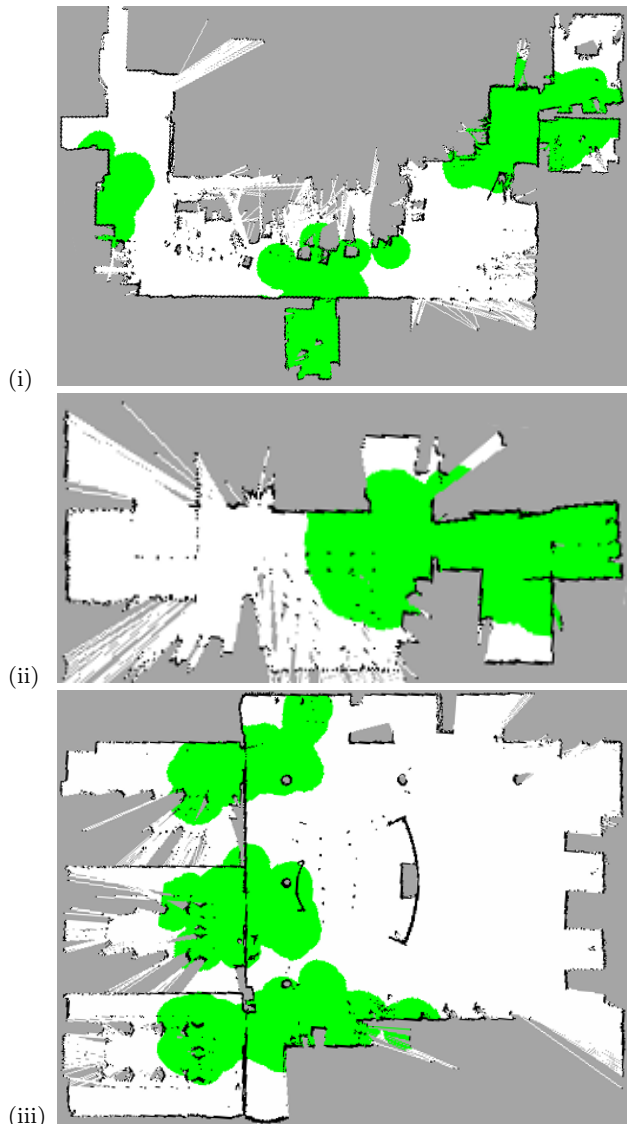
Fig. 10 Free space prediction using human trajectories. Green areas denote the true positives within floors (i)-(iii).

encountered by domestic mobile robots when deployed in dynamic environments alongside humans.

Our contributions are based on the development of intelligent interactions among conceptually different mapping levels, namely, the metric, social and semantic levels. The presented experiments highlighted a number of strong dependences among these levels and the way in which they can be used to enhance individual performances and in turn the global robot operation.

While the immediate application of our work is within the domain of assistive robotics, we further find it pertinent to other areas of robotics where humans and robots share the same spaces. Our ultimate goal is to optimize the benefit that a robot can obtain by treating the human as an information source rather than an ob-
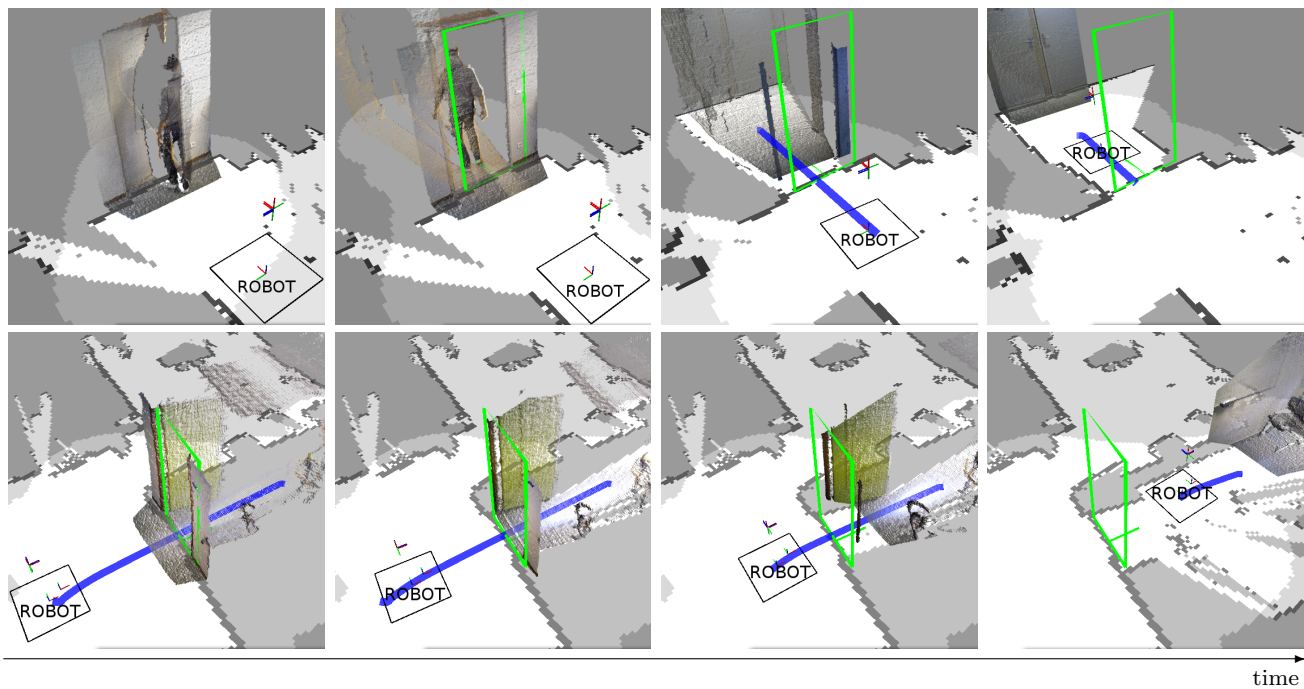
Fig. 11 Path planning and execution within an extended horizon as a result of free space prediction based on humans observation. The computed path is shown as the blue curve.

stacle or sensor interference. Towards this goal, we are interested in exploiting higher-level, multi-party human activity cues and ways to alleviate the increased levels of noise that are susceptible to. In this effort, the earlier work presented in [24] sets the general context of our future developments by formulating human interactions of arbitrary numbers of participants in compliance with sociological studies.

References

1. PAL (Personally Assisted Living) project, https://pal.inria.fr/
2. Alempijevic, A., Fitch, R., Kirchner, N.: Bootstrapping navigation and path planning using human positional traces. In: IEEE Int. Conf. on Robotics and Automation (2013)
3. Althaus, P., Christensen, H.: Automatic map acquisition for navigation in domestic environments. In: IEEE Int. Conf. on Robotics and Automation (2003)
4. Azzalini, A.: A class of distributions which includes the normal ones. Scandinavian Journal of Statistics **12**(2), 171–178 (1985)
5. Azzalini, A., Valle, A.D.: The multivariate skew-normal distribution. Biometrika **83**(4), 715–726 (1996)
6. Chapoulie, A., Rives, P., Filliat, D.: Topological segmentation of indoors/outdoors sequences of spherical views. In: Int. Conf. on Intelligent Robots and Systems (2012)
7. Diosi, A., Taylor, G., Kleeman, L.: Interactive slam using laser and advanced sonar. In: IEEE Int. Conf. on Robotics and Automation (2005)
8. Drouilly, R., Papadakis, P., Rives, P., Morisset, B.: Local map extrapolation in dynamic environments. In: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on (2014)
9. Fouhey, D., Delaitre, V., Gupta, A., Efros, A., Laptev, I., Sivic, J.: People watching: Human actions as a cue for single view geometry. International Journal of Computer Vision **110**(3), 259–274 (2014)
10. Fox, D., Burgard, W., Thrun, S.: The dynamic window approach to collision avoidance. IEEE Robotics Automation Magazine **4**(1), 23–33 (1997)
11. Grisetti, G., Stachniss, C., Burgard, W.: Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE Transactions on Robotics **23**(1), 34–46 (2007)
12. Grzonka, S., Dijoux, F., Karwath, A., Burgard, W.: Mapping indoor environments based on human activity. In: Int. Conf. on Robotics and Automation (2010)
13. Grin-Lajoie, M., Richards, C.L., Fung, J., McFadyen, B.J.: Characteristics of personal space during obstacle circumvention in physical and virtual environments. Gait and Posture **27**(2), 239 – 247 (2008)
14. Hahnel, D., Triebel, R., Burgard, W., Thrun, S.: Map building with mobile robots in dynamic environments. In: Int. Conf. on Robotics and Automation (2003)
15. Hall, E.: The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd (1966)
16. Kim, S., Cheong, H., Park, J.H., Park, S.K.: Human augmented mapping for indoor environments using a stereo camera. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2009)
17. Kostavelis, I., Gasteratos, A.: Semantic mapping for mobile robotics tasks: A survey. Robotics and Autonomous Systems **66**, 86 – 103 (2015)

18. Kruijff, G.J.M., Zender, H., Jensfelt, P., Christensen, H.I.: Clarification dialogues in human-augmented mapping. In: ACM SIGCHI/SIGART Conference on Human-robot Interaction (2006)

19. Kruse, T., Pandey, A., Alami, R., Kirsch, A.: Human-aware robot navigation: A survey. Robotics and Autonomous Systems **61**(512), 1726–1743 (2013)

20. Li, G., Zhu, C., Du, J., Cheng, Q., Sheng, W., Chen, H.: Robot semantic mapping through wearable sensor-based human activity recognition. In: IEEE Int. Conf. on Robotics and Automation, pp. 5228–5233 (2012)

21. O'Callaghan, S., Singh, S., Alempijevic, A., Ramos, F.: Learning navigational maps by observing human motion patterns. In: IEEE Int. Conf. on Robotics and Automation (2011)

22. Ogawa, Y., Wang, Z., Wada, T., Hirata, Y., Kosuge, K.: Generating human motion transition map in indoor environment and analyzing human behavior by geographical clustering. In: IEEE Int. Conf. on Robotics and Automation (2014)

23. Papadakis, P., Rives, P., Spalanzani, A.: Adaptive spacing in human-robot interactions. In: IEEE Int. Conf. on Intelligent Robots and Systems (2014)

24. Papadakis, P., Spalanzani, A., Laugier, C.: Social mapping of human-populated environments by implicit function learning. In: IEEE Int. Conf. on Intelligent Robots and Systems (2013)

25. Pronobis, A., Jensfelt, P.: Large-scale semantic mapping and reasoning with heterogeneous modalities. In: Int. Conf. on Robotics and Automation (2012)

26. Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.: Ros: an open-source robot operating system. In: ICRA Workshop on Open Source Software (2009)

27. Rios-Martinez, J., Spalanzani, A., Laugier, C.: From proxemics theory to socially-aware navigation: A survey. International Journal of Social Robotics **7**(2), 137–153 (2015). DOI 10.1007/s12369-014-0251-1

28. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition (2011)

29. Sundaram, S., Mayol-Cuevas, W.: What are we doing here? egocentric activity recognition on the move for contextual mapping. In: IEEE Int. Conf. on Robotics and Automation (2012)

30. Topp, E., Christensen, H.: Topological modelling for human augmented mapping. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2006)

31. Wada, T., Wang, Z., Ogawa, Y., Hirata, Y., Kosuge, K.: Incremental human motion map system and human walking behavior representation in indoor environment. In: IEEE Int. Conf. on Robotics and Biomimetics (2012)

32. Wong, L.L., Kaelbling, L.P., Lozano-Pérez, T.: Not seeing is also believing: Combining object and metric spatial information. In: IEEE Int. Conf. on Robotics and Automation (2014)

33. Yanco, H., Drury, J.: Classifying human-robot interaction: an updated taxonomy. In: Int. Conf. on Systems, Man and Cybernetics (2004)

34. Zender, H., Mozos, O.M., Jensfelt, P., Kruijff, G.J., Burgard, W.: Conceptual spatial representations for indoor mobile robots. Robotics and Autonomous Systems **56**(6), 493 – 502 (2008)

35. Zhao, C., Mei, W., Pan, W.: Building a grid-semantic map for the navigation of service robots through human-robot interaction. Digital Communications and Networks (2015)

36. Ziaeefard, M., Bergevin, R.: Semantic human activity recognition: A literature review. Pattern Recognition **48**(8), 2329 – 2345 (2015)