



Contents lists available at ScienceDirect

Expert Systems with Applications

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

Simultaneous people tracking and motion pattern learning



Sarath Kodagoda*, Stephan Sehestedt

Centre for Autonomous Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia

ARTICLE INFO

Article history:

Available online 29 May 2014

Keywords:

People tracking
Motion pattern learning
Human Robot Interaction

ABSTRACT

The field of Human Robot Interaction (HRI) encompasses many difficult challenges as robots need a better understanding of human actions. Human detection and tracking play a major role in such scenarios. One of the main challenges is to track them with long term occlusions due to agile nature of human navigation. However, in general humans do not make random movements. They tend to follow common motion patterns depending on their intentions and environmental/physical constraints. Therefore, knowledge of such common motion patterns could allow a robotic device to robustly track people even with long term occlusions. On the other hand, once a robust tracking is achieved, they can be used to enhance common motion pattern models allowing robots to adapt to new motion patterns that could appear in the environment. Therefore, this paper proposes to learn human motion patterns based on Sampled Hidden Markov Model (SHMM) and simultaneously track people using a particle filter tracker. The proposed simultaneous people tracking and human motion pattern learning has not only improved the tracking robustness compared to more conservative approaches, it has also proven robustness to prolonged occlusions and maintaining identity. Furthermore, the integration of people tracking and on-line SHMM learning have led to improved learning performance. These claims are supported by real world experiments carried out on a robot with suite of sensors including a laser range finder.

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Successful Human Robot Interaction (HRI) requires a robot to have advanced abilities to carry out complex tasks. One such ability is robust people tracking. It has been identified as an important tool in HRI not only for safe operation (Schulz, Burgard, Fox, & Cremers, 2003) but also for collision avoidance (Bennewitz, Burgard, Cielniak, & Thrun, 2005) or to implement following behaviors (Bolić & Fernández-Caballero, 2011; Gockley, Forlizzi, & Simmons, 2007; Prassler, Bank, & Kluge, 2002). Conventional way of people tracking is to use known motion models (Kluge, Kohler, & Prassler, 2001; Montemerlo, Thrun, & Whittaker, 1999) including probabilistic motion models (Tadokoro, Hayashi, Manabe, Nakami, & Takamori, 1995; Zhu, 1991). However, those methods have limited ability to model agile human movements. Therefore, some researchers opt not to use motion models (Francesc Serratosa & Amézquita, 2012; Kluge et al., 2001). Another type of techniques can track people in the vicinity of sensors (Montemerlo et al., 1999; Schulz, Burgard, Fox, & Cremers, 2003). However, they do not provide solutions for tracking with long term occlusions.

Rosencrantz, Gordon, and Thrun (2003) have overcome the problem of tracking with temporary occlusions, however the techniques is not reliable when it is tracking outside the sensory ranges.

It has been observed that human motion often follows place dependent patterns as it is influenced by a combination of social, psychological and physiological constraints (Altman, Rapoport, & Wohlwill, 1980; Arechavaleta, Laumond, Hicheur, & Berthoz, 2006; Dean, 1996; Hall, 1969). Therefore, if such a model could be learned by a robot, it could subsequently be used to improve people tracking.

Approaches to the learning of place dependent motion pattern models have been proposed in the past. Bennewitz et al. (2005) used a network of laser scanners to learn motion patterns of individual occupants of an office environment. After collecting a data set, Expectation Maximization (EM) was used to cluster trajectories for building a Hidden Markov Model (HMM). The HMM was used to implement collision avoidance behaviors. However, this technique relies on environment mounted sensors, which needs infrastructure modifications and hence it leads to difficulty in deployment. In Kanda, Glas, Shiomi, and Hagita (2009), multiple laser scanners were used to learn activity patterns of people in a shopping mall. The idea was to automatically identify potential customers in order to communicate with them. The approach

* Corresponding author. Tel.: +61 295147629; fax: + 61 295142655.

E-mail addresses: Sarath.Kodagoda@uts.edu.au (S. Kodagoda), Stephan.Sehestedt@uts.edu.au (S. Sehestedt).

seems appealing, however it uses six infrastructure mounted laser scanners and off-line learning framework which may limit the feasibility to be used in many other robotic applications. Using stationary laser scanners, [Luber, Tipaldi, and Arras \(2011\)](#) proposed to learn a spatial affordance map based on a Poisson process. The approach was shown to be a viable extension to Multi Hypothesis Tracking (MHT) with improved tracking capabilities and robustness. However, it uses stationary sensors. [Vasquez, Fraichard, and Laugier \(2009\)](#) proposed Growing Hidden Markov Models to learn motion patterns. Although it implements an on-line learning algorithm, the technique requires the observer to be stationary. [Lookingbill, Lieb, Stavens, and Thrun \(2005\)](#) used a small helicopter with a camera as a semi-stationary observer to monitor a single roundabout to build an activity histogram. This histogram was shown to improve target tracking notably, however it may not be feasible nor suitable to be used in the indoor scenarios like the one proposed in our paper. [Bruce and Gordon \(2004\)](#) used training data to learn goal locations in a small environment. This information was then used to provide a particle filter based tracker with an improved prediction model, which was shown to perform better than Brownian motion for prediction. This work relies on previously learned goal locations in a given environment.

In general, the methods proposed in the literature have many limitations as described in the previous paragraphs such as the requirement of infrastructure based sensors, difficulty of operation under partial observability or occlusions, limited on-line adaptability, and on-line operation. Further, they are not capable of simultaneously improving both the model learning and tracking. Therefore, here we propose simultaneous people tracking and motion pattern learning based on our previous work on SHMMs ([Sehestedt, Kodagoda, & Dissanayake, 2010](#)), which uses robot mounted sensors for on-line learning and can effectively handle partial observations with limited FOV of the sensors. This framework allows the robot to improve its tracking abilities with the availability of a learned model, while improving the model learning with the feedback from the improved tracker. The SHMM based framework was tested in an office environment using the robot LISA shown in [Fig. 1](#) with appealing results. More specifically, the contributions of this paper are, 1. Formulation of the Sampled Hidden Markov Model for effective handling of on-line learning and model adaptation to capture the changes in the human motion patterns. 2. Synthesis of the theoretically sound probabilistic algorithm for the simultaneous people tracking and motion pattern learning for improving both aspects. 3. Implementing and testing the algorithm and obtaining superior results for long term occlusions when comparing with conventional (model based) methods.

This paper is organized as follows. Section 2 gives a brief introduction to SHMMs for on-line learning of common human motion patterns. Section 3 introduces and formulates the simultaneous



Fig. 1. The LISA robot.

people tracking and motion pattern model learning. Section 4 presents experimental results showing the viability and effectiveness of the proposed approach. Finally, Section 5 summarizes the contributions and briefly discusses current and future work.

2. Sampled Hidden Markov Models

In this section, a brief introduction to SHMM learning is given. SHMMs provide a sparse representation of common motion patterns and can be learned on-line using a mobile robot's on-board sensors. The importance of this can be seen in [Fig. 2](#), which illustrates a robot as a red circle and its current laser scan as a red outline. It can be observed that the field of view (FOV) of the robot is a small fraction of the size of the operating environment. Thus, any learning algorithm that can be used in such environments needs the ability to incrementally learn with partial observations.

An SHMM is defined by its states S and state transition matrix A where each state is represented by a set of weighted samples. Although, these sample sets can represent arbitrary probability distributions, for notational simplicity, the states are defined by their means μ and covariances Σ as,

$$S = s^{(i)} = \begin{bmatrix} \mu^{(i)} \\ \Sigma^{(i)} \end{bmatrix} \quad 1 \leq i \leq N \quad (1)$$

where N is the number of states in the model. Note that the model is time dependent as learning is done incrementally, however it is omitted for convenience of notation. The state transition matrix contains the probabilities of transitions from state i to state j as

$$A = a^{(ij)} = \begin{bmatrix} K^{(ij)} \\ P(s^{(j)}|s^{(i)}) \end{bmatrix} \quad 1 \leq i \leq N \quad (2)$$

where $K^{(ij)}$ is the number of times a transition was observed, from which the probability of the transition $P(s^{(j)}|s^{(i)})$ can be calculated.

A particle filter based people tracker is used for learning the SHMM. Consider the situation in [Fig. 3\(a\)](#) where a person walked along the trajectory indicated by the arrow. The tracking algorithm produced a series of sample clusters, each of which could be interpreted as one state of an SHMM. To obtain a more sparse representation of the observed trajectory, a subset of those sample clusters was used as shown in [Fig. 3\(b\)](#), where means, covariances and transition of states are shown as red squares, red ellipses and red lines respectively. It is to be noted that the state transitions could be directly derived from the sequence of sample clusters as the temporal order of the clusters are known.

Suppose another person is observed walking along the trajectory shown by the arrow in [Fig. 3\(c\)](#), where the information needs

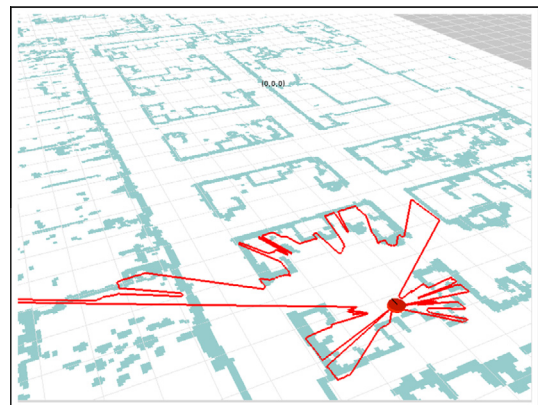


Fig. 2. A small robot observing its environment.

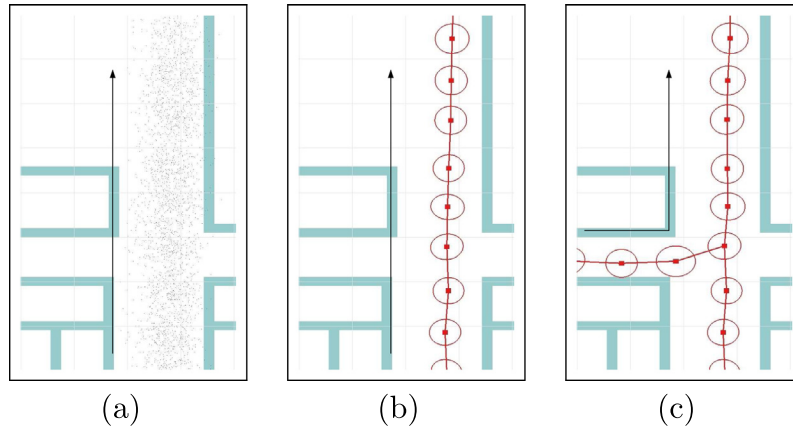


Fig. 3. A basic example of SHMM learning.

to be immediately added to the available motion pattern model. As part of the new trajectory is overlapped with the existing trajectory, a fusion mechanism is essential. Further, the non-overlapping states need to be appropriately added to the model. The symmetric Kullback–Leibler divergence (KLD) (Kullback & Leibler, 1951) is used for this purpose, which finds associations between states in the SHMM and sample clusters obtained from people tracking as,

$$KLD(s^{(i)} \| s^{(j)-}) = KLD_s(s^{(i)} \| s^{(j)-}) + KLD_{s-}(s^{(j)-} \| s^{(i)}) \quad (3)$$

with $1 \leq i \leq N$ and $1 \leq j \leq K$, where K is the number of sample clusters (candidate states, $s^{(j)-}$) reported by the tracker. Whenever an association is found, the new information is added to the corresponding state and state transitions can be updated by counting. The non-associated sections are added as new states.

Now the probabilistic formulation of the SHMM learning approach is given by the belief of motion patterns D_t at time t conditioned on the robot's location estimate and people tracking results. That is the belief,

$$Bel(D_t) = P(D_t | \zeta_t, \zeta_t, z_t, \dots, \zeta_0, \zeta_0, z_0) \quad (4)$$

is to be computed, where ζ_t is the robot localization hypothesis, ζ_t is the people tracking result and z_t the sensor reading at time t .

From this an incremental update rule can be derived using the Bayes theorem as,

$$Bel(D_t) = \frac{P(\zeta_t | D_t, \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0) \cdot P(D_t | \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0)}{P(\zeta_t | \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0)} \quad (5)$$

Since the denominator is independent of D_t , it can be written as,

$$Bel(D_t) = \eta P(\zeta_t | D_t, \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0) \cdot P(D_t | \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0) \quad (6)$$

where $\eta = P(\zeta_t | \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0)$ is a constant.

This is the belief of D at time t given all past observations, sensor readings and observer poses. Obviously, this is not an efficient solution for updating the belief since all observations of moving people, sensor data and observer poses up until time t would have to be remembered. Therefore, it is assumed that observations and poses are conditionally independent of past observations and poses given ζ_t and D_t , i.e. the system is Markov. Therefore,

$$Bel(D_t) = \eta P(\zeta_t | D_t, \zeta_t, z_t) \overbrace{P(D_t | \zeta_t, z_t, \zeta_{t-1}, \dots, \zeta_0, \zeta_0, z_0)}^{\text{prior belief}} \quad (7)$$

In fact the last term of this equation is the belief at time $t-1$ and thus the final update rule is written as

$$Bel(D_t) = \eta \overbrace{P(\zeta_t | D_t, \zeta_t, z_t)}^{\text{people tracking}} Bel(D_{t-1}) \quad (8)$$

As given in the above equation, the belief $Bel(D_t)$ can now be updated using the most recent observations of moving people.

3. Simultaneous people tracking and motion pattern learning

In this section, the formulation is extended to simultaneous people tracking and motion model learning. The idea is to tightly integrate learning and tracking in order to improve both performances. Therefore, the goal is to estimate

$$P(D_t, \zeta_t, \zeta_t | z_t) \quad (9)$$

which is the joint probability of the motion pattern model D_t , the position estimates of people in the robot's FOV ζ_t and the robot's location estimate ζ_t at time t given the latest sensor reading z_t . Assuming independence between these variables, it can be shown that

$$P(D_t, \zeta_t, \zeta_t | z_t) = P(D_t | \zeta_t, z_t) P(\zeta_t | z_t) \prod_{k=1}^K P(\zeta_t^{(k)} | z_t) \quad (10)$$

where K denotes the number of tracked people. It is to be noted that the first term on the right hand side is conditioned on ζ_t . Furthermore, it can be observed that the estimate of D_t is conditionally independent of z_t given ζ_t which is estimated from the latest sensor reading, leading to

$$P(D_t, \zeta_t, \zeta_t | z_t) = P(D_t | \zeta_t) P(\zeta_t | z_t) \prod_{k=1}^K P(\zeta_t^{(k)} | z_t) \quad (11)$$

The second term on the right hand side defines robot localization and the estimate is commonly conditioned on control input given as u_t resulting in

$$P(D_t, \zeta_t, \zeta_t | z_t, u_t) = P(D_t | \zeta_t) P(\zeta_t | z_t, u_t) \prod_{k=1}^K P(\zeta_t^{(k)} | z_t) \quad (12)$$

Intuitively, this equation could be solved from the right to the left, i.e. after states of people have been estimated, the result could be used to improve the robot localization. The localization data in turn would be used to determine the global positions of detected people. Moreover, the result could be used in the first term to update the model of motion patterns. However, to fully exploit the idea of an incrementally learned model of motion patterns D_t , the simultaneous utilization and learning of the model is desirable. To accomplish this, the last term of above equation could be made dependent on D_t .

$$P(D_t, \xi_t, \zeta_t | z_t, u_t) = P(D_t | \xi_t) P(\zeta_t | z_t, u_t) \prod_{k=1}^K P(\zeta_t^{(k)} | z_t, D_t) \quad (13)$$

Note that when people tracking is performed, although time has already incremented to the current discrete time step t , belief at this point is still $D_t = D_{t-1}$, which is indicated by the notation D_t . We now have a formulation of SHMM learning, which explicitly accounts for a mobile observer and takes advantage of improved tracking while learning.

3.1. Motion prediction

There is no general model for human motion prediction as such a very conservative prediction models like Brownian motion or constant velocity model is commonly used (Luber et al., 2011). In

rare cases, less conservative learned models have been presented (Bennewitz et al., 2005; Bruce & Gordon, 2004; Luber et al., 2011), however, they often use stationary observers or off-line learning, which limits the potential mobile robotic applications. Here, we propose to utilize the learned motion models at the prediction stage of a particle filter (PF) based people tracker (Arulampalam, Maskell, & Gordon, 2002) as

$$P(\xi_t | \xi_{t-1}, D_t) \quad (14)$$

Consider the scenario given in Fig. 4(a) with a person (blue ellipse) associated with state “a” in the model, state ellipses (red) and state transition lines (thicknesses are proportional to the magnitude of transition probabilities). Then in the next state, the location of the person can be predicted as indicated by the purple arrow based on dx and dy .

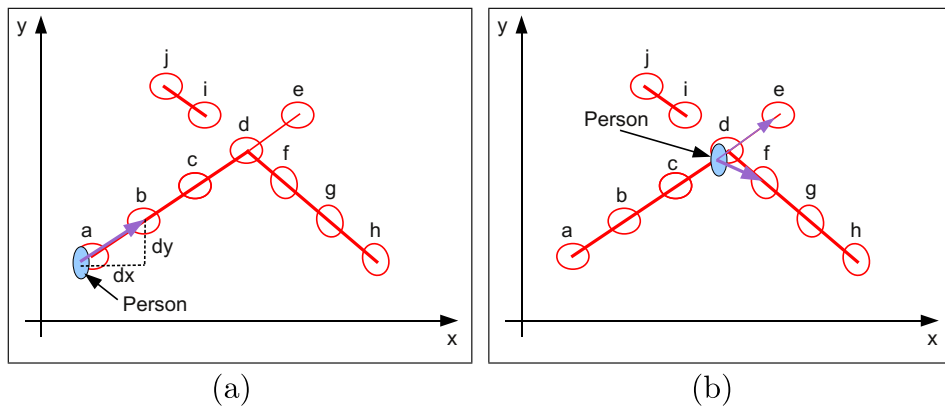


Fig. 4. Motion prediction using an SHMM.



Fig. 5. The map used for experiments. Desk areas, corridors and common are marked as “D”, “H” and “C” respectively. LISA’s pose is shown by a red circle and the current laser reading is shown as a red outline. Note the limited FOV of LISA. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

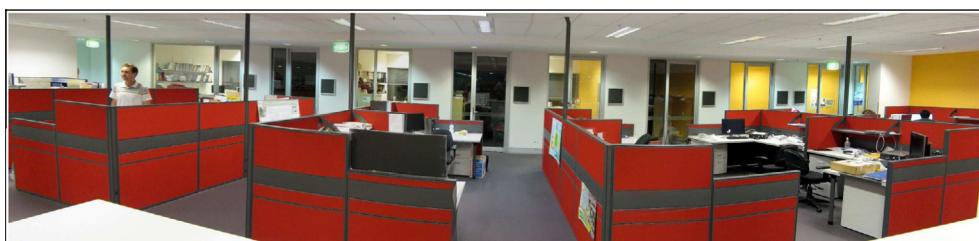


Fig. 6. A panoramic view of the large open office environment.

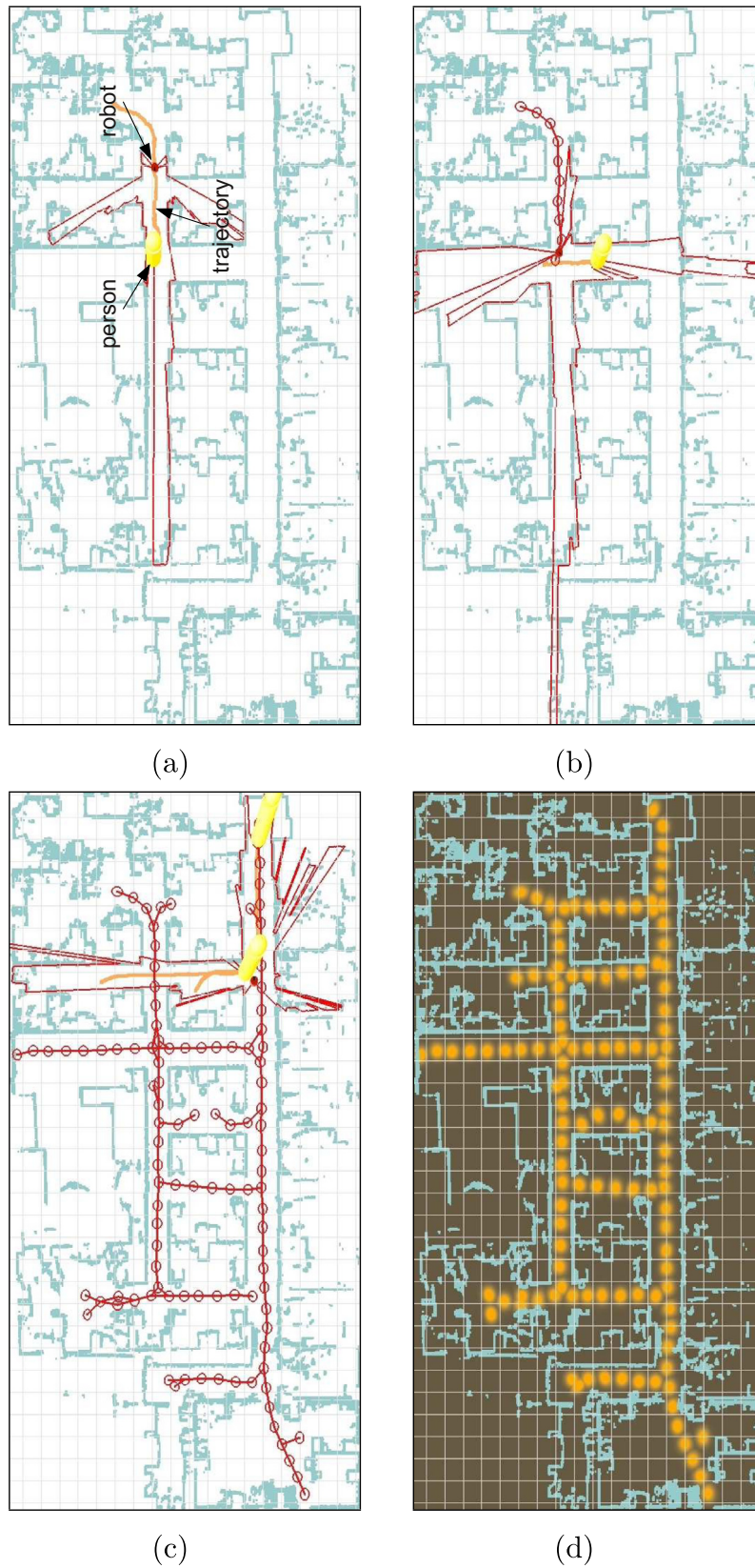


Fig. 7. Model learning with real world data.

In our current implementation, half of the samples are dedicated to SHMM based prediction and the other half is used for the prediction based on the constant velocity model to cater for some previously unlearned motion patterns. In more complex

situations, there can be more state transitions based on the learned model (see Fig. 4(b)). There the person is associated with state d has two possible transitions, e and f . As the transition probabilities are available in the model, they are used to determine the number

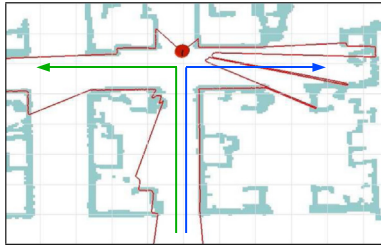


Fig. 8. Experimental set up for the evaluation of the tracking performance. Two trajectories (blue and green) were considered. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Table 1
Number of track losses in 22 trials.

Experiment	Sample size				
	50	100	200	500	1000
Stage 1	13	8	5	2	0
Stage 2	3	0	0	0	0
Stage 3	22	12	9	4	1
Stage 4	2	0	0	0	0

of samples that are allocated to each prediction. That is, the predictions with higher transition probabilities are allocated more samples than that of the prediction with lower transition probabilities. Hence the number of samples are estimated based on

$$N_D^{(j)} = P(s^{(j)}|s^{(i)}) \cdot N_D, \quad 1 \leq i \leq N \quad (15)$$

$$1 \leq j \leq N$$

where N is the number of states of the SHMM, $N_D^{(j)}$ denotes the number of samples dedicated to the transition from the current state i to state j .

3.2. Long term prediction

Long term prediction of people locations without observations is a complex and a non-trivial problem due to the agile and complicated nature of human movements and the presence of occlusions. Without having prior knowledge about movements nor recent

observations, the conventional model based predictions tend to deviate from the actual people movements leading to track losses. However, this may be solved to a greater extent, if the knowledge of human motion patterns are available.

In our approach, we utilize the learned human models based on SHMM in the prediction stage of the particle filter as described in the previous section. However, it is important to note that there is an inherent uncertainty in the learned model due to the gross nature of motion pattern learning. It gives rise to slight growing uncertainty of the estimator over time. However, as the tracker is still managed to follow the general trend of people motion, it still has significantly lower track losses.

4. Experimental results

Experiments were conducted using the Lightweight Integrated Social Autobot (LISA) shown in Fig. 1. LISA is a low cost robotic platform designed for fast and easy deployment. The base is an iRobot create equipped with a small Intel Atom X86-64 computer together with a Hokuyo UTM-30LX laser range scanner (<http://www.hokuyo-aut.jp>) for perception. The software development environment is Player/Stage (<http://playerstage.sourceforge.net/>) and all the algorithms were implemented in C++ within the Orca software framework (Brugali et al., 2007).

Fig. 5 shows a Simultaneous Localization and Mapping (SLAM) generated map of the environment where desk areas and corridors are marked appropriately. The LISA robot is shown as a red circle with the red outline illustrating the observed laser reading. Being a small robot, it has a significantly limited field of view due to the presence of furniture. The map spans approximately 32 m × 20 m. Fig. 6 shows a panoramic view of the office environment. It is important to note the complexity of the environment with large amount of clutter and semi-static objects like trash cans. Furthermore, some of the walls are made out of glass contributing to perception issues.

4.1. Learning motion patterns

In this section SHMM learning is presented with the robot LISA in the aforementioned office environment. Ten different subjects were included in this experiment and no modifications were done

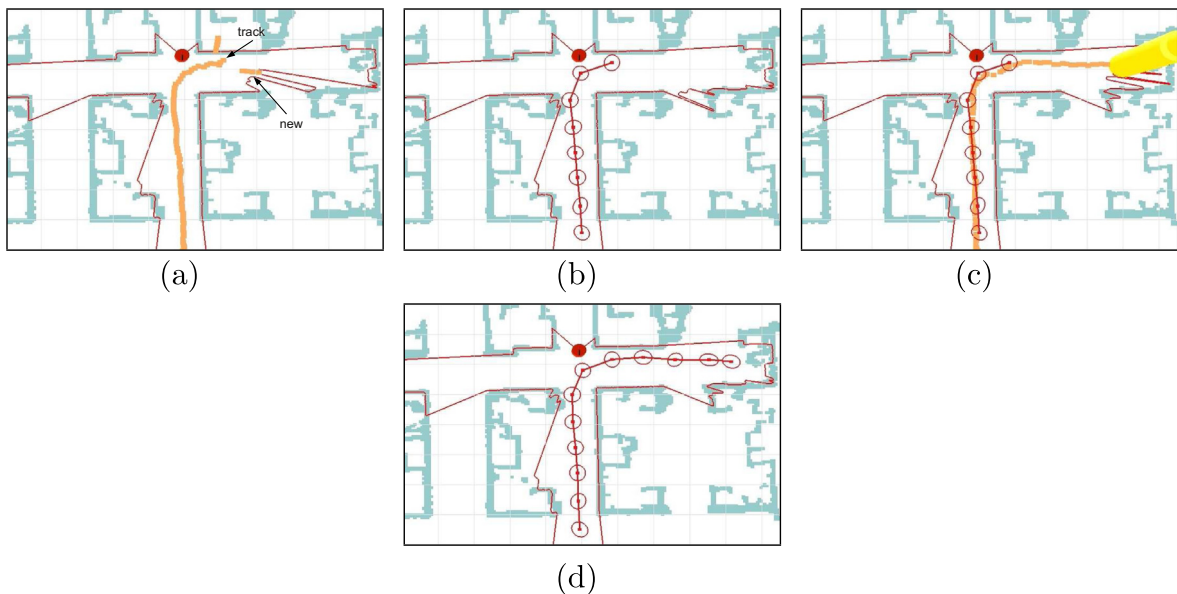


Fig. 9. Learning while track losses occur.

to the environment. The limited observability at most times means that the robot had to explore the environment to build a model of motion patterns. Furthermore, in order to observe longer trajectories the robot had to follow people, hence it was a truly mobile observer.

The series of figures in Fig. 7 show the evolution of an SHMM. Fig. 7(a) shows the robot following a person, where the person is represented by a yellow cylinder and the trajectory is shown as an orange line. The robot is shown as a red circle, where the red outline indicates the observed reading of the forward looking laser sensor. The observed trajectory exhibits a typical human motion and accordingly it is represented in the initial model in Fig. 7(b).

Fig. 7(c) shows the model after more than 70 observed trajectories while the robot was on the move. The trajectories were successfully joined and compactly represented. The final represen-

tation including more than 80 trajectories are shown in the (Fig. 7(d)) as a unimodal Gaussian distribution. It could be noted that trajectories are positioned correctly on free spaces rather than through obstacles on the map. Further, compared to grid based representations of motion patterns, a greater efficiency is achieved as the belief has to be maintained only in the relevant areas (with human motion) of interest rather than over the entire space.

4.2. Tracking robustness

In this experiment, the robot was positioned at an intersection while observing people trajectories as shown in the Fig. 8. The test data consists of 22 observed trajectories for the right turn and the same for the left turn (44 in total). The experiment was divided into four stages. In the first stage (stage 1) people followed the

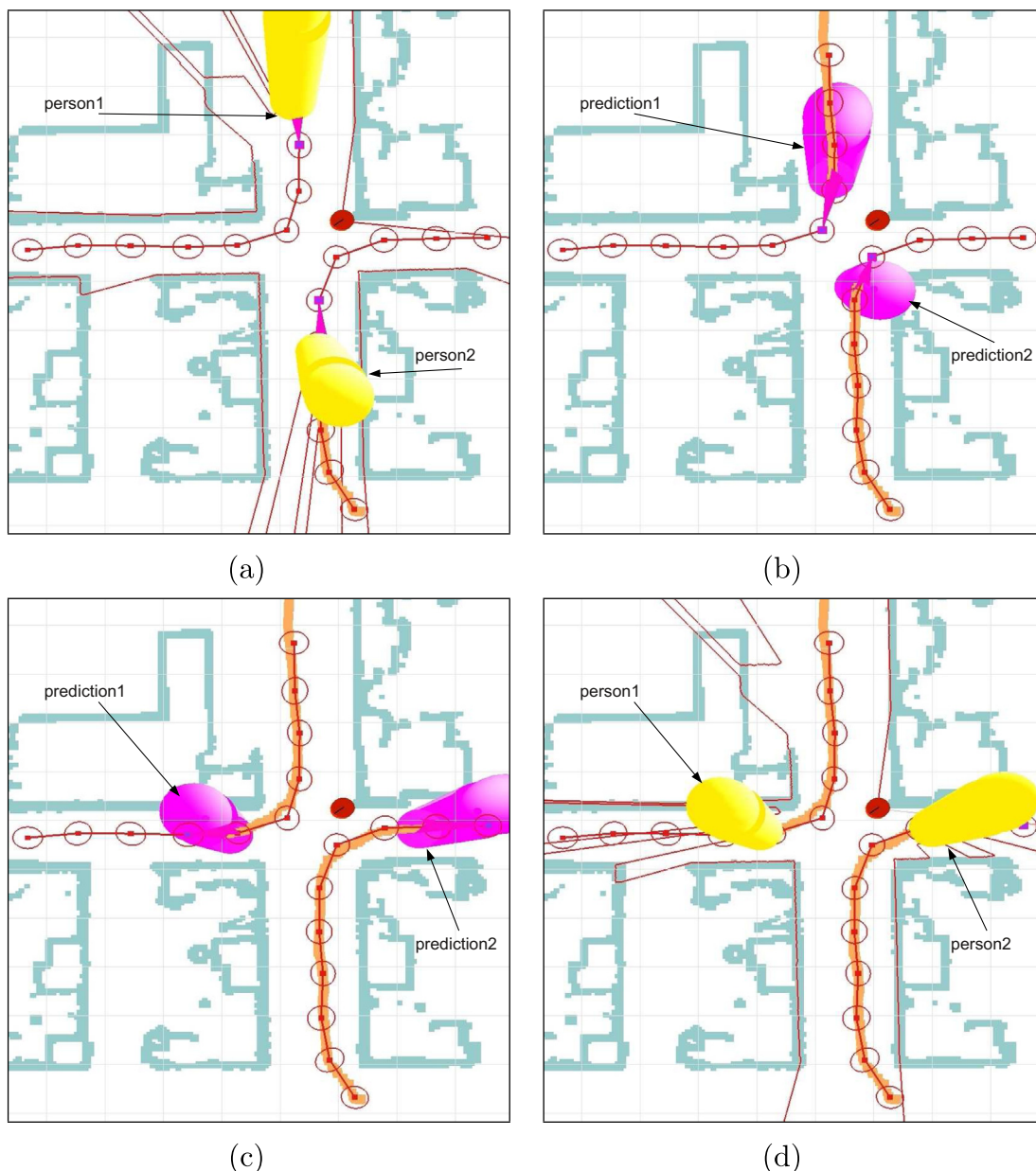


Fig. 10. Motion prediction during sensor failure in a complex situation. (a) Two people were tracked and associated with an SHMM with two disconnected trajectories. (b) LISA's sensor fails. Consequently, the position estimates of people were updated based on the SHMM. (c) Over an extended period of time, the tracks and identities were maintained using the information from the SHMM. (d) After the sensor resumed operation, both tracks could be recovered successfully and the identities were correctly maintained.

trajectory indicated by the blue arrow (right turn) to evaluate the tracking performance. In the second stage (stage 2,) a model of the blue trajectory was learned and used in the tracker to predict the motion. In the third stage (stage 3), the model from stage two was used, however, people followed the green trajectory (left turn), i.e. they exhibited a drastically different behavior compared to the previous observations. In the fourth stage (stage 4) of the experiment, the model was extended to include both types of trajectories and tested on people following either trajectory. Based on these tests, the tracking performance was evaluated in terms of the number of track losses as shown in Table 1.

In stage 1, the subjects took a sharp right turn, as indicated by the blue arrow in Fig. 8, while being observed by the robot. As there was no earlier learned SHMM, tracking was done using the near constant velocity model for motion prediction. Sample sizes of 50, 100, 200, 500 and 1000 were considered and the observed track loss rates are summarized in the first row of Table 1. Not surprisingly, it can be seen that the tracking with fewer number of samples provided large number of track losses. This is due to the sharp turn, where the constant velocity model could not adequately provide a reasonable prediction. However, with the increasing number of samples, the tracker could achieve lower track losses. In stage 2, the SHMM model was learned first with sharp right turn data before being used in the tracker. The algorithm was tested by replaying the data in stage 1 of the experiment. As shown in Table 1, the tracker performance was significantly improved with no track losses for more than 100 samples used in the PF.

In stage 3, data of the people following the right turn was used to learn the model, however tested on people following a left turn, i.e. their motion pattern would diverge from the learned model. As expected the number of tracks losses has increased significantly, even more than that of using a constant velocity model (stage 1). It can also be seen that the tracking performance has an adverse effect with the small number of samples used. Finally, in stage 4, after adaptively learning the left turn with the previously learned right turn, the tracker performed significantly better with only loss of 2 tracks even with 50 samples used in the PF.

4.3. Learning with track losses

The findings in stage 1 and 3 of above experiment can be further analysed to understand the process of learning in the presence of track losses. As discussed before, in stage 1, LISA observed people following the blue trajectory in Fig. 8 without having a priori motion pattern model. The People were tracked with a sample size of 100 in this experiment. Fig. 9(a) shows the tracking result of the first observed person, where the track was lost during the sharp turn. Fig. 9(b) illustrates the learning result based on that trajectory. As could be seen in Table 1, most of the tracks were lost during the process without contributing much to the learned model. However, once an extended track was observed as shown in Fig. 9(c), the result was immediately used to update the model as shown in Fig. 9(d). This model was further tuned with other few more extended tracks to achieve the gross motion model. This alleviates the need to keep all the trajectories by just keeping the combined trajectory.

4.4. Occlusion handling and long term prediction

Consider a part in an office environment with two strictly separated motion trajectories as shown in Fig. 10(a). The figure further shows two people who were tracked by LISA while predicting both motions based on the motion pattern model. Then, in Fig. 10(b) a sensor failure has occurred, which is indicated by the absence of the red outline of the laser scan. From this point, the two people

tracks were maintained purely based on the predictions made by the motion pattern model and the predicted estimates are shown as transparent purple cylinders in the figure. The sensor failure persisted for a while and the two people turned to their respective rights. While the constant velocity model based tracker failed to track both targets, our SHMM based tracker could successfully track in such agile situations as shown in Fig. 10(c). Furthermore, it could successfully maintain the identities of people. Finally, as in Fig. 10(d), after the sensor resumed it's normal operation, the

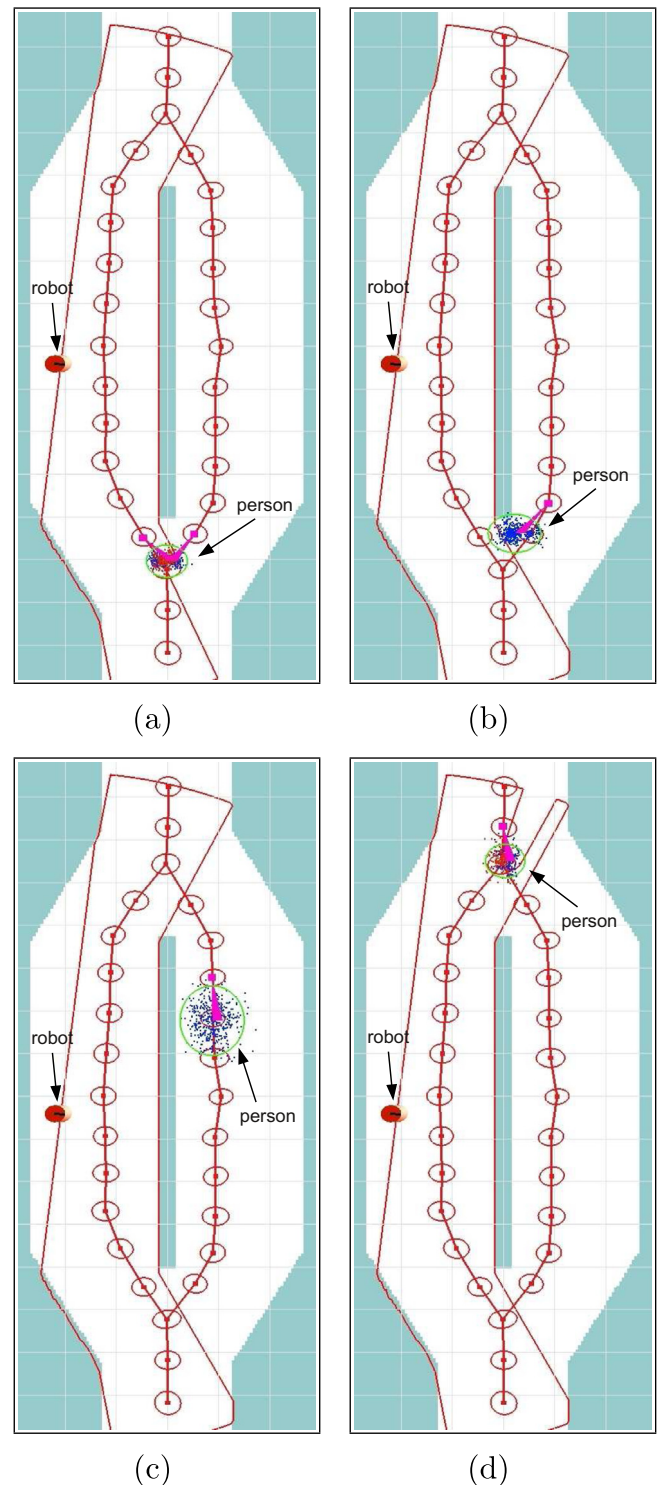


Fig. 11. Long term motion prediction.

two people were re-detected correcting associated uncertainties of the filter.

Consider a hallway which is partly divided by a wall as shown in Fig. 11(a), where the dividing wall is approximately eight meters long. A motion pattern model has been learned for this environment. At one time, there was a person appeared while the robot was observing the hallway from the left side. The person moved upwards in the figure and took the path right of the dividing wall leaving the field of view of the robot. Once the person was occluded by the wall, as there were no observations, all samples were predicted according to the SHMM (Fig. 11(b)) for continuous tracking. Although, unavailability of observations lead to slight increase of uncertainty (which is indicated by the green covariance ellipse in Fig. 11(c)), it did not contribute to track losses. The constant velocity model based implementation lost track within a couple of iterations of the particle filter. Once the observations were available towards the end of the track, the filter converged successfully as shown in Fig. 11(d).

5. Conclusions and future work

5.1. Conclusions

Model based conventional trackers have significant difficulties in tracking with long term occlusions. Therefore, this paper proposed to use human motion pattern models in people tracking within a probabilistic framework. We presented simultaneous people tracking and human motion pattern learning based on Sampled Hidden Markov Models and particle filter tracker. Learning is achieved on a mobile observer where only the robot's on-board sensors were used without relying on cumbersome infrastructural sensors. Since there was no dedicated learning phase, all new information was incorporated in the model immediately improving the adaptability. This property is of central importance which particularly distinguish our work with many of the other previously presented approaches. The representation of motion as an SHMM is more memory efficient than grid based approaches as the model only represents motion in areas of interest rather than the whole space. Furthermore, as the model is not fully connected, the transition matrix can be replaced by more compact data structures.

The knowledge of motion patterns has been shown to be useful in not only handling long term occlusions but also maintaining identities. While a reduction in the tracking errors could be observed with SHMMs integrated tracker, the most significant improvement was observed in the tracking robustness especially in complex situations like a sudden change of direction. The number of lost tracks were greatly reduced even when using small sample sizes for PF tracking. These effects in turn led to a better convergence in motion pattern learning, which again highlights the benefits of on-line learning of human motion patterns. The integration of tracking and learning was shown to have mutual benefits.

5.2. Future work

Once the robot is placed in a new environment, it takes some time for learning the motion patterns. Until a reasonable motion pattern model is learned, the tracking performance of the current algorithm suffers in accuracy. Therefore, learning the human behaviors directly from the environmental features rather than relying on observing humans in the environment (through object affordance and hallucinated humans) is a research direction that is of future interest. Further, in the context of learning of human spatial behavior, this paper has presented some of the fundamental

abilities needed for lifelong learning such as on-line learning, adaptability and the ability to include incomplete knowledge in incremental learning. One another key competency for lifelong learning under uncertainty in a changing environment is the forgetting of obsolete information, which is not addressed in this research and is considered as a great future direction.

Acknowledgments

This work was supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

References

- Altman, I., Rapoport, A., & Wohlwill, J. (1980). *Environment and culture*. Springer.
- Arechavaleta, G., Laumond, J. P., Hicheur, H., & Berthoz, A. (2006). The nonholonomic nature of human locomotion: A modeling study. In *IEEE/RAS-EMBS international conference on biomedical robotics and biomechanics*.
- Arulampalam, M. S., Maskell, S., & Gordon, N. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50, 174–188.
- Bennewitz, M., Burgard, W., Cielniak, G., & Thrun, S. (2005). Learning motion patterns of people for compliant robot motion. *International Journal of Robotics Research*, 24.
- Bolić, J. M. G., & Fernández-Caballero, A. (2011). Agent-oriented modeling and development of a person-following mobile robot. *Expert Systems with Applications*, 38, 4280–4290.
- Bruce, A., & Gordon, G. (2004). Better motion prediction for people-tracking. In *IEEE international conference on robotics and automation (ICRA)*, Citeseer.
- Brugali, D., Brooks, A., Cowley, A., Côté, C., Domínguez-Brito, A., Létourneau, D., Michaud, F., & Schlegel, C. (2007). *Trends in component-based robotics*. Springer tracts in advanced robotics (Vol. 30). Berlin/Heidelberg: Springer.
- Dean, D. (1996). *Museum exhibition: theory and practice*. Routledge.
- Francesc Serratosa, R. A., & Amézquita, N. (2012). A probabilistic integrated object recognition and tracking framework. *Expert Systems with Applications*, 39, 7302–7318.
- Gockley, R., Forlizzi, J., & Simmons, R. (2007). Natural person-following behavior for social robots. In *ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 17–24). New York, NY, USA: ACM <http://doi.acm.org/10.1145/1228716.1228720>.
- Hall, E. T. (1969). *The hidden dimension – man's use of space in public and private*. London: The Bodley Head Ltd.
- Kanda, T., Glas, D., Shiomi, M., & Hagita, N. (2009). Abstracting people's trajectories for social robots to proactively approach customers. *IEEE Transactions on Robotics*, 25, 1382–1396. <http://dx.doi.org/10.1109/TRO.2009.2032969>.
- Kluge, B., Kohler, C., & Prassler, E. (2001). Fast and robust tracking of multiple moving objects with a laser range finder. In *IEEE international conference on robotics and automation (ICRA)* (pp. 1683–1688).
- Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22, 79–86.
- Lookingbill, A., Lieb, D., Stavens, D., & Thrun, S. (2005). Learning activity-based ground models from a moving helicopter platform. In *IEEE international conference on robotics and automation (ICRA)* (pp. 3948–3953).
- Luber, M., Tipaldi, G. D., & Arras, K. O. (2011). Place-dependent people tracking. *The International Journal of Robotics Research* <http://dx.doi.org/10.1177/0278364910393538>.
- Montemerlo, M., Thrun, S., & Whittaker, W. (1999). Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *IEEE international conference on robotics and automation (ICRA)* (pp. 695–701).
- Prassler, E., Bank, D., & Kluge, B. (2002). Motion coordination between a human and a mobile robot. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 1228–1233).
- Rosencrantz, M., Gordon, G., & Thrun, S. (2003). Locating moving entities in dynamic indoor environments with teams of mobile robots. In *Second joint international conference on autonomous agents and multi agent systems* (pp. 233–240).
- Schulz, D., Burgard, W., Fox, D., & Cremers, A. B. (2003). People tracking with mobile robots using sample-based joint probabilistic data association filters. *The International Journal of Robotics Research*, 22, 99–116.
- Sehestedt, S., Kodagoda, S., & Dissanayake, G. (2010). Models of motion patterns for mobile robotic systems. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 4127–4132).
- Tadokoro, S., Hayashi, M., Manabe, Y., Nakami, Y., & Takamori, T. (1995). On motion planning of mobile robots which coexist and cooperate with human. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 518–523).
- Vasquez, D., Fraichard, T., & Laugier, C. (2009). Growing hidden markov models: An incremental tool for learning and predicting human and vehicle motion. *The International Journal of Robotics Research*, 28, 1486–1506.
- Zhu, Q. (1991). Hidden markov model for dynamic obstacle avoidance of mobile robot navigation. *IEEE Transactions on Robotics and Automation*, 7, 390–397.