Decision-Theoretic Planning with Person Trajectory Prediction for Social Navigation

Ignacio Pérez-Hurtado, Jesús Capitán, Fernando Caballero and Luis Merino

Abstract Robots navigating in a social way should reason about people intentions when acting. For instance, in applications like robot guidance or meeting with a person, the robot has to consider the goals of the people. Intentions are inherently non-observable, and thus we propose Partially Observable Markov Decision Processes (POMDPs) as a decision-making tool for these applications. One of the issues with POMDPs is that the prediction models are usually handcrafted. In this paper, we use machine learning techniques to build prediction models from observations. A novel technique is employed to discover points of interest (goals) in the environment, and a variant of Growing Hidden Markov Models (GHMMs) is used to learn the transition probabilities of the POMDP. The approach is applied to an autonomous telepresence robot.

Keywords Markov decision processes · Social robot navigation · GHMM

1 Introduction

Social robots are becoming a strong trend in the last years. It is clear that future robotic applications will require robots to coexist with human beings. In scenarios populated with people, robots should behave in a *social* manner [2]. This includes not only considering humans in a different way as other "obstacles", but also reasoning about people intentions and reacting to them. For instance, in applications like robot guidance or meeting with a person, robots have to consider people's goals and commitment in order to actuate in advance [6]. Moreover, robots may want to avoid certain places when humans intend to go not to disturb them [14].

I. Pérez-Hurtado(✉) · L. Merino
Pablo de Olavide University Seville, Seville, Spain
e-mail: iperde@upo.es

J. Capitán · F. Caballero
University of Seville, Seville, Spain

In this paper, we consider the application of telepresence robots in scenarios like meeting a person at a particular place. The objective is to increment the social intelligence and autonomy of the telepresence robot, so that the robot can execute the low-level navigation tasks and the user can concentrate in the interaction with his/her peers. This is relevant, as it has been observed that one of the problems of telepresence systems is the cognitive overload that arises by having to take low-level (navigation commands) and high-level decisions (interaction) at the same time. This may lead to mistakes at low level and to give less attention to the high-level tasks [15].

These scenarios are uncertain by nature: robot actions are not always deterministic; the environment may be dynamic; and sensors are noisy. Furthermore, the main source of uncertainty comes from people intentions, which are non-observable and should be modeled probabilistically. For this reason, Partially Observable Markov Decision Processes (POMDPs) are proposed in this paper for decision making in these setups. POMDPs provide a sound mathematical framework for decision-theoretic problems in uncertain domains, and have already been used for social applications [2, 3, 6]. Even though they have traditionally faced scalability issues and can be computationally costly, recent advances in online [12, 13] and offline [8] solvers are making POMDPs increasingly practical for robot planning in large domains.

POMDPs use prediction models to infer the state. For instance, people motion models are required in most social tasks. However, in most works, basic or handcrafted models for people movement are used [2, 6]. However, motion patterns and people intentions are affected by points of interest in the environment, and follow repetitive patterns: people move between doors and corridors following common trajectories; places of interest are common goals or affect people behavior, attracting them (e.g., vending machines) or repelling them (e.g., grass lawns); etc. Therefore, machine learning techniques are considered in the literature to estimate such *interesting* points and learn models for people intentions [3, 17].

In this paper, we contribute by using a POMDP model for a social task of a robot meeting with a person, where human motion and intentions are modeled automatically by an extension of Growing Hidden Markov Models (GHMMs) [16]. This GHMM model is first learnt from data observed by the robot and then integrated within the POMDP in order to predict the person intentions or goals. Later, an offline solver is used to compute an approximate optimal policy for the robot. During the execution phase, as the robot interact with the person, a probability distribution over the person position and intention is estimated using the learnt GHMM. That *belief* is also used to feed the POMDP policy that selects best actions for the robot in order to meet the person or avoid him/her depending on his/her goal.

We show results to prove the feasibility of the method in an indoor scenario where a telepresence robot [11] has to navigate autonomously around meeting people at certain points of interest (e.g., coffee machine) and not bothering them at others (e.g., toilets).

The remainder of the paper is as follows: Section 2 defines the problem as a social task for robot navigation; Section 3 describes our approach and models for decision making; Section 4 provides experimental results; and Section 5 discusses conclusions and future work.

Fig. 1 The telepresence robot is located in a meeting area, and the task consists of approaching people that go to typical interaction points (like coffee machines). These points of interest are learnt previously from data.

2 Problem Definition

In this work, we focus on a social task where a telepresence robot needs to interact with people in the environment. The main objective of a telepresence robot is to act as an *avatar* for a remote user, carrying a video-conference system on board and allowing that remote user to *sense* and *interact* with the environment from the distance.

As commented, the objective is that the robot carries out the low-level navigation tasks, allowing the user to concentrate on the interaction through the robot. In particular, the task we consider here allows the user to connect to the robot, which operates autonomously in a certain area (for instance, a meeting room or coffee area), waiting for people to appear (see Fig. 1). Then, the robot automatically should go and catch the person at their destination so that the remote user can establish a conversation. For that, the robot must reason about the possible intentions of the person and distinguish between two types of destinations: adequate and inadequate spots for having a conversation. For example, there are some places where people go to interact with others, like a coffee machine or a rest area. However, the robot should not disturb people when they intend to go to the toilet or exit the area.

The robot can detect and track people nearby, but its only available information for the task is a map of the scenario. The intention of each person entering in the operational area is not observable, so the robot needs to plan where to go taking into account uncertainties in people's positions and intentions. Moreover, we want the robot to discover automatically which are the locations of the *hot* spots of the scenario, where people intend to go, either to interact with others or to leave the scenario.

We propose a POMDP to model and solve this social task, since it allows the robot to deal with the uncertainties associated with its sensors, actions and the people around in a compact manner. POMDPs are also adequate for different multi-objective problems if their reward and cost functions are designed properly. Furthermore, we aim to learn the spatial structure of human trajectories in a specific environment by building an Instantaneous Topological Map (ITM) that can be viewed as a dynamic occupancy grid map. A Hidden Markov Model (HMM) is then built over this ITM and used as the transition model for the former POMDP. The details are described in the next section.

3 A POMDP for Social Navigation

The problem in Section 2 can be modeled as a POMDP where the robot maintains a belief over a person detected and its intentions. Here, we assume that the robot can only reason about a person at once, so when there are multiple people around, it only focuses on a particular one. We use a GHMM to model person motion patterns and discover automatically the points of interest in the scenario by observing people around.

3.1 POMDP Preliminaries

Formally, a discrete POMDP is defined by the tuple $\langle S, A, Z, T, O, R, h, \gamma \rangle$ [5].

- The *state space* is the finite set of possible states $s \in S$, for instance robot and people poses.
- The *action space* is defined as the finite set of possible actions that the robot can take, $a \in A$.
- The *observation space* consists of the finite set of possible observations $z \in Z$ from the onboard sensors.
- After performing an action a, the state transition is modeled by the conditional probability function $T(s', a, s) = p(s'|a, s)$, which indicates the probability of reaching state s' if action a is performed at state s.
- The observations are modeled by the conditional probability function $O(z, a, s') = p(z|a, s')$, which gives the probability of getting observation z given that the state is s' and action a is performed.
- The reward obtained for performing action a at state s is $R(s, a)$.

The state is not fully observable; at every time instant the agent has only access to observations z which give incomplete information about the state. Thus, a belief function b is maintained by using the Bayes rule. If action a is applied at belief b and observation z is obtained, a new belief b' is given by:

$$b'(s') = \eta O(z, a, s') \sum_{s \in S} T(s', a, s) b(s), \qquad (1)$$

where the normalization constant:

Fig. 2 The space is discretized by learning a topological map from people tracks obtained with the robot sensors (in blue). Some of the nodes are identified during the learning phase as goals (in red).

$$\eta = p(z|b,a) = \sum_{s' \in S} O(z,a,s') \sum_{s \in S} T(s',a,s)b(s) \tag{2}$$

gives the probability of obtaining a certain observation z after executing action a for a belief b.

The objective of a POMDP is to find a policy that maps beliefs into actions in the form $\pi(b) \rightarrow a$, so that the *value* is maximized. This value function represents the expected total reward earned by following π during h time steps starting at the current belief b: $V^{\pi}(b) = E\left[\sum_{t=0}^{h} \gamma^t r(b_t, \pi(b_t))|b_0 = b\right]$, where $r(b_t, \pi(b_t)) = \sum_{s \in S} R(s, \pi(b_t))b_t(s)$. Rewards are weighted by a discount factor $\gamma \in [0, 1)$ to ensure that the sum is finite when $h \rightarrow \infty$. Therefore, the optimal policy π^* is the one that maximizes that value function: $\pi^*(b) = \arg\max_{\pi} V^{\pi}(b)$.

3.2 States

The state s of our POMDP consists of three factors: the robot position, the person position and the person goal. As we employ a discrete POMDP, the scenario is divided into non-overlapping regions in order to discretize the robot and person positions (see Fig. 2). Each region has a centroid and all regions are combined into a topological map that is discovered automatically, as it will be described in the next section. Also, there is a finite set of goals (each goal corresponds to a region) where the person can go, which are discovered automatically too, as explained later.

We assume that the localization system of the robot is good enough to be able to determine its region with high certainty. Therefore, the robot position is assumed observable, being just necessary to keep a belief over the person position and intention. Note that the most relevant uncertainty of the problem comes from the person intentions which are non-observable by nature.

3.3 State Transitions: A Growing Hidden Markov Model

As described in Section 3.1, the POMDP planner needs a probabilistic transition function of the state $T(s', a, s)$, which models the dynamics of people locations and motion intentions. Instead of handcrafting this transition model (a Markov model), we have developed an extension of GHMMs [16] to learn this transition function from data.

In a GHMM, there is a discrete representation of the space, which is divided into regions. Transitions are only allowed between neighboring regions. The learning process consists of estimating the best space discretization, and identifying neighboring regions and transition probabilities from observed data. Thus, first a topological map is built with the ITM algorithm [4]; then, an HMM is built from the ITM, and its transition (and prior) probabilities are trained with the incremental Baum-Welch technique [7]. Once the GHMM has been trained, it is integrated with the POMDP before starting the task execution.

Learning Phase. The ITM algorithm structures the space in a graph whose nodes represent the centroid of Voronoi regions where people have been observed and edges represent connections between adjacent/neighboring regions (in the mathematical sense). In order to enforce an average geometrical distance between nodes, a threshold τ to insert new nodes is defined. Each node maintains and updates a Gaussian distribution in relation with the observations (2D positions of people) within its corresponding region. We propose a variant of the original ITM algorithm where a bivariate Gaussian distribution $\mathcal{N}(\mu_n, \Sigma_n)$ is updated for each region n after each new observation, where μ_n and Σ_n are respectively the mean and the covariance matrix of all the observations (x, y) related to n. Thus, instead of using a fixed covariance matrix for all the nodes, each node stores and updates a specific covariance matrix, so the observation model can adapt to the characteristics of the different parts of the scenario. The centroids are also updated in a different manner as in the original ITM algorithm, and each centroid is computed as the mean of its associated observations.

People goals are automatically discovered by applying hypothesis testing (t-test), there are two types of goals: entry/exit points where people appear or disappear and standing points where people stop longer than usual in the scene. The algorithm is adaptive, i.e, nodes, edges and goals are created, erased and updated dynamically as more people are observed. A $p - value$ threshold is defined in order to accept or refuse new goals.

Finally, a HMM is used to model all the transition state probabilities, being the state of a person its node of the graph and its goal or intention. In particular, prior and transition probabilities are computed by applying the Baum-Welch algorithm, including people positions and velocities as observations. A sampling ratio parameter T_s is used to sample the observed trajectories at a constant rate and feed the Baum-Welch algorithm (hence, each state transition corresponds with a time T_s). In the GHMM framework, the HMM can be trained several times during the learning phase. For this paper, the HMM has been generated and trained once after the creation of the topological map and goal discovery, since we are using an offline approach. For more details about the learning phase, please refer to [9].

Belief Estimation. Once the transition probabilities of the GHMM have been learnt (for the person position and goal), the belief over the person position and intention can be updated each T_s with Equation 1. Note that the robot actions do not affect people positions nor intentions in our model. Moreover, the robot positions are considered observable and its transition probabilities for each action are hand-coded.

3.4 Observations, Actions and Rewards

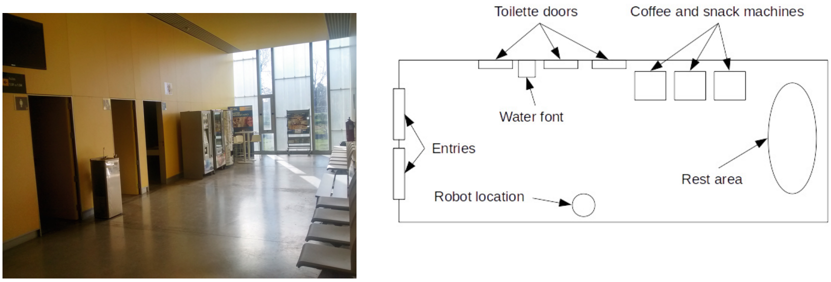
The robot has sensors onboard to measure its own pose and estimate the person position. At each moment, it can either determine the region where the person is or not detect anything. Sensors are noisy, and the probability of non-detecting the person (false negative) is p_f. Moreover, if the person is in a certain region, it could be detected in the adjacent regions.

This probability of erroneous detection depends on the distance between the centroids of the actual region and the observed region. The probability of detecting a person for each region is modeled by a Gaussian centered in the centroid of the region. Those Gaussian distributions vary for each region and are learnt together with the GHMM [9].

In addition, the robot can take movement actions at each iteration of the planner. In particular, the robot can decide either to stay where it is or to move to an adjacent region. Those transitions are not modeled as deterministic and there is certain probability that the robot may end up in an erroneous region. Moreover, there is a cost associated with moving to an adjacent region, whereas there is no cost associated with staying in the same region.

The objective of the robot is to come across the person in order to have a conversation, therefore the reward function is designed with this purpose. For that, two different types of goals are considered: adequate or inadequate. *Adequate* goals are those where the robot can go and have an interaction with the person. *Inadequate* goals are those where the robot should not bother the person and go to its *home* position, defined beforehand. Thus:

- If the person intends to go to an adequate goal and the robot is there, it gets a positive reward R_{pos}.
- If the person intends to go to an inadequate goal and the robot is at *home* position, it also gets a positive reward R_{pos}.
- If the person intends to go to an adequate goal and the robot is not there when the person arrives, it gets a negative penalty R_{neg}.
- If the person intends to go to an inadequate goal and the robot is there, it gets a negative penalty R_{neg}.



(a) (b)

Fig. 3 (a) Experimental area at university. (b) Schematic view with the main points of interest and the *home* position for the robot.

The *home* position is a region more or less centered in the scenario where the robot can wait for people to arrive. The reward function tries to encourage the robot to catch the people who go to adequate goals and to arrive there before them. It also forces the robot to go back to *home* if the person intends to go to an inadequate goal. The fact that the person comes across the robot in an inadequate place is penalized because it may be considered disturbing.

4 Experiments

In this section, we present some experimental results to show the feasibility of our approach. We implemented our decision-making algorithm in a real telepresence robot.

4.1 Experimental Setup

The scenario used for our social task is one of the rest areas in Pablo de Olavide University (Fig. 3), which is a space of 4.30×11.80 meters with a single entry/exit point at one side and several points of interest: a spot with a coffee and a snack machine, a door to the toilets, a water font and a rest area with magazines.

We implemented our methods for people motion model and decision making in C++ under the Robot Operating System (ROS) framework. We used for the experiments the TERESA robot [11], which is equipped with two laser-scanners (front and back) and a video-conference system. The robot had a map of the scenario and was able to localize itself and navigate between waypoints thanks to the ROS navigation stack (`amcl` and `move_base` packages).

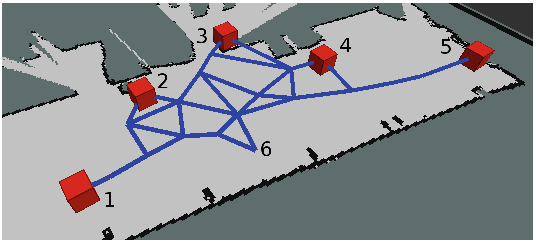


Fig. 4 Topological map (blue) and discovered goals (red) for the scenario. Node 6 of the topological map is used as *home* position for the robot.

First, we placed the robot in the scenario without moving from the *home* position, just observing trajectories of people passing by. With that information, we ran the algorithm described in Section 3.3 to learn a topological map of the scenario with the centroids of the regions, the possible goals for people and a GHMM with the transition probabilities.

The robot used the two laser-scanners for person detection and tracking, applying the algorithm in [1] and a Kalman Filter for temporal tracking and velocity estimation. More than 200 people trajectories were recorded in a dataset and used to train the models[1], generating a topological map with 21 nodes, 32 edges and 5 discovered goals (see Fig. 4). The algorithm was able to discover as goals all the points of interest in the scene: (1) entry/exit door; (2) water font, (3) toilets, (4) coffee/snack machines, (5) rest area. The corresponding GHMM was trained by sampling the people trajectories at $T_s = 1$ Hz, resulting in 105 states and 1,705 transition probabilities.

Once the GHMM was learnt, we implemented our POMDP model[2] to obtain a policy for the robot. In this case, we used an offline POMDP solver, Symbolic Perseus [10]. During the experiments, we ran two different modules: a module using the GHMM to estimate the belief of the person position and intention, and a module to determine the best action for the robot at each time given the current belief. The estimator module is executed at 1 Hz whereas the decision-maker at 0.33 Hz. Moreover, the decision-maker commands the robot to stay at the same region or to go to adjacent ones, which means sending to the `move_base` navigator the corresponding waypoint (centroid of the destination region).

4.2 Results

In order to evaluate the behavior of the robot with the computed policy, we ran different trials where people were appearing at the scenario and going to different places. In general, we observed a common behavior: the robot waits before moving

[1] The ITM algorithm was executed with $\tau = 1$ meter for node insertion and $p - value = 10^{-4}$ for hypothesis testing.

[2] The parameters were set as $p_f = 0.1$, $R_{pos} = 10$ and $R_{neg} = -10$.

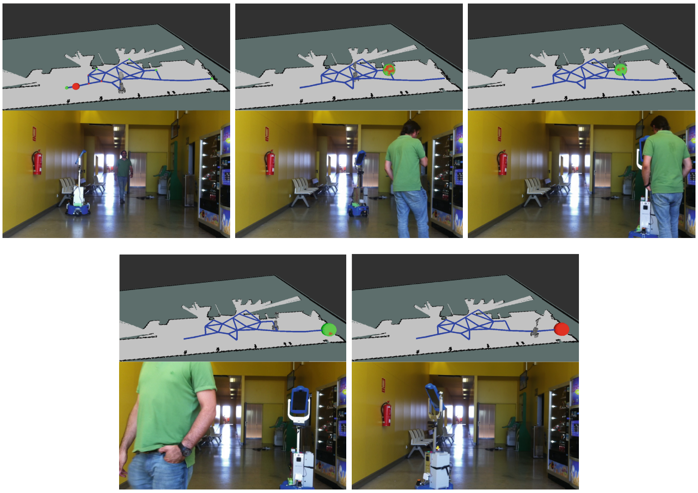


Fig. 5 Several snapshots with the person going to the coffee machine and later to the rest area.

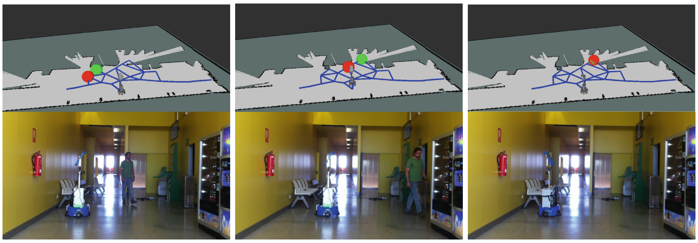


Fig. 6 Several snapshots with the person going to the toilet.

when it is not very sure about the person intention (high uncertainty in the goal belief); however, when the certainty on the person intention increases, the robot goes and perform its task, i.e., meeting the person if he/she goes to an adequate destination or coming back to the *home* position if he/she goes to an inadequate destination.

In Figures 5, 6 and 7, three sequences of snapshots of an experiment are depicted. Due to space limitations, only a few examples of the robot behavior are shown. Their corresponding representations with the RVIZ visualization tool are also shown, where the belief over the person positions is depicted with red spheres at the nodes of the topological map (the bigger the sphere, the higher the probability), and the belief over the intentions with green spheres at the goal positions.

In particular, it can be seen in Figure 5 how the robot waits until it knows with certainty where the person goes, and then it goes to the coffee machine to meet him. Later, it also follows him to the rest area. In Figure 6, the robot beliefs at first that the person goes to the water font and starts moving there, but then it changes its belief to the toilet and decides not to go back to the *home* position. Finally, in Figure 7, the

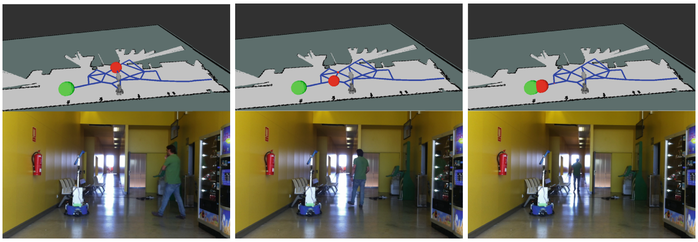


Fig. 7 Several snapshots with the person going out of the toilet and leaving the room. Real images and RVIZ representations are combined together.

person gets out of the toilet and its intention is clear from the beginning, so the robot does not move and let the person exit the room.

5 Conclusions

This paper has presented a decision-theoretic approach for social navigation under uncertainties. The social task (where a robot needs to interact with people) and its associated uncertainties are modeled using a POMDP, which allows the robot to reason in a principled manner about the possibilities in the future, and about the uncertain (non-observable) intentions of people. Moreover, a prediction model for people motion is learnt from observed data and used to train the POMDP policy. In particular, an extension of the GHMM framework is proposed, which allows the robot to discover automatically points of interest in the scenario and the transition probabilities for people movements.

The paper applies the concepts to a social task with a telepresence robot, and it presents results obtained with a real system. These results show how the robot is able to accomplish its task in a satisfactory way even with noisy sensors and reasoning on future people intentions. We believe that our approach should behave in a more robust manner than other greedy policies where the planner does not consider uncertainties in the future. If the robot only takes into account the current belief on the person intentions, it will be more prone to erroneous decisions based on incorrect beliefs.

As future work, we plan to perform a benchmarking comparison of our policy against other simpler (greedy) policies using systematic simulations. Also, we would like to explore the use of online approaches, where the POMDP prediction model and policy are learnt during the execution of the task.

Acknowledgements This work is partially funded by the EC-FP7 under grant agreement no. 611153 (TERESA) and by the Junta de Andalucia through the project PAIS-MultiRobot (TIC-7390). I. Pérez-Hurtado is also supported by the Postdoctoral Junior Grant 2013 co-funded by the Spanish Ministry of Economy and Competitiveness and the Pablo de Olavide University.

References

1. Arras, K.O., Mozos, O.M., Burgard, W.: Using boosted features for the detection of people in 2D range data. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3402–3407. Roma, April 2007
2. Bai, H., Cai, S., Ye, N., Hsu, D., Lee, W.S.: Intention-aware online POMDP planning for autonomous driving in a crowd. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 454–460, May 2015
3. Foka, A., Trahanias, P.: Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. International Journal of Social Robotics **2**(1), 79–94 (2010). http://dx.doi.org/10.1007/s12369-009-0037-z
4. Jocksuch, J., Ritter, H.: An instantaneous topological map for correlated stimuli. In: Proceedings of the International Joint Conference on Neural Networks, vol. 1, pp. 529–534. Washington (1999)
5. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial Intelligence **101**, 99–134 (1998)
6. Merino, L., Ballesteros, J., Pérez-Higueras, N., Ramón-Vigo, R., Pérez-Lara, J., Caballero, F.: Robust person guidance by using online POMDPs. In: Armada, M.A., Sanfeliu, A., Ferre, M. (eds.) ROBOT 2013: First Iberian Robotics Con-ference, Advances in Intelligent Systems and Computing, vol. 253, pp. 289–303. Springer International Publishing (2014)
7. Neal, R., Hinton, G.: A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan, M. (ed.) Learning in Graphical Models (1998)
8. Ong, S.C., Png, S.W., Hsu, D., Lee, W.S.: POMDPs for robotic tasks with mixed observability. In: Proceedings of the Robotics: Science and Systems Conference. Seattle (2009)
9. Perez-Hurtado, I., Capitan, J., Caballero, F., Merino, L.: An extension of GHMMs for environments with occlusions and automatic goal discovery for person trajectory prediction. In: European Conference on Mobile Robots. p. To appear. Lincoln (2015)
10. Poupart, P.: Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes. Ph.D. thesis, University of Toronto (2005)
11. Shiarlis, K., Messias, J., van Someren, M., Whiteson, S., Kim, J., Vroon, J., Englebienne, G., Truong, K., Evers, V., Perez-Higueras, N., Perez-Hurtado, I., Ramon-Vigo, R., Caballero, F., Merino, L., Shen, J., Petridis, S., Pantic, M., Hedman, L., Scherlund, M., Koster, R., Michel, H.: Teresa: a socially intelligent semi-autonomous telepresence system. In: Workshop on Machine Learning for Social Robotics at ICRA-2015. Seattle (2015)
12. Silver, D., Veness, J.: Monte-Carlo planning in large POMDPs. In: Advances in Neural Information Processing Systems (2010)
13. Somani, A., Ye, N., Hsu, D., Lee, W.S.: DESPOT: Online POMDP planning with regularization. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, pp. 1772–1780. Curran Associates, Inc. (2013)
14. Tipaldi, G.D., Arras, K.O.: Planning problems for social robots. In: Proc. International Conference on Automated Planning and Scheduling (ICAPS 2011). Freiburgu (2011)
15. Tsui, K., Desai, M., Yanco, H., Uhlik, C.: Exploring use cases for telepresence robots. In: 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 11–18, March 2011
16. Vasquez, D., Fraichard, T., Laugier, C.: Growing Hidden Markov Models: An Incremental Tool for Learning and Predicting Human and Vehicle Motion. The International Journal of Robotics Research **28**(11–12), 1486–1506 (2009)
17. Xie, D., Todorovic, S., Zhu, S.: Inferring "dark matter" and "dark energy" from videos. In: IEEE International Conference on Computer Vision (ICCV), pp. 2224–2231 (2013)