Multimodal Human-Robot Interaction using Gestures and Speech Recognition

João Garcia
Instituto Superior Técnico
University of Lisbon
Lisbon, Portugal
Email: joao.p.garcia@tecnico.ulisboa.pt

***Abstract*—This work proposes a Decision-Theoretic (DT) approach to problems involving interaction between robot systems and human users, which takes into account the latent aspects of Human-Robot Interaction (HRI), e.g., the user's status. The presented approach is based on the Partially Observable Markov Decision Process (POMDP) framework, which efficiently handles uncertainty in planning problems involving physical agents, extended with information rewards (POMDP-IR) to optimize the information-gathering capabilities of the system. The approach is formalized into a framework which considers: observable & latent variables; gesture & speech rooted observations; and action factors which are related to the agent's actuators or to the information gain goals (Information-Reward (IR) actions). Under the proposed framework, the robot system is able to: actively gain information and react according to hidden features, inherent to HRI settings; effectively achieve the goals of the task in which the robot is employed; and follow a socially appealing behavior. Finally, the framework was thoroughly tested in a socially assistive scenario, in a realistic apartment testbed and resorting to an autonomous mobile social robot. The experiments' results prove the validity of the proposed approach for problems involving robot systems in HRI scenarios.**

I. Introduction

Recent technological developments have extended robotics to social settings. Beyond the basic capabilities of moving and acting autonomously, robots in Human-Robot Interaction (HRI) scenarios need to communicate and interact with human users in a social and engaging manner. In this context, socially intelligent robotics emerged with the purpose of creating robots capable of exhibiting natural social qualities.

Social robots need to be capable of developing affective interactions and to empathize with human users [1]. This requirement involves the ability to infer and react according to latent (hidden) variables: the user's affective and motivational status.

The agent acting in a HRI scenario must take into account the effects of its actions in the human user, which are uncertain, and the sensory information it receives, which is noisy. Planning under these conditions is attainable through Partially Observable Markov Decision Processes (POMDPs) [2]. POMDPs, through the transition and observation models, deal with the aforementioned uncertainty, by statistically representing the possible outcomes of the agent's different actions and the accuracy of the sensory information. Furthermore, the problem of empathizing with the human user adds the goal of information gain on latent variables, which is addressed by the extensions to POMDPs introduced by Partially Observable Markov Decision Processes with Information Rewards (POMDPs-IR) [3].

A. Related Work

In HRI scenarios, Decision-Theoretic (DT) approaches to planning based on the POMDP framework are found in assistive scenarios, such as the robotic wheelchair [4], where the goal is to recognize the intention of the user but do not include social capabilities to improve recognition. Also, in socially assistive settings, the POMDP framework models the social interaction between robot and human users in, for instance, nursing homes [5], although not taking into account the user's status. Finally, the POMDP was used to model problems with latent variables and adapt the agent's behavior accordingly in an automated hand-washing assistant [6]. However, the agent in the latter work does not actively seek to gain information on the user's status, and is, therefore, limited to react based on a possibly uncertain belief on the hidden variables.

None of the aforementioned works is based on the POMDP-IR framework, which provides the means to reward low uncertainty and actively gain information on the features of interest. DT planning based on POMDP-IR has only been applied to the problem of active cooperative perception [3]. The present work, however, is focused on the problem of accomplishing a given task, in a HRI scenario, while actively inferring hidden variables of interest.

B. Contributions

The current project studies planning under uncertainty in HRI scenarios. In this context, this work presents a Decision-Theoretic (DT) approach to the problem of decision making in the aforementioned scenario, based on POMDPs-IR. This approach allows to handle uncertainty in the state of the environment, introduced by noisy sensors and the latent aspects of HRI. Furthermore, the proposed approach is tested in a set of experiments which consider a Network Robot System (NRS) in a socially assistive setting.

Summarizing, the contributions of this work are:

- A framework for DT planning under uncertainty in HRI problems, which allows the agent to accomplish a given

task, actively infer latent variables of interest and adapt its behavior accordingly;
- The development of a NRS employed in a socially assistive task.

This work starts to provide a brief overview on the DT models relevant to its comprehension, in Section II. Subsequently, Section III introduces the DT framework and Section IV applies the framework to a socially assistive scenario: the robot therapist. Section V explains and discusses the experiments performed in order to validate the proposed approach and, finally, Section VI draws final conclusions and discusses potential directions for future research.

II. Background

Markov Decision Processes (MDPs) provide the mathematical framework to formulate and solve decision-making problems in stochastic environments, for a single agent, and assuming full knowledge of the environment.

A. Partially Observable Markov Decision Processes

The assumption that the state of the environment is known to the agent is not realistic in many practical scenarios involving physical agents. The POMDP extends the MDP framework to partially observable environments and provides a framework for planning in problems which combine partial observability, uncertain action effects, unknown environment dynamics and multiple objectives.

Formally, a POMDP is a tuple (S, A, T, R, Ω, O), in which:

- $S = \{s^1, \ldots, s^N\}$ is a finite set of mutually exclusive states, defining the model of the world;
- $A = \{a^1, \ldots, a^M\}$ is a finite set of actions at the disposal of the agent;
- T is the transition function $T : S \times A \times S \rightarrow [0, 1]$. The Transition model $T(s, a, s') = P(s'|s, a)$ represents the probability of reaching state s' from s if action a is performed;
- R is the reward function $R : S \times A \rightarrow \mathbb{R}$. The reward model $R(s, a)$ defines the numeric reward for the agent to perform action a while in state s. This model reflects the agent's goals or preferences;
- $\Omega = \{o^1, \ldots, o^W\}$ is a finite set of observations that correspond to features of the environment which can be directly perceived by the agent's sensors.
- O is the observation function $O : S \times A \times \Omega \rightarrow [0, 1]$. The Observation model $O(o, a, s') = P(o|a, s')$ corresponds to the probability of observing o after performing action a and reaching state s'.

In a partially observable environment, the system does not satisfy the Markov property, since observations do not uniquely identify the state of the environment and a direct mapping of observations to actions does not suffice to achieve optimal behavior. Therefore, an history of executed actions and perceived observations would be necessary to infer the current state. Storing all actions and observations would require increasing memory over time, rendering it impractical for long planning horizons.

A solution is to encode the aforementioned history in a probability distribution over all states: the belief state. This belief $b(s)$ is a vector that denotes the probability that the state of the environment is $s \in S$. b is dynamically updated by the Bayes' rule, every time the agent performs an action $a \in A$ and observes $o \in \Omega$:

$$b^{ao}(s') = \frac{P(o|s', a)}{P(o|b, a)} \sum_{s \in S} P(s'|s, a) b(s). \tag{1}$$

In Equation 1, $P(s'|s, a)$ and $P(o|s', a)$ are defined by the Transition and Observation model, respectively. $P(o|b, a)$ is a normalizing constant, defined by:

$$P(o|b, a) = \sum_{s' \in S} P(o|s', a) \sum_{s \in S} P(s'|s, a) b(s). \tag{2}$$

The belief respects the Markov property, i.e., at each transition, the next belief state only depends on the current belief, action and observation. Also, it is a sufficient statistic of the history, meaning the agent's performance is not affected by not memorizing the complete sequence of actions and observations.

The goal of the agent is to choose actions to successfully complete its task following the best behavior possible, i.e., to compute the optimal policy. A policy $\pi(b)$ maps belief states to actions, indicating the action to perform for each belief. It is, therefore, a function over a continuous set of probability distributions over the state space S. The evaluation of a policy is done through the value function $V^{\pi}(b)$, defined as the expected future discounted reward given to the agent by following the policy π, starting from belief b:

$$V^{\pi}(b) = \mathbf{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(b_t, \pi(b_t)) \Big| b_0 = b\right], \tag{3}$$

where:

$$R(b_t, \pi(b_t)) = \sum_{s \in S} R\big(s, \pi(b_t)\big) b_t(s). \tag{4}$$

Equation 4 defines the reward given to the agent in time step t, according to the belief state while following policy π.

The policy that maximizes the value function is the optimal policy π^*. It indicates the optimal action to perform in the current time step for each belief state b, assuming the agent will also act optimally onward. The value of the optimal function is called the optimal value function V^*.

The optimal value function satisfies the Bellman optimality equation $V^* = H_{POMDP} V^*$:

$$V^*(b) = \max_{a \in A} \left[R(b, a) + \gamma \sum_{o \in O} P(o|b, a) V^*(b^{ao})\right], \tag{5}$$

with b^{ao} given by Equation 1, $P(o|b, a)$ as defined in Equation 2 and $R(b, a)$ given by Equation 4.

Solution methods for POMDPs differ from exact solution algorithms (e.g., Monahan's enumeration algorithm [7]), intractable for large problems, to approximate policy optimization (e.g., Point-based Value Iteration (PBVI) [8]). The method

of reference in solving POMDPs throughout this work is *PERSEUS* [9], a randomized PBVI algorithm.

B. Factored Models

The environment, for a given problem, can be represented through certain features of interest (e.g., location of the human user and battery level of the robot). If each feature is associated with a variable X_i, with domain D_i, the state space becomes the cross product of the variables related to the features of the environment:

$$S = \{D_1 \times D_2 \times \cdots \times D_k\},$$

considering an environment with k features.

Similarly, each actuator or actuation feature (e.g., the direction and speed of the agent) define the action space as a combination of variables A_j. Finally, each sensor or type of information transform the observation space in a likely manner.

Models with this representation are denoted factored models [10]. These models exploit the structure of the decision-making problem to solve it more efficiently.

C. POMDP with Information Rewards

The traditional POMDP model defines a state-based reward function, which does not reward information gain. Consequently, if information gain is one of the objectives of a task, the POMDP framework needs to be extended to allow rewarding low-uncertainty beliefs. This extension is provided through the POMDP-IR framework [3].

The information gain goal is achieved via the inclusion of Information-Reward (IR) actions, which allow rewarding the agent for achieving a certain knowledge on particular features of the environment, namely specific state factors. Therefore, the standard POMDP action space, denoted as A_d, is extended with an IR action for each state factor of interest. That is, for a state factor of interest X_i with domain $\{x_1, x_2, \cdots, x_j\}$, the corresponding IR action is:

$$A_i = \{commit_1, commit_2, \cdots, commit_j, null\},$$

and the action space of the POMDP-IR becomes:

$$A^{IR} = A_d \times A_1 \times A_2 \times \cdots \times A_l,$$

where l is the number of IR actions.

At each time step, the agent performs a domain-level action $a \in A_d$, and chooses an extra information action for each state factor of interest. The latter do not influence the transition nor the observation model, but change the reward given to the agent:

$$R^{IR}(X, A) = R_d(X, A_d) + \sum_{i=1}^{l} R_i(X_i, A_i),$$

where R_d is the reward function of the POMDP model and R_i is the information reward.

Every time step, the agent either makes no assumption regarding the information objectives, by choosing the IR action $A_i = null$, or collects the reward for its belief over X_i, through $A_i = commit_k$, $1 < k < j$. The rewards given to the agent for a correct or an incorrect assumption are $r_i^{correct}$ and $-r_i^{incorrect}$, respectively. The presence of information rewards influence the policy towards lowering the uncertainty associated with the state factor of interest.

The threshold of belief regarding a particular state factor ($b(X_i = x_k)$) above which the IR action is $commit_k$, is denoted β:

$$\beta = \frac{r_i^{incorrect}}{r_i^{correct} + r_i^{incorrect}}. \tag{6}$$

The exact values of $r_i^{correct}$ and $r_i^{incorrect}$ depend on the problem and need to take into account the rewards given for other tasks, such as R_d. Rewarding too much the IR action in regard to the other actions, might induce the agent to ignore the other tasks. Also, the value of β should be such that $b(X_i = x_k) > \beta$ is reachable. Therefore, the value of β is dependent on the sensory limitations of the agent, particularly on the ability of the agent to observe the state factor of interest X_i.

III. Framework Description

The proposed approach models the considered problem under the POMDP-IR theory [3]. Figure 1 represents the projected framework as a two-stage Dynamic Bayesian Network (DBN), which depicts the dynamics of the HRI problem.

A. States and Transitions

The agent acting in a HRI scenario considers two types of state factors: the *task* variables T and the *person* variables P. The *task* variables model the environment features that provide information on the progress of the tasks. On the other hand, the *person* variables track the human status and are inherently latent. The latter are used to gain information on the human user's affective and motivational status and adapt the robot behavior accordingly.

The number of variables depend on the amount of features essential to represent the environment and is, therefore, dependent on the specific task. The criteria for the selection of states involve a trade-off between operational complexity and predicted system performance, since operational complexity increases with the number of states.

Furthermore, depending on the objectives of the agent acting in a HRI setting, the *task* variables might not exist. This is the case when the single goal of the agent is to gain information on the human user, e.g., a robot psychologist.

A *person* variable can be constant if its value does not change during the task. This is the case of personal traits (e.g., *Personality* and *Preferences*), which are relevant for the robot behavior and do not alter for the duration of the interaction. In Figure 1, P_k represents a constant *person* variable. The value of P_k at each time step only depends on the value of the same variable in the previous time step. Otherwise, *person* variables are inferred from the user's behavior (factors P_1 to P_j in Figure 1), which is represented in the model's observations. These may consist of state factors of interest, according to the POMDP-IR framework.

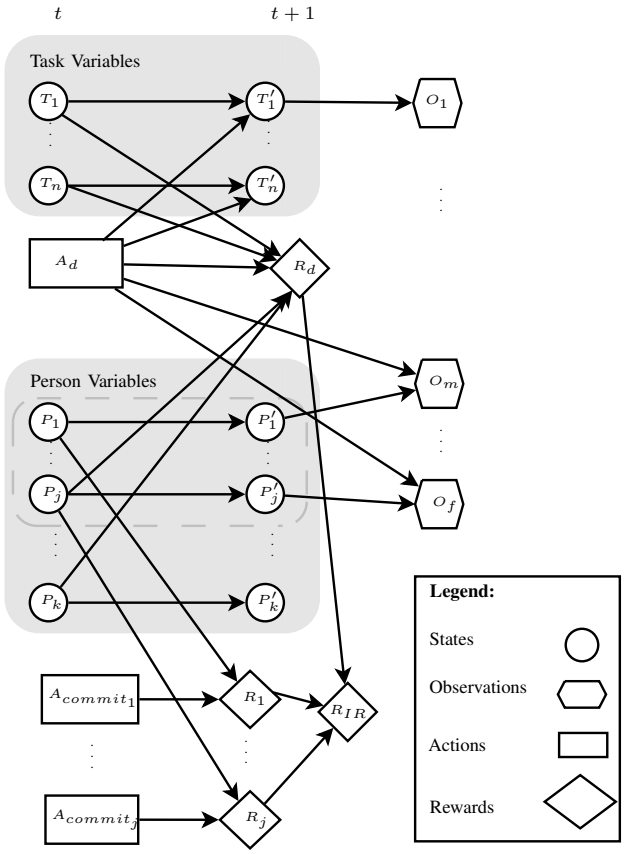


Fig. 1. DBN representation of the proposed DT model.

B. Observations and Observation Model

In a social HRI setting, observations reflect the user's behavior. This behavior is used to monitor the progress of the task and infer the user's affective and motivational status.

Observations are discrete, symbolic values, classified from sensory data, which correspond to features of the environment that are observable in a given state.

The observation factors are contingent on the sensory capabilities of the robot system. Nevertheless, the correct understanding of the user's status relies on the agent being capable of recognizing human communication methods. Consequently, the robot system ought to be able to recognize speech and gestures in order to understand the human user's affective and motivational status.

The observation model is of key importance in the achievement of the information gain goals of the agent. It reflects the probability of receiving a certain observation, given the state of the environment and the action performed. Certain actions, such as questioning or approaching the user, increase the probability of perceiving certain observations. This fact is of utter importance to actively gain information on the user's status. The dependency on the action is represented in observations O_m to O_f in Figure 1.

C. Actions

The model of Figure 1 comprehends two types of actions: A_d and A_{commit}. The first have an effect on the environment and is dependent on the actuators of the agent, while the latter are used for the information gain goals of the agent.

Typically, the action domain A_d contains the minimum set of functionalities which allow the agent to complete its tasks. Social robots, however, need to communicate in a natural, easily understandable way with the human users. To achieve this objective, the robot must be able to express different moods and emotions. Consequently, the action domain A_d of a social robot ought to include speech and/or gestural capabilities and/or graphical emotion displays.

Following the POMDP-IR framework, besides the domain-level action factor A_d, the model has additional action factors A_{commit} for each state factor of interest. The state factors of interest, in the problem under study, are included in the *person* variables, as these contain the aforementioned affective and motivational state of the human user. The actions A_{commit} allow rewarding the agent for decreasing the uncertainty regarding particular features of the environment.

D. Reward Model

Generally, there is no definitive criteria to define the reward model, as rewards are defined over the abstract states and actions of the DT model. Therefore, a policy with satisfactory practical quality is usually obtained through a process of trial and error, where different reward models are used.

In the DT model of Figure 1, rewards are either associated with task objectives: R_d, or with the information gain goals: $R_i, i = 1, \ldots, j$. The sum of these rewards, R_{IR}, constitute the reward awarded to the agent at each time step.

The behavior of the robot consists of the sequence of domain actions A_d the agent performs. In the social HRI scenario, and in order to adapt the robot's behavior to the user's affective and motivational status, the reward assigned to an action depends not only on the *task* variables but also on the *person* variables.

The information rewards R_i influence the behavior of the agent, with the purpose of achieving a low uncertainty regarding certain *person* variables. The value of these rewards are dependent on the threshold of knowledge required, according to the POMDP-IR framework.

E. Estimation of the Stochastic Models

According to the POMDP framework, the model of Figure 1 requires the definition of the transition T and observation O functions. To obtain these functions, the problem designer needs to estimate the respective probability distributions.

One way to estimate the transition and observation models of a POMDP is by collecting experimental data. In this situation, the problem is similar to estimating the structure of a Hidden Markov Model (HMM), and the problem designer might use the Baum-Welch algorithm [11]. However, in the social HRI scenario, learning the model structure from data

might prove a difficult task, due to the lack of well labeled data.

Another common method of modeling the stochastic environment is to simulate the physical system. This approach allows to collect a large number of transition/observation samples and to simplify the estimation problem, since the exact state is accessible. Nevertheless, robotic simulators are not capable of simulating humans and their stochastic behavior as of yet.

Finally, the probability distributions can reflect common knowledge on the problem under study. In these cases, the models can be approximately estimated by means of the expertise of the problem designer, resulting, nevertheless, in policies with good practical quality.

IV. Selected Application

The proposed approach was tested in a case study which considers a socially assistive task: rehabilitation therapy.

A. Scenario

Rehabilitation therapy includes passive or active exercises. In the first, the therapist (human or robot) physically assists the patient to move the affected limb. On the other hand, in active exercises, the patient moves the affected limb by him/herself, while the therapist has the functions of coaching and motivating.

Up to date research in rehabilitation robotics covers mainly passive exercises. Nevertheless, social robots provide a way to approach active rehabilitation exercises, representing an innovative way to monitor, motivate and coach patients.

Overall, the goals of the robot therapist in the considered rehabilitation scenario are:

- To help the user in the given setting, by monitoring the patient's movements (e.g., encourages the patient to continue if he/she stops performing the exercise);
- To adapt its behavior and, consequently, the therapy style (e.g., nurture or challenge the patient), in accordance with the patient's affective and motivational status.

B. Decision-Theoretic Model for the Robot Therapist

The application of the proposed framework to the robot therapist scenario results in the DT model represented in Figure 2.

1) States: The significant features of the environment in which the robot is to operate are related to the human user. The fulfillment of the task's objectives require that the agent keeps track of the user's movements, possesses knowledge regarding relevant personal traits of the user and infers his/hers affective status. Therefore, the proposed DT model considers the state space represented, in factored form, in Table I.

The user's movement is encoded in the *task* state factor *Exercise* ($Exer.$). When the exercise is performed as prescribed, the state factor assumes the value *correct*: $Exer. = Correct$. Otherwise, if the movement is inappropriately or not performed, $Exer. = Incorrect$. The state factor *Personality* ($Pers.$) is a constant *person* variable, known beforehand by the problem designer, which represents the patient's behavioral personality, as *Introverted* or *Extroverted*. Finally, the *Fatigue* state factor ($Fat.$) is a measure of the patient's weariness, caused by the physical exercise. It assumes the values *Tired* or *Energized* whether the patient shows signs of fatigue or liveliness, respectively.

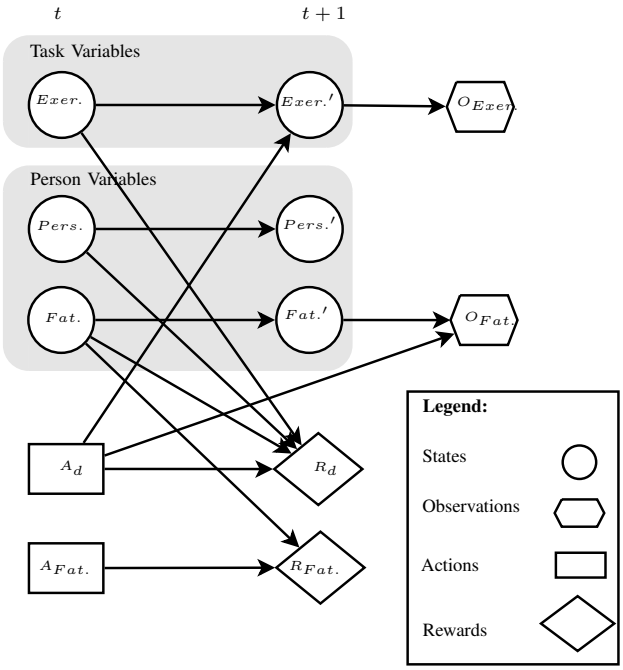


Fig. 2. DBN representation of the DT model for the robot therapist.

2) Observations: The observation space is represented, in factored form, in Table I. Observations reflect the relevant behavior of the patient, in accordance with the task's goals. In the present case study, the agent ought to evaluate the movement performed by the patient and to infer his/hers affective status.

The gesture-related observation factor $O_{Exer.}$ is used to evaluate the exercise and assumes, as a result, the values *Proper* or *Wrong*. $O_{Exer.} = Proper$ whenever the agent perceives that the patient performed the movement as prescribed. Otherwise, $O_{Exer.} = Wrong$ if the agent perceives that the patient did not perform the movement or performed it incorrectly.

The observation factor $O_{Fat.}$, which is related to the affective status of the patient represented in state factor *Fatigue*, assumes the values *Weary*, *Energetic* or *None*. $O_{Fat.} = Weary$ or $O_{Fat.} = Energetic$ when the patient demonstrates feeling tired or lively, respectively. Otherwise, $O_{Fat.} = None$ if the agent does not perceive any relevant information regarding the affective status of the patient.

In this case study, $O_{Exer.}$ is obtained by visual classification of the patient's gestures and $O_{Fat.}$ through speech interaction, i.e., through classification of the user's verbal responses.

3) Actions: The proposed DT model considers two action factors: the *Action Domain* A_d and the *IR Action* $A_{Fat.}$. At each time step, the agent chooses one value for each

TABLE I
STATE, OBSERVATION AND ACTION SPACES FOR THE ROBOT THERAPIST CASE STUDY

	Factors	Values
States	$Exer.$	Correct, Incorrect
	$Pers.$	Introverted, Extroverted
	$Fat.$	Tired, Energized
Observations	$O_{Exer.}$	Proper, Wrong
	$O_{Fat.}$	Weary, Energetic, None
Actions	A_d	Nurture, Challenge, Query Patient, End Therapy, None
	$A_{Fat.}$	Commit Tired, Commit Energized, Null

action factor. The possible values for the action factors are represented in Table I.

The IR action is defined according to the POMDP-IR framework, with a *commit* action for each value of the related state factor ($Fat.$), and a *null* action. $A_{Fat.}$ allows rewarding the agent for reducing the uncertainty regarding the state factor $Fat.$, related to the patient's fatigue.

The *Action Domain* A_d contains the set of functionalities which allow the agent to achieve its goals, which are, in this case study, to monitor and motivate the patient in an active physical rehabilitation exercise. That is, whenever the patient stops performing the exercise or performs it incorrectly, the robot encourages him/her to proceed with the exercise.

The therapy style, i.e., the robot's approach to the patient changes as a function of his/hers *Fatigue* and *Personality*. Dependent on these factors, the encouragement is classified as $Nurture$ or $Challenge$ whether the agent opts for a softer (e.g. "You are doing great! Keep on the good work.") or a more defiant approach (e.g., "You can do better than that!").

Since the therapy style is dependent on the *person* variables, it is important to gain information and maintain a low uncertainty regarding the state factors $Pers.$ and $Fat.$. As $Pers.$ is constant, the agent only actively seeks to reduce uncertainty on the state factor $Fat.$, through the $Query\ Patient$ action. This action consists of verbally interacting with the patient to infer his/hers *Fatigue*.

Moreover, the agent ought to end the exercise ($End\ Therapy$) when the patient persistently shows he/she is not able to proceed with it. Finally, at each time step, the agent might choose to do nothing ($None$).

Besides adapting speech in conformance with the behavior of the patient, the robot also modifies the emotion displayed to the more appropriate in the given situation.

4) Transition, Observation and Reward Functions: The proposed framework allows to take into account the effects of time in the states of the DT model. Namely, in the current case study, the transition function T encodes that $b(Fat.) = Tired$ increases at each time step in the absence of opposing observations ($O_{Fat.} = Energetic$). That is, the agent realistically believes that the patient is feeling more tired over time. Also, the transition function of this case study dictates that the probability of the patient correctly performing the exercise ($Exer. = Correct$) increases with the motivation actions ($Nurture$ or $Challenge$).

The observation function O encodes the error in sensory data classification. This means, for instance, that even if the patient's gesture is classified as incorrect ($O_{Exer.} = Wrong$), the agent's belief on $Exer. = Incorrect$ is not 100% and the robot might require more information before motivating the patient. Furthermore, the probabilities in O take into account that information-gathering actions (such as $Query\ Patient$) increase the probability of perceiving a verbal response from the user (e.g., $O_{Fat.} = Weary$).

The time step of the synchronous decision-making loop (i.e., the time that elapses between decision episodes), needs to consider the rate of classification of sensory data. The agent has multiple sensors and respective classification systems, operating at different frequencies. Consequently, the decision-making loop rate needs to be equal or lower than the lowest sensory operating frequency. In the present case-study, the time step of the decision-making loop is 5 seconds.

The DT model of Figure 2 rewards IR actions ($R_{Fat.}$) and A_d actions (R_d). The information rewards are defined, in accordance with the POMDP-IR framework, so that the agent actively seeks to have a belief on $Fat. = Tired$ or $Fat. = Energized$ greater than 75% (i.e., $\beta = 0.75$ in accordance with the POMDP-IR framework).

Actions in A_d are rewarded in accordance with the state of the environment: *Encouragement* actions ($Nurture$ and $Challenge$) are rewarded 0.2 whenever the patient is incorrectly performing the exercise or 0.1 when he/she shows signs of feeling tired, and penalized -0.1 otherwise. The reward given to each action also depends on the state factor $Pers.$: for an *Introverted* person, the $Nurture$ action is preferred while the $Challenge$ action is favored for an *Extroverted* person; The $Query\ Patient$ action is penalized with -0.2; $None$ is not rewarded nor penalized; $End\,Therapy$ receives high penalization (-1) when the patient feels energetic and a reward of 0.1 otherwise. The discount factor in this case study is 0.9.

V. EXPERIMENTS

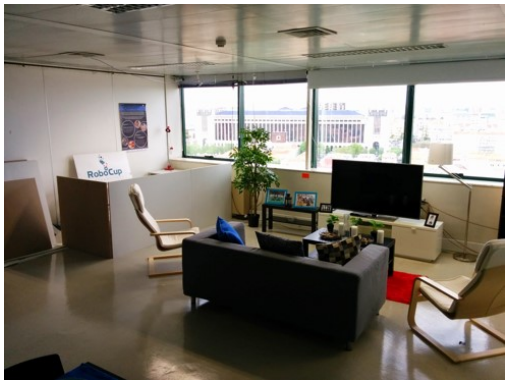
The robot therapist case study was implemented in a real mobile social robot and interacted, in different experiments, with several persons, in a realistic apartment testbed.

A. Experimental Setup

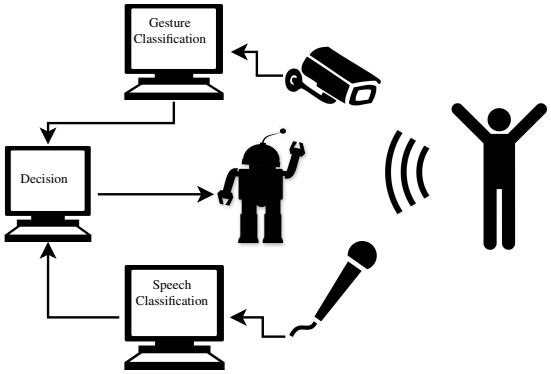
The networked robot system used in the present case study consists of the MOnarCH robot platform, represented in Figure 3(a) and an external Kinect camera. The robot platform provides the actuating capabilities required to implement the domain actions A_d and the sensors necessary for the speech related observations $O_{Fat.}$. The Kinect camera is strategically located for a clear view of the patient's movements and is used, therefore, for the classification of the exercise $O_{Exer.}$.

The sensory information is, after classification, used as input to the decision system that controls the actuators of the robot platform.

(a) Robot Platform used in the experiments.



(b) Living room area of the ISRoboNet@Home testbed.



(c) Components of the experimental setup: Decision System, Gesture & Speech Classification, Robot and sensors. Arrows represent directions of communication.

Fig. 3. Experimental setup for the robot therapist case study.

The experiments within this case study took place in the ISRobotNet@Home Testbed[1], which is represented in Figure 3(b). This testbed provides the infrastructure to implement networked robot systems in a domestic environment.

In accordance with the observation space of the DT model, the patient's movement is to be classified as *Correct* or *Incorrect*, at each time step. Likewise, in order to infer the patient's affective status, and as the preferred means of communication is through verbal interaction, the patient's speech is classified as: Demonstrative of the patient feeling $Weary$ or $Energetic$; $None$ if it does not add relevant information.

1) Gesture Classification: Classification of the patient's movement is achieved resorting to a Kinect-based application, which makes use of a gesture database previously built through the Visual Gesture Builder (VGB)[2] tool, available in the *Kinect for Windows* Software Development Kit (SDK). First, within VGB, the system designer tags frames in recorded video clips, which are related to meaningful gestures. These tagged frames are, then, used as inputs to the detection algorithm during the training stage. On the application runtime, the detection technologies detect discrete and continuous gestures. Discrete gesture classification outputs a Boolean indicating if the user is performing a trained gesture and a confidence level on the Boolean classification. Continuous gesture classification results in a float indicating the progress of the user as he/she performs the gesture.

2) Speech Classification: Automatic Speech Recognition (ASR) is based on VoCon Hybrid[3], a state of the art commercial solution. This speech recognition engine is based on context-free grammars, written in Backus-Naur Form (BNF), which encode the utterances to be recognized. The grammars are created with prior knowledge of the scenarios that the robot needs to understand. Speech understanding follows the definition of a corpus over the context-free grammars, which spawns the possible sentences the ASR recognizes.

3) Decision System: The decision system used in the present case study is based on the Symbolic Perseus solver [12]. Symbolic Perseus uses the *PERSEUS* algorithm [9], and Algebraic Decision Diagrams (ADD) [13] as the underlying data structure.

The policy is computed offline, in order to save computational resources during the online execution of the task. The online computation, in the decision system, consists on the belief update and the selection of the action in accordance with the updated belief and the previously calculated policy.

B. Experimental Results

The experiments performed in this case study intend to prove that the proposed DT model is able to perform a given task with different persons, while taking into account the *person* (latent) variables to adapt the agent's behavior.

Each experiment considers a different user, which is classified according to his/hers personality (i.e., as introverted or extroverted), and with regard to his/hers ability to perform the exercise (athletic or unfit).

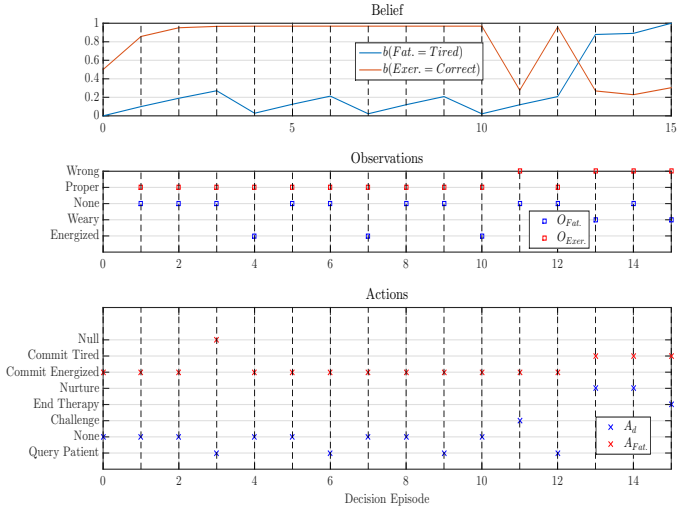
The experiments carried out within this work were recorded and are available at https://www.youtube.com/playlist?list=PLp1xxEiDsjtBcYGJuoiwBytoK5ZdMDvIv.

Figure 4 represents the data obtained in the experiments, namely the observations, actions and belief on the two key state factors considered: $Fat.$ and $Exer.$. Figure 5 represents episodes of the different experiments where the robot interacts with the user.

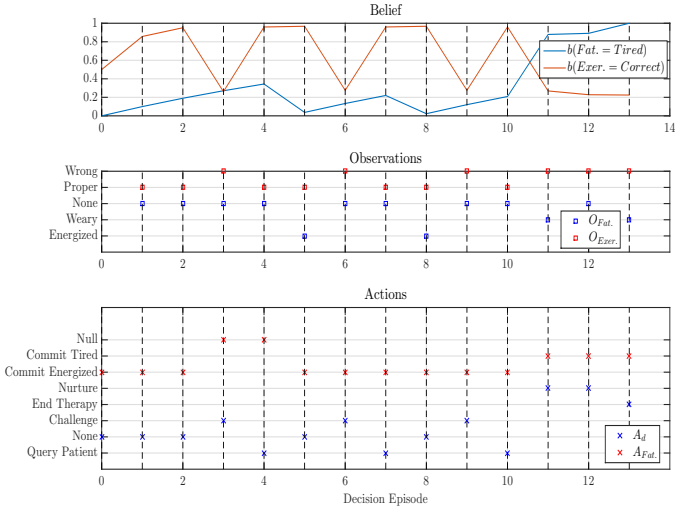
1) Experiment A: This experiment considers a user which is classified as extroverted ($Pers. = Extroverted$) and athletic. The user feels energetic for the first fifty seconds (decision step 10), approximately, and tired afterwards.

At the beginning, the robot chooses not to act, since the exercise is well performed and the agent has a low uncertainty regarding the *fatigue* status of the user. This uncertainty on the state factor $Fat.$, however, increases over time, driving

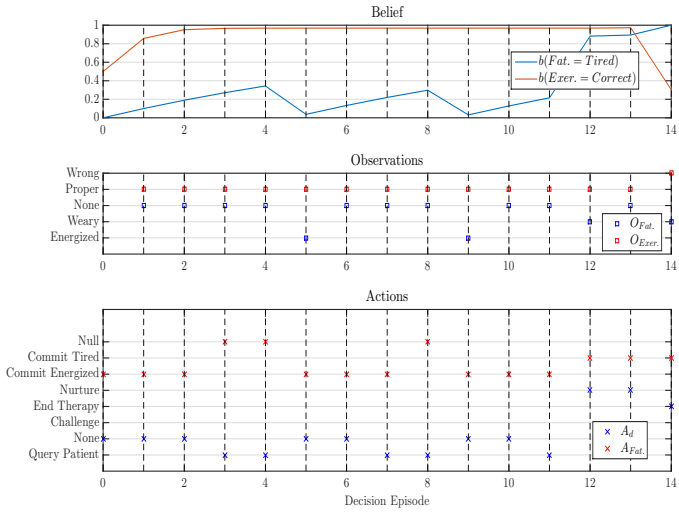
[1] http://welcome.isr.tecnico.ulisboa.pt/isrobonet/
[2] https://msdn.microsoft.com/en-us/library/dn785529.aspx
[3] http://www.nuance.com/for-business/speech-recognition-solutions/vocon-hybrid/index.htm



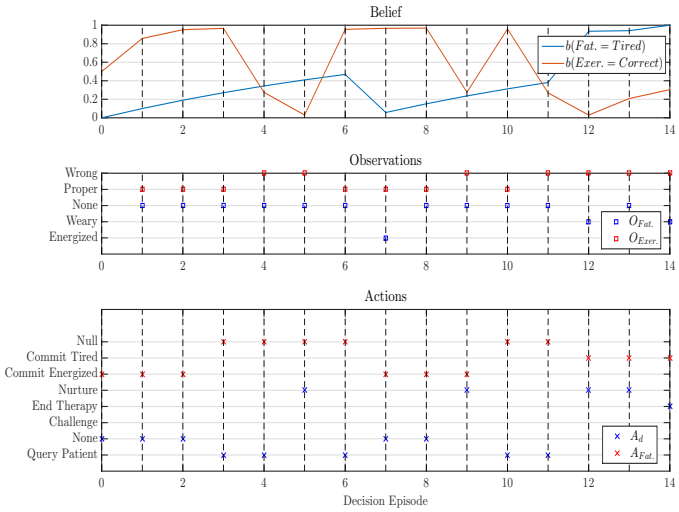
(a) Experiment A: Evolution of the Belief on the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed.



(b) Experiment B: Evolution of the Belief on the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed.



(c) Experiment C: Evolution of the Belief on the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed.



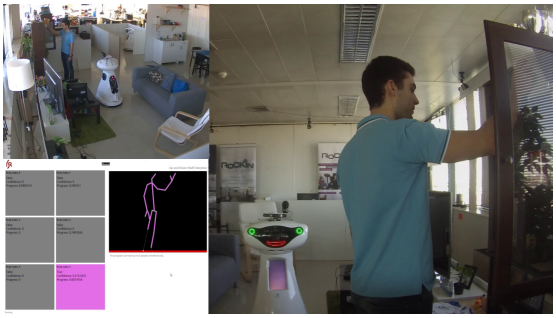
(d) Experiment D: Evolution of the Belief on the states $Fat.$ and $Exer.$ w.r.t. the decision episode, the observations received and the actions performed.

Fig. 4. Data obtained in the experiments

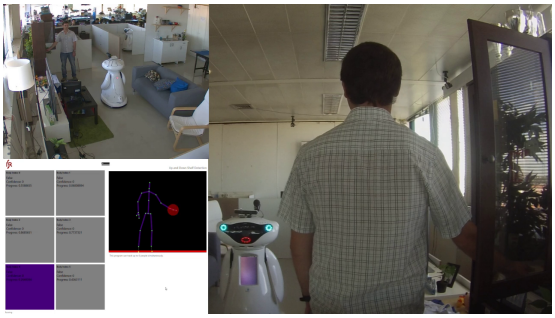
the robot to actively seek to reduce it, by querying the user (decision step 3). The answer ($O_{Fat.} = Energetic$), informs the robot that the user is still active and motivated, increasing the certainty on $Fat. = Energized$. This behavior is repeated until the user does not perform correctly the exercise ($O_{Exer.} = Incorrect$) in decision step 11. Then, the robot motivates the person through a challenging approach due to the considered *personality* of the user and the current *fatigue* status. Following these events, the agent's uncertainty on the state factor $Fat.$ increased and the robot queries the user in decision step 12. After receiving information that the user now feels tired ($O_{Fat.} = Weary$), the robot changes therapy style and adopts a nurturing approach. As the user continuously shows not being able to carry out the exercise and the certainty on $Fat. = Tired$ increases, the robot finally chooses to end the therapy in decision step 15.

2) Experiment B: This experiment considers a user classified as extroverted ($Pers. = Extroverted$) and unfit. The user feels energetic for the first forty seconds, approximately, and tired afterwards.

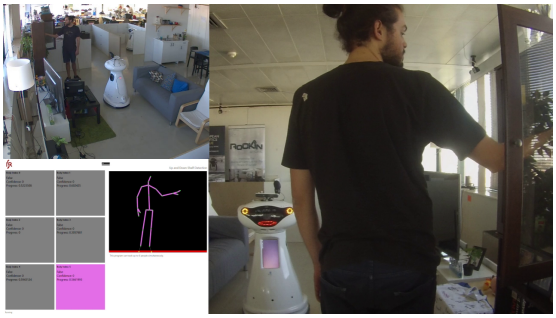
The behavior of the robot is similar to the previous experiment while the user shows feeling energetic and correctly performs the exercise. Nonetheless, the user incorrectly performs the exercise more often, at which occasions the robot acts in motivating with a challenging approach, while the agent believes the user feels motivated/energetic. Even though motivating the user, the robot keeps track of his/hers *fatigue* and reacts when the uncertainty on $Fat.$ is high, i.e., $b(Fat. = Tired) < 0.75$ and $b(Fat. = Energized) < 0.75$. Finally, the agent ends the therapy once it persistently observes the user is not performing the exercise and feels tired.



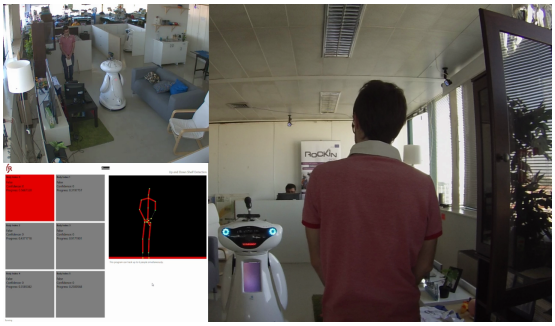
(a) Scene of experiment A: Robot challenges the user to proceed with the exercise.



(b) Scene of experiment B: Robot queries the user with regard to his status.



(c) Scene of experiment C: Robot motivates the user through a nurturing approach.



(d) Scene of experiment D: Robot ends the therapy.

Fig. 5. Episodes of the experiments where the robot interacts with the user. In each figure: Right and top left images show different views of the ISRobotNet@Home Testbed; Bottom left image represents the interface of the gesture classification application.

3) Experiment C: This experiment considers a user classified as introverted ($Pers. = Introverted$) and athletic. The patient feels energetic up to, approximately, 45 seconds (decision step 9), and tired afterwards.

The behavior of the robot is heavily dependent on its knowledge regarding the fatigue status of the user. While the uncertainty on the $Fat.$ state factor is high, the robot queries the user. Since the uncertainty on $Fat.$ increases over time, the agent performs the action $Query\,Patient$ until it perceives an answer $O_{Fat} = Energetic$ or $O_{Fat} = Weary$ (decision steps 3 & 4 / 7 & 8). Nevertheless, the robot performs the therapy task while actively gathering information on the environment, motivating the user once the belief on $b(Fat. = Tired)$ is high, and ending the therapy appropriately.

4) Experiment D: This experiment considers a user which is classified as introverted ($Pers. = Introverted$) and unfit. The user feels energetic for the first 40 seconds (decision step 8), approximately, and tired onward.

The behavior of the robot changes in accordance with its belief on the states of the environment. In the present experiment, there is a "trade-off" between motivating or querying the user depending on the belief over the state factors $Fat.$ and $Exer.$. In decision step 3, the agent queries the agent due to the high uncertainty on $Fat.$. Afterwards, the agent perceives no answer but observes the user incorrectly performed the movement. This observation does not translate, however, into an absolute certainty on the exercise having been incorrectly performed ($b_4(Exer. = Correct) \approx 0.3$), since the DT framework takes into account sensor related noise. The agent, then, queries the user once again (decision step 4), due to the increasing uncertainty on the *fatigue* of the user. Once again, the NRS receives no answer ($O_{Fat.} = None$), and observes the user incorrectly performed the movement. This time, the agent's belief on $Exer. = Incorrect$ is higher ($b_5(Exer. = Incorrect) \approx 0.95$) and it motivated the user. Nevertheless, the uncertainty on $Fat.$ is still high on decision step 6 and the robot once again queries the user, perceiving this time an answer.

During the rest of the experiment, the robot once again queries the user when the uncertainty on $Fat.$ is high (decision steps 10 & 11) and motivates the user in accordance with the beliefs on the variables $Exer.$ and $Fat.$ (decision steps 9, 12 & 13). Finally, the agent ends the therapy in decision step 14.

C. Discussion

Table II details the behavior of the robot for each experiment. As expected: the number of motivation actions is higher for the users classified as unfit, which incorrectly perform the exercise more often than the athletic users; and the number of query actions is higher for the users classified as introverted.

The robot detected the fatigue status change from $Energized$ to $Tired$ in all the experiments, taking between, approximately, 15 seconds (experiments 1, 2 & 3) to 20 seconds (experiment 4), to have a high belief on $Fat. = Tired$

TABLE II
BEHAVIOR OF THE ROBOT WITH REGARD TO THE EXPERIMENT

	Motivation actions	Query actions	Time elapsed until agent detected change of user's status	Time elapsed until agent ends therapy since it detected user is tired	Duration of the experiment
Experiment A	3	4	15 s	10 s	75 s
Experiment B	5	3	15 s	10 s	65 s
Experiment C	2	5	15 s	10 s	70 s
Experiment D	4	5	20 s	10 s	70 s

($b(Fat. = Tired) > 0.8$) from the time the user started to feel tired. Moreover, the agent motivated the user upon detection of faulty movements, either immediately after observing $O_{Exer.} = Wrong$ (experiments 1, 2 and 3) or after two consecutive observations (experiment 4) depending on the belief over the state factors $Fat.$ and $Exer.$. Finally, the agent ended the therapy when consistently observing the user was not capable of proceeding with the exercise, after 10 seconds, approximately, of having a high certainty on the user feeling tired.

Overall, the DT approach to planning in the robot therapist resulted in a behavior capable of achieving the task and information goals, adaptive to the user's status and socially appealing.

VI. CONCLUSION

Building on the POMDP-IR framework, this work introduced a DT approach to planning in social HRI. Under the proposed framework, the agent is capable of achieving its task and information goals while following a socially appealing behavior. Moreover, the agent adapts its behavior in accordance with the affective and motivational status of the user. The properties of the DT framework were demonstrated in the robot therapist case study. The experiments' results prove the validity of the proposed framework for problems involving robot systems in HRI scenarios.

In future work and to further validate the framework developed within this work, further experiments ought to be performed, in particular considering cases of social assistance with real patients. This would involve the implementation of a DT-based NRS in a real therapy scenario, in order to evaluate the behavior of the robot system as to what concerns the accomplishment of its goals, its social qualities and adaptability. Moreover, the proposed framework ought to be tested in a scenario which considers: more latent variables, in particular hidden variables of interest; more information-gathering actions; and more complex actions, e.g., manipulation.

Solution methods for MDP-based models, such as the framework proposed in this work, present an important practical issue since they assume complete knowledge of the stochastic models (Transition and Observation models). Besides, any change to the parameters of these models imply a recalculation of the DT policy. On the other hand, Reinforcement Learning (RL) approaches [14] assume either absent or imperfect knowledge on the environment dynamics. The DT policies are learned, in this case, from the interaction of robotic agents with their environment. RL methods can support on the structure and properties of the proposed model to overcome the aforementioned implementation issues.

REFERENCES

[1] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[2] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1, pp. 99 – 134, 1998.

[3] M. T. J. Spaan, T. S. Veiga, and P. U. Lima, "Decision-theoretic planning under uncertainty with information rewards for active cooperative perception," *Autonomous Agents and Multi-Agent Systems*, vol. 29, no. 6, pp. 1157–1185, 2015.

[4] T. Taha, J. V. Miro, and G. Dissanayake, "Pomdp-based long-term user intention prediction for wheelchair navigation," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, May 2008, pp. 3920–3925.

[5] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Special issue on Socially Interactive Robots, Robotics and Autonomous Systems*, vol. 42, no. 3 - 4, pp. 271 – 281, 2003.

[6] J. Hoey, P. Poupart, A. v. Bertoldi, T. Craig, C. Boutilier, and A. Mihailidis, "Automated Handwashing Assistance for Persons with Dementia Using Video and a Partially Observable Markov Decision Process," *Computer Vision and Image Understanding*, vol. 114, no. 5, pp. 503–519, May 2010.

[7] G. E. Monahan, "A survey of Partially Observable Markov Decision Processes: Theory, models, and algorithms," *Management Science*, vol. 28, no. 1, pp. 1–16, January 1982.

[8] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for pomdps," in *International Joint Conference on Artificial Intelligence (IJCAI)*, August 2003, pp. 1025 – 1032.

[9] M. T. J. Spaan and N. Vlassis, "Perseus: Randomized point-based value iteration for pomdps," *J. Artif. Int. Res.*, vol. 24, no. 1, pp. 195–220, Aug. 2005.

[10] C. Boutilier and D. Poole, "Computing optimal policies for partially observable decision processes using compact representations," in *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*, ser. AAAI'96, 1996, pp. 1168–1175.

[11] S. Koenig and R. Simmons, "Xavier: A robot navigation architecture based on partially observable markov decision process models," in *Artificial Intelligence Based Mobile Robotics: Case Studies of Successful Robot Systems*, R. B. D. Kortenkamp and R. Murphy, Eds. MIT Press, 1998, pp. 91 – 122.

[12] P. Poupart, "Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes," Ph.D. dissertation, University of Toronto, Toronto, Ont., Canada, 2005.

[13] J. Hoey, R. St-aubin, A. Hu, and C. Boutilier, "Spudd: Stochastic planning using decision diagrams," in *In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 1999, pp. 279–288.

[14] T. Jaakkola, S. P. Singh, and M. I. Jordan, "Reinforcement learning algorithm for partially observable markov decision problems," in *Advances in Neural Information Processing Systems 7*. MIT Press, 1995, pp. 345–352.