



Human Intent Prediction Using Markov Decision Processes

Catharine L. R. McGhan,* Ali Nasir,† and Ella M. Atkins‡
 University of Michigan, Ann Arbor, Michigan 48109

DOI: 10.2514/1.1010090

This paper describes a modeling method for predicting a human’s task-level intent through the use of Markov Decision Processes. Intent prediction can be used by a robot to improve decision-making when human and robot operate in a shared physical space. This work presumes human and robot goals are independent such that the robot seeks to avoid interfering with the human rather than directly assisting the human. The proposed human intent prediction system transforms goal sequences the human is expected to complete, a limited past action history, and a correlation of observed behaviors with actions into a prediction of the in-progress or next action the humans is most likely to take. An intra-vehicle activity space robotics application example is presented.

Nomenclature

A	=	$\{1, 2, \dots, n_a\}$, set of actions for the Markov Decision Process
A^i	=	$\{a_1^i, a_2^i, \dots, a_{n_a}^i\}$, $a_k^i \in A$, history of recently executed actions in state i
F^i	=	$\{f_1^i, f_2^i, \dots, f_{n_f}^i\}$, set of binary flags for high-priority interruptive goal states (on/off) in state i
G^i	=	$\{g_1^i, g_2^i, \dots, g_{n_g}^i\}$, set of binary flags indicating goal status (complete/incomplete) in state i
p_{kx}	=	$p(g_x^i = 0 g_x^i = 0, A^i, a_k)$, probability of no change for an incomplete mission goal g_x^i due to a completing action a_k in state s^i with action history A^i
$R(s^i, a_k)$	=	reward function for performing action a_k at state s^i ; a_k is an optional term
s^i	=	$\{G^i, A^i, F^i\}$, a single state of the Markov Decision Process
S	=	$\{s^1, s^2, \dots, s^{n_s}\}$, set of Markov Decision Process states
$T(s^i, a_k, s^j)$	=	$p(s^j s^i, a_k)$, probability of transitioning from state s^i to s^j by executing action a_k
α_i, β_i, k	=	known reward weights and constants for all j
λ_{kx}	=	transition probability weight for impact of action a_k on mission goal g_x^i
γ	=	Bellman equation discount factor
$(1 - p_{kx})$	=	$p(g_x^i = 1 g_x^i = 0, A^i, a_k)$, the probability of goal objective g_x^i becoming one (completed) due to a completing action a_k in state s^i with action history A^i and goal g_x^i

I. Introduction

FUTURE intelligent space robotic systems will support autonomous human–robot interaction missions nonorbit and on planetary surfaces. This paper studies automation capable of predicting a human astronaut’s intent without explicit communication during physically proximal human–robot operations. This situation would exist when a human and robot occupy a shared workspace [e.g., an International Space Station (ISS) habitation module] and are assigned distinct sequences of task-level goals. In this scenario, a persistent objective of the robot is to use predicted intent to minimize the probability that the robot’s goal-seeking actions will interfere with the human’s work rather than directly assist him or her with their work. To predict human intent, the robot must first identify the human in the workspace, observe the human over time, and then make informed decisions about its own motion. In this work, human intent prediction (HIP) is performed using knowledge of the astronaut’s goals as well as observed physical states. The robot seeks to optimize its goal-seeking behaviors while avoiding conflict with its human companion. The outcome assumed in this work is that the human is not impacted by the robot, and thus can completely ignore the robot while working. Because models and observational data are uncertain, a Markov Decision Process (MDP) framework is adopted.

Significant prior research has been conducted in human action recognition. Hidden Markov models (HMMs) can be used to identify gestures in real time [1]. HMM outputs can be augmented with regression [2] and classification methods [3]. Karami et al. [4,5] discussed the use of partially observable Markov Decision Processes (POMDPs) for robot action selection. Karami et al. showed the results for a case study where the human had two possible mission tasks in [4], although they used the POMDP approximations to predict human behavior within a combined robot decision-making POMDP and compared these decisions with baseline model outputs in [5]. POMDPs have also been proposed [6] to compute the strongest beliefs of human preferences using active communication to enable explicit requests for model updates. Although HMM and POMDP methods have become popular for robot decision making due to the inclusion of hidden or belief states, there are drawbacks. Including uncertain human state data in a combined model of human intent prediction and robot action choice explodes the size of the state space, even relative to the baseline MDP that assumes an observable state. Approximate POMDP solution methods have been shown to be feasible for real-time use in simple robotics applications [7], but these have only been shown to work for state-space sizes of up to $\sim 10^3$. It is unclear whether use of the belief state will, in practice, improve results, particularly for highly structured environments such as would exist given a highly trained astronaut sharing a workspace with a robot in the International Space Station. Lasota et al. [8] used MDPs for human action (intent) prediction also but focuses on convergence on a collaboration strategy and required both the human and robot be able to commit the same actions for goal completion. Conversely, we use the simplification that tasks are independent to allow for a larger set of use cases due to improved scalability and responsiveness.

Presented as Paper 2012 at the AIAA Infotech@Aerospace 2012, Garden Grove, CA, 19–21 June 2012; received 11 December 2012; revision received 11 August 2014; accepted for publication 9 November 2014; published online 18 May 2015. Copyright © 2014 by Catharine L. R. McGhan. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. Copies of this paper may be made for personal or internal use, on condition that the copier pay the \$10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 2327-3097/15 and \$10.00 in correspondence with the CCC.

*Ph.D. Candidate, Aerospace Engineering Dept., Autonomous Aerospace Systems Lab. Student Member AIAA.

†Ph.D. Fellow, Aerospace Engineering Dept., Autonomous Aerospace Systems Lab. Member AIAA.

‡Associate Professor, Aerospace Engineering Dept., Autonomous Aerospace Systems Lab. Associate Fellow AIAA.

This paper describes the formulation of a predictive human intent model meant for use within a decision-making framework that uses linked but separate MDPs to guide robot decision making in a shared workspace. Robot decision making occurs in two separate stages: human intent prediction and robot action choice (RAC). The results of HIP are fed into RAC as if they were deterministic observations, enabling the decomposition of these two MDPs. With the MDP formulation, we assume the human's actions are fully observable. The tradeoff is that, although our MDPs have lower complexity than a POMDP formulation, we assume the next action selected by a human intent prediction MDP is observable. To simplify decision making, we assume the HIP MDP does not require a model of the robot, because we have also assumed that the robot is working unobtrusively on independent tasks, and thus does not impact human decision making. This paper subsequently focuses on the HIP stage of decision making, reserving the specifics of RAC for future work.

To support full observability, we assume a closed-world environment and sufficient robot memory to store a partial action history. We also encode a model of human preferences within the reward function of the HIP MDP. We hypothesize that a robot can predict companion intent by first identifying actions based on observations rather than relying on explicit communication, and then recognizing those observed actions as part of a finite sequence. Selected MDP policy actions represent the human's intent given the assumption that the MDP model and costs adequately match those of the human. To predict the next action, an updated human state is sensed from the environment, and that human's goal objectives and previous actions are given to the MDP model policy. Use of the MDP requires the Markov assumption: that is, independence of future state from past state. However, an abbreviated state history captured within the current state is used when knowledge of some past history is required. Environmental information is included in the form of sensed events that are folded into the human's goal/objectives before inclusion in the MDP state vector. The human intent prediction MDP formulation is described next, followed by an intervehicle activity (IVA) ISS case study in which a robot and astronaut work independently in a shared habitation module workspace.

II. Human Intent Prediction Model

For our architecture, a human intent prediction Markov Decision Process is constructed to enable translation of the robot's observations, including observations of its human companion, into predictions of the next action the human will accomplish. A robot action choice MDP will subsequently use this information to select the appropriate action for the robot to take.

A Markov Decision Process is defined as [9,10]

$$\text{MDP} = \{S, A, T(s^i, a_k, s^j), R(s^i, a_k)\} \rightarrow \pi(s^i) \quad (1)$$

where S denotes the set of states s^i ; A denotes the set of available (or allowable) actions a_k ; and $R(s^i, a_k)$ is the set of state-dependent rewards for performing action a_k at state s^i . $T(s^i, a_k, s^j) = p(s^j | s^i, a_k)$ is the set of all transition probabilities, with the probability of a state s^j occurring as an outcome of action a_k executed from state s^i . We assume that the optimal policy is time invariant.

We solve the MDP as an infinite horizon problem using a Gauss–Seidel value iteration [9]. We set the Bellman equation discount factor γ to 0.95 and the convergence bound ϵ to 10^{-5} .

The HIP MDP is formulated as

$$\begin{aligned} S &= \{s^1, s^2, \dots, s^{n_s}\} \\ s^i &= \{G^i, A^i, F^i\} \\ G^i &= \{g_1^i, g_2^i, \dots, g_{n_g}^i\}, & g_k^i &\in \{0, 1\}, & k &= \{1, \dots, n_g\} \\ F^i &= \{f_1^i, f_2^i, \dots, f_{n_f}^i\}, & f_k^i &\in \{0, 1\}, & k &= \{1, \dots, n_f\} \\ A^i &= \{a_1^i, a_2^i, \dots, a_{n_a}^i\}, & a_k^i &\in A, & A &= \{1, 2, \dots, n_a\}, & k &= \{1, \dots, n_h\} \end{aligned} \quad (2)$$

Each state s^i in S includes the human's task-level mission goal states G^i , actions A^i , and interruptive goal states F^i . A^i represents an abbreviated past history of n_h human companion actions. High-priority interruptive goal states F^i are activated by sensed events that can override mission goals. Mission and high-priority goals are binary valued, set to complete (designated as 1) or incomplete (designated as 0). Actions are integer-valued attributes taken from action set A . The total number of possible states is $n_s, n_s = 2^{n_g} * n_h^{n_a} * 2^{n_f}$. For an action history of length $n_h = 4$, our chosen baseline IVA scenario with three task-level goals, one high-priority goal, and five actions has state-space size $n_s = 2^3 * 5^4 * 2^1 = 10,000$. The robot's state is not included, as we assume for simplicity that robot actions do not influence the human directly. The human's goal objectives g_k^i and f_k^i are assumed independent.

We define an action a_k as a task that may require multiple motions and that will complete without interruption. We assume each goal can be completed by a particular action or ordered set of actions, but some goals may have many satisficing n -tuple action sequences that are order invariant. Mapping a sequence of actions to a larger task requires knowledge of the sequence of events. In our model, the human can only attempt to fulfill one high-priority goal objective at a time, but some actions may impact fulfillment of multiple task-level goals. In cases where an action cannot change a goal, corresponding probabilities are set to zero. Action history length n_h in this work is set by the user and allows transitions to be conditioned on a sequence of up to n_h actions.

The MDP selects optimal actions in each state based on transition probability T and reward R . The transition probability map, specified as a tensor or set of action-specific matrices, capitalizes on the fact that, for the astronaut–robot domain, the astronaut will be highly likely to follow step-by-step procedures for many activities conducted on the space station. For transitions that are impossible, we set $p = 0$. For cases where there are known procedures (chains of actions from which we expect little to no deviation), transition probabilities are set to near one for the expected state–action outcomes. State–action combinations with high probabilities will initially be identified from published procedures to follow; these probabilities can be estimated in advance and remain constant or be seeded with large initial values:

$$R(s_i) = \sum_{j=1}^{n_g} \alpha_j r_1(g_j^i) + \sum_{j=1}^{n_f} \beta_j r_2(f_j^i) \quad (3)$$

Reward function $R(s_i)$ is based on fulfillment of recurrent human objectives, providing a straightforward representation of human preference for the MDP. Next, we test reward weights α_j, β_j in the range $[0 \ 1]$. Function r_1 represents the nonnegative reward due to task-goal achievement is represented by r_1 , whereas r_2 represents reward (or cost) of a high-priority goal being achieved (or not achieved). Weights are scaled based on maximum possible values of r_1 and r_2 . If a high-priority goal flag is active but not complete, the value of r_2 is negative. Note that our simple reward

function does not contain costs (e.g., fuel, energy) for action completion of a_k . Such costs could be included, so long as we also include a no-op (“no-operation”) action that is selectable from every state in which other actions are energy consuming.

Because the goals are binary-valued $\{0,1\}$ attributes [see Eq. (2)], we can use the goal flags directly when calculating reward r_1 . Equation (4) shows an example reward function; we assume that $|k| > 1$ to encourage high-priority goal achievement:

$$\begin{aligned} r_1(g_j^i) &= g_j^i \\ r_2(f_j^i) &= \begin{cases} 1 & \text{if } f_j^i = 1 \\ -k & \text{if } f_j^i = 0 \end{cases} \end{aligned} \quad (4)$$

In our HIP models, reward is only given once an objective is met; there is no reward for partial success. If a goal requires multiple actions to accrue in the action history, those interim states receive no reward. Specific action-history content enables factoring action reward into the optimization, as does the expected future value term of the Bellman equation. To encode human preference with respect to goal-driven behaviors, we must address whether each behavior is driven by goals without context or whether context drives human adaptation to the special circumstances of the environment [11]. We define structured context as prior knowledge specific to the environment: the human and robotic agent’s placement and movements within it, according to the goals that must be achieved. Nominally, we encode what we know of the structured context of a goal-specific behavior into MDP transition probabilities. We then are free to use the reward function strictly to encode the relative value of the goals.

III. Case Study

For our case study, we model a simple intervehicle activity scenario including the actions of eating ($a_k^i = 1$), drinking ($a_k^i = 2$), interacting with a computer ($a_k^i = 3$), high-priority button pushing ($a_k^i = 4$), and no op ($a_k^i = 5$). These five tasks are required for astronauts in IVA as well as for humans on Earth. The state has three mission goals of sate hunger (g_1^i), sate thirst (g_2^i), and complete a general mission-oriented work effort (g_3^i), as well as one high-priority goal of button inactive (f_1^i) that indicates that a button needs to be pushed (corresponding to a safety-critical mission task that might need to be completed). We have conducted previous human subject experiments of these tasks with a safe robotic manipulator arm that confirms the feasibility of a shared workspace [12].

Equation (5) gives our domain representation:

$$s^i = \{g_1^i, g_2^i, g_3^i, a_1^i, \dots, a_{n_h}^i, f_1^i\} \quad a_k^i \in \{1, 2, 3, 4, 5\} \quad (5)$$

The actions are integers from 1 to 5, corresponding to the labels given previously. The mission goals and high-priority goal are binary valued, with 0 corresponding to incomplete and 1 corresponding to complete. A 0 value indicates that a high-priority goal needs to be satisfied (a high cost is incurred for remaining in that state), and a 1 indicates that the button is inactive and does not need to be pushed again (a modest reward is offered for remaining in this safe state). The action-history lengths explored in this case study range from $n_h = 0$ to $n_h = 5$:

$$R(s_i) = \begin{cases} \alpha_1 g_1^i + \alpha_2 g_2^i + \alpha_3 g_3^i + \beta_1 & \text{if } f_1^i = 1 \\ \alpha_1 g_1^i + \alpha_2 g_2^i + \alpha_3 g_3^i - k * \beta_1 & \text{if } f_1^i = 0 \end{cases} \quad (6)$$

The preceding reward function for our case study is consistent with the form shown in Eqs. (3, 4). To equally weight all goals and a nonactive high-priority goal, we set all weighting factors to 1. We then make k large positive constant to prioritize completion of the high-priority goal; in this example, so long as $k > 3$, the MDP will prioritize high-priority task completion above any mission-related action, even if such an action may contribute to the completion of multiple mission goals.

Transition probabilities for this case study are given in Eq. (7), and they are dependent on all aspects of the current state as well as the predicted next action choice a_k :

$$\begin{aligned} T(s^i, a_k, s^j) &= \prod_{z=1}^{n_f} p(f_z^j | s^i, a_k) * \prod_{z=1}^{n_g} p(g_z^j | s^i, a_k) \\ T(s^i, a_k, s^j) &= p(f_1^j | f_1^i, a_k) * \prod_{z=1}^3 p(g_z^j | g_z^i, A^i, a_k) \end{aligned} \quad (7)$$

This product formulation for the MDP probability tensor presumes conditional independence between all goal and high-priority goal flags, with the general form shown on the first line and the specific form for the case study shown on the second line. Not shown in this equation is that we presume the action history in s^j will contain action a_k as the “previous action” with 100% probability at the next iteration. The first term of Eq. (7) is the effect of the high-priority goal flag on the transition probability. For this case study, we give the act of high-priority button pushing a 100% probability of the button becoming inactive postaction, which is a simplification consistent with the expectation that high-priority goals would necessarily be given the time and concentration to be completed correctly the first time the completing action was pursued.

The second term in Eq. (7) characterizes the mission goal properties, which are generally a function of goals already accomplished, the action history, and the current selected action. However, there are several simpler cases that capitalize on conditional independence when possible. Mission goal probability may become $p(g_z^j | g_z^i)$ when the goal flag is set and may not reset ($g_z^j = g_z^i$). Mission goal probability may be $p(g_z^j | a_k)$ when only the current choice of action matters in determining goal satisfaction (e.g., for a coin flip), $p(g_z^j | g_z^i, a_k)$ when the Markov property truly holds (thus, the action history has no impact), and $p(g_z^j | g_z^i, A^i)$ when there is a delayed reaction time (for instance, after taking a medication). In the final case, the importance of a_k is still reflected through its insertion in the action history; this is because of our earlier assumption that the only valid transition for the action-history ($A^i \rightarrow A^j$) is an “update” process that includes a_k . For this case study, the oldest action is forgotten and the newest action is set to the chosen action a_k , such that $A^i = \{a_1^i, a_2^i, \dots, a_{n_h}^i\}$ becomes $A^j = \{a_2^j, \dots, a_{n_h}^j, a_k^j\}$.

For our case study, there are, at most, only 10 possible transitions for each state, and the no-op action is included for use when no goal requires accomplishment. The push-button and no-op actions each have only one possible valid next state with transition probability 1. Eating and mission-related problem-solving actions have at most two outcomes each: staying in the same state (no change) or transitioning to corresponding goals being completed. If the mission goal has already been completed, it stays in the absorbing (completed goal) state with 100% probability. The

“drink” action has up to four possible outcomes, due to a coupling of the drink action with both the hunger and thirst goal objectives; the action of taking a drink, in this case, may sate a person’s feeling of hunger. For convenience, we refer to a transition for a mission goal g_x^i due to a completing action a_k as having probability p_{kx} in the no-change case from 0 to 0, and $(1 - p_{kx})$ for transitioning from 0 to 1:

$$(1 - p_{11}) = p(g_1^j | g_1^i, A^i, a_k) = \lambda_{11} \frac{(a_1^i == a_1) + (a_2^i == a_1) + (a_3^i == a_1) + (a_k == a_1)}{3 + 1}, \quad g_1^i = 0, \quad g_1^j = 1 \quad (8)$$

An example of probability calculation is given in Eq. (8) for the goal of sating hunger, with an action history of length $n_h = 3$. Here, we look to see how many times the completing action a_1 appears in the actions chosen. Every time a match is found, a “1” is tallied and we sum the number of times the action appears. We divide by the number of terms to normalize. We then multiply by a weighting factor in the range $[0 \ 1]$ to scale this value; if $\lambda_{11} = 1$, then eating a chip four times in a row will cause hunger to be sated with 100% probability. More complex formulations could give unequal weight to each term in the numerator, such that the time at which an action occurs in the history may give it more or less impact, or involve the use of a weighted product of terms instead of a weighted sum.

We evaluate MDP formulations by comparing policy outputs. We assess the impact of reward function weights by varying reward weightings of two goals at a time, and we assess the impact of action-history length by varying the values of n_h ; we assess the level of transience of these impacts by comparing the outcomes of the same reward and action-history length variations against different choices of probabilities p_x in the transition probabilities. We examine the tradeoff between different types of goal interdependencies, e.g., two independent goals (eating and math), two partially dependent goals (eating and drinking), and task-level versus high-priority goals (eating and button pushing). Currently, values for transition probabilities are manually chosen but could be determined from experimental results. An example of probabilities with an action-history length of $n_h = 0$ is that the eating action has 25% probability of sating hunger (and 75% not); drinking has 25% probability of sating hunger, 25% of sating thirst, and 25% of sating both hunger and thirst (and 25% chance of neither hunger nor thirst transitioning); and problem solving has 75% probability of completing the work effort. Note that, in this case, the drinking action has a probability $p_{21} = 0.5$ of hunger remaining the same and a probability $p_{22} = 0.5$ of thirst remaining the same; probabilities for this set of four outcomes could be arbitrary, so long as they sum to 100%. Simulation results have shown that the HIP MDP outputs sensible policies but can be sensitive to stored action-history size [13].

Complexity tradeoffs are central to any discussion of the value of the MDP formulation for this work. Consider the HIP MDP in the context of the larger decision-making problem. The human-aware robot decision-making process we recommend uses linked but separate MDPs: one for human intent prediction and one for robot action choice. In RAC, the robot’s state and actions must be modeled as well as the human’s state and actions; the decoupling we propose is only possible because of our assumption that the human will not be influenced by the robot’s actions [12]. Suppose the robot is executing a simple set of button-pressing tasks. The RAC state space must contain button status and a model of robot state (e.g., unstowed but retracted, extended to a particular button). As described in [14], a simple RAC domain might have three goals related to button pressing and keeping the area clear for the human (high priority), and four actions to choose between (no-op, pressing button 1, pressing button 2, and return to unstow). Additional flag data indicate whether robot action is in progress or complete and whether the human and robot are becoming too close to each other. The robot’s state space also must include information on the human’s actions to predict and avoid conflict or physical collision. This RAC state-space size, as described in [14], is $^R n_s = 2^3 * 3^1 * 2^1 * 2^1 * 5^1 * 4^1 = 1920$, alongside the HIP MDP of size $^H n_s = 10,000$. If we used a combined MDP, the human action set need not be included (HIP already includes the observed human action), but the next-predicted action would need to be determined internally. Thus, $n_s = (2^3 * 2^1 * 5^4) * (2^3 * 3^1 * 2^1 * 2^1) = 960,000$ for the combined MDP, which is manageable but large. Consequently, problems formulated independently for each actor with our modeling scheme require on the order of 100,000–500,000 states and, in other work [14], McGhan determined that the complementary MDP for robot action choice given deterministic HIP input need only increase in complexity as a function of the size of the possible human intent set.

From this discussion, we draw the following conclusions: the decomposition of HIP and RAC supports significant computational savings (nearly two orders of magnitude for our example) over a combined MDP formulation. The MDP-based robot decision-making problem still has the potential to become computationally intensive as the number of human and robot goals and actions grows, but complexity will scale more slowly with the split between intent prediction and robot action choice. Although we did not create a POMDP model of our case study, addition of the POMDP belief state is known to increase complexity, which in turn reduces scalability [15]. We believe relaxation of the observability assumption is not necessary due to the rigorous astronaut training program and detailed ISS environment models.

IV. Conclusions

A MDP framework for predicting human intent during human-robot operations in a common workspace has been presented. Given scripted and rehearsed astronaut and robot activities in a space environment, a key assumption can be made: the human can ignore the robot in its choice of actions, enabling the robot to act both without explicitly communicating and without disturbing or endangering its human companion. The output from the simulated human HIP MDP described in this work can provide human intent data to a robot action planners.

The proposed Human Intent Prediction MDP framework requires a best-fit model of the human. Generation of an accurate model will require long-term observations of each astronaut as well as incorporation of data from mission task scripts and initial interviews with each astronaut. A priori MDP parameter determination is likely possible for the ISS shared workspace domain given the rigorous and repeated training conducted on Earth as well as the potential to solicit extensive input from the astronauts based on questionnaires as well as long-term observations of individual astronaut preferences. Such training may be sufficient and even preferred given that HIP behaviors will be repeatable as a function of astronaut observations. This information could also be used to initially seed the HIP, but then additional in situ data could be used to further train the HIP MDP.

While this work assumes task assignments are independent and fixed prior to the start of the work session, it is possible for task assignments to evolve. With the model described above, the HIP MDP would need to be re-run with each new task set; extension of the MDP to model changes in task preferences within the state may allow for a smoother evolution in task preferences at the cost of increased state-space complexity. Modelling is discussed for a well-known environment where astronauts perform near-deterministic action sequences due to the level of scripting in their day-to-day activities. This assumption applies to any structured environment where humans have been well-trained in agreed-upon task sets. Future exploration of cases where the decoupled HIP and robot action choice model breaks down will help better define when robot action choice must feed back into HIP, as well as conditions under which explicit communication is required.

References

- [1] Aggarwal, J. K., and Cai, Q., “Human Motion Analysis: A Review,” *Computer Vision and Image Understanding*, Vol. 73, No. 3, March 1999, pp. 428–440. doi:10.1006/cviu.1998.0744

- [2] Poppe, R., "Vision-Based Human Motion Analysis: An Overview," *Computer Vision and Image Understanding*, Vol. 108, Nos. 1–2, 2007, pp. 4–18. doi:10.1016/j.cviu.2006.10.016
- [3] Poppe, R., "A Survey on Vision-Based Human Action Recognition," *Image and Vision Computing*, Vol. 28, No. 6, 2010, pp. 976–990. doi:10.1016/j.imavis.2009.11.014
- [4] Karami, A.-B., Jeanpierre, L., and Mouaddib, A.-I., "Human–Robot Collaboration for a Shared Mission," *Proceedings of the 5th ACM/IEEE International Conference on Human–Robot Interaction (HRI'10)*, IEEE, Piscataway, NJ, 2010, pp. 155–156.
- [5] Karami, A.-B., Jeanpierre, L., and Mouaddib, A.-I., "Partially Observable Markov Decision Process for Managing Robot Collaboration with Human," *21st International Conference on Tools with Artificial Intelligence (ICTAI '09)*, IEEE Computer Society, Washington, D.C., 2009, pp. 518–521.
- [6] Matignon, L., Karami, A. B., and Mouaddib, A. I., "A Model for Verbal and Non-Verbal Human–Robot Collaboration," *Papers from the 2010 AAAI Fall Symposium: Dialog with Robots*, Vol. FS-10-05, AAAI, 2010.
- [7] Pineau, J., Gordon, G. J., and Thrun, S., "Anytime Point-Based Approximations for Large POMDPs," *Journal of Artificial Intelligence Research*, Vol. 27, No. 1, 2006, pp. 335–380.
- [8] Lasota, P., Nikolaidis, S., and Shah, J., "Developing an Adaptive Robotic Assistant for Close Proximity Human–Robot Collaboration in Space," *AIAA Infotech@Aerospace*, AIAA Paper 2013-4806, Aug. 2013.
- [9] Puterman, M. L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed., Wiley, New York, 1994, pp. 18–22, 166.
- [10] Russell, S. J., and Norvig, P., *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice–Hall/Pearson Education, Upper Saddle River, NJ, 2003, pp. 613–618.
- [11] Casner, S. A., "Understanding the Determinants of Problem-Solving Behavior in a Complex Environment," *Human Factors*, Vol. 36, No. 4, Dec. 1994, pp. 580–596.
- [12] McGhan, C. L. R., and Atkins, E. M., "Human Productivity in a Workspace Shared with a Safe Robotic Manipulator," *Journal of Aerospace Information Systems*, Vol. 11, No. 1, Jan. 2014, pp. 1–18. doi:10.2514/1.54993
- [13] McGhan, C. L. R., Nasir, A., and Atkins, E. M., "Human Intent Prediction Using Markov Decision Processes," *Infotech@Aerospace Conference*, AIAA Paper 2012-2445, June 2012.
- [14] McGhan, C. L. R., "Safe and Efficient Robot Action Choice Using Human Intent Prediction in Physically-Shared Space Environments," Ph.D. Dissertation, Univ. of Michigan, Ann Arbor, MI, 2014.
- [15] Lusena, C., Goldsmith, J., and Mundhenk, M., "Nonapproximability Results for Partially Observable Markov Decision Processes," *Journal of Artificial Intelligence Research*, Vol. 14, June 2001, pp. 83–103.

A. Banerjee
Associate Editor