MOMDP-based target search mission taking into account the human operator’s cognitive state
Paulo Eduardo Ulbadino de Souza, Caroline Ponzoni Carvalho Chanel, Frédéric Dehais

To cite this version:
Paulo Eduardo Ulbadino de Souza, Caroline Ponzoni Carvalho Chanel, Frédéric Dehais. MOMDP-based target search mission taking into account the human operator’s cognitive state. 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2015), Nov 2015, Vietri sul Mare, Italy. Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on, pp.729-736, 2015. <hal-01264111>
Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: [http://oatao.univ-toulouse.fr/](http://oatao.univ-toulouse.fr/)

**Eprints ID**: 14666

**To link to this paper**: DOI:10.1109/ICTAI.2015.109

**URL**: [http://dx.doi.org/10.1109/ICTAI.2015.109](http://dx.doi.org/10.1109/ICTAI.2015.109)

**To cite this version**: Ulbadino de Souza, Paulo Eduardo and Ponzoni Carvalho Chanel, Caroline and Dehais, Frédéric *MOMDP-based target search mission taking into account the human operator’s cognitive state*. (2015) In: 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2015), 9 November 2015 - 11 November 2015 (Vietri sul Mare, Italy)

Any correspondence concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)
MOMDP-based target search mission taking into account the human operator’s cognitive state

Paulo E. U. Souza, Caroline P. Carvalho Chanel, Frédéric Dehais
Institut Supérieur de l’Aéronautique et de l’Espace (ISAE-SUPAERO)
Université de Toulouse, 31055 TOULOUSE Cedex 4
name.surname@isae-supero.fr

Abstract—This study discusses the application of sequential decision making under uncertainty and mixed observability in a mixed-initiative robotic target search application. In such a robotic mission, two agents, a ground robot and a human operator, must collaborate to reach a common goal using, each in turn, their recognized skills. The originality of the work relies in considering that the human operator is not a providential agent when the robot fails. Using the data from previous experiments, a Mixed Observability Markov Decision Process (MOMDP) model was designed, which allows to consider aleatory failure events and the partial observable human operator’s state while planning for a long-term horizon. Results show that the collaborative system was in general able to successfully complete or terminate the mission, even when many simultaneous sensors, devices and operators failures happened. So, the mixed-initiative framework highlighted in this study shows the relevancy of taking into account the cognitive state of the operator, which permits to compute a policy for the sequential decision problem which prevents to re-planning when unexpected (but known) events occurs.

Keywords—MOMDP, robotics, mixed initiative, operators cognitive state estimation

I. INTRODUCTION

Unmanned Vehicles (UVs) are becoming increasingly present in a wide variety of operational contexts such as military operation, border security, inspection of contaminated area for prevent human from hazard exposure. Most of scientific and technical efforts have focused on the implementation of smart sensors, complex embedded systems and autonomy to enhance the efficiency of the UVs [1], especially when the human operator can not analyze or access visual data [1]–[4]. However these developments were generally achieved without questioning the integration of the human operators in the control loop [5]: the human operator is considered as a providential agent that will be able to take over when sensors or automations fail [2]–[4]. Yet, poor user interface design, complexity of automation and high operational pressure can leave the human operator ill-equipped when mental workload exceeds human capacity [6]. For instance, careless design of authority sharing can lead to human-automation conflicts when the human operator misunderstand the automation behavior [7], [8]. The occurrence of such situation is critical as long as it may cause “mental confusion” (i.e. the human operator is unable to glance and process the relevant parameters) [8] or attentional tunneling (i.e. the human operator is excessively focused on a single display) [9] yielding to irrational behavior [10]. Not surprisingly, a safety analysis report [11] revealed that human factors issues were involved in 80% of accidents. This trend has led Cummings and Mitchell (2008) to state: “Because of the increased number of sensors, the volume of information, and the operational demands that will naturally occur in a multiple-vehicle control environment, excessive cognitive demands will likely be placed on operators. As a result, efficiently allocating attention between a set of dynamic tasks will be critical to both human and system performance. - p. 451”.

A promising avenue to deal with these issues is to consider that robot and human abilities are complementary and are likely to provide better performance when joined efficiently than when used separately. This approach, known as mixed-initiative [12], [13] defines the role of the human and artificial agents according to their recognized skills. It allows the human and the robot to set the appropriate level of autonomy of the robot [14]. An interesting mixed-initiative proposition, presented by [15], relies on a statistical approach to determine which entity (i.e human or UVs) is the most efficient for a given task. Interestingly enough, this approach paves the way for allocating roles and sharing authority between the human and artificial agents. In [3], a mixed-initiative planning approach is proposed to monitoring the system and to coordinate operator’s interventions in a rescue scenario. In this work, the mixed-initiative planning continuously coordinates, integrates and monitors the operator’s interventions and decisions. Another example can be found in [16], in which a robot and an human operator collaborate for an urban search and rescue mission in order to detect and report objects of interest.

A key issue to design a mixed-initiative system is to implement a decision system. This latter defines the role and the authority of human and artificial agents, while estimating capabilities of evolved human (intention, situation awareness, sensor’s failure perception) and robotic agent (sensor’s status, mission task, etc). Such decision-making
system can be governed by a policy resulting from the resolution of a Partially Observable Markov Decision Process (POMDP), as proposed by [17], which is able to adapt itself to the user’s intention getting feedback from the user in terms of satisfaction. A different way to drive interaction using POMDPs is studied in [18] for assisting persons with dementia during hand-washing. Note that, the state vector of the robot can be often considered as fully observable while the operator’s cognitive state is, by definition, partial observable. Such decomposition can be addressed using a Mixed Observability Markov Decision Process (MOMDP) [19], which is a stochastic model derived from the POMDP [20]. The MOMDP is a formal framework that considers fully and partially observable state variables under probabilistic uncertainties while decreasing the computational cost required to produce an optimal policy [19]. In addition, these two types of agents may face unexpected random situations during the mission. It can be modeled as probabilistic effects of actions. Moreover, this kind of model allows the inclusion of the uncertainty in the observations of the agents’ states (i.e. the cognitive state of the human operator) and the environment. The MOMDP aim to achieve a policy that maps an optimal action for each belief state – composed by the observable state and the partially observable state estimation. Thus, it is expected that the resulting policy could help to implement a genuine adaptive interaction, because this formalism is perfectly suited to maintain a state estimation and to decide of the men-robot system dynamics. We note $\Delta$ the belief state space. At each time step $t$, the agent updates its belief state defined as an element $b_t \in \Delta$ using the Bayes’ rule [20].

$$b_t^o(s') = \frac{p(o|s',a) \sum_s p(s'|s,a)b(s)}{\sum_s p(o|s',a) \sum_s p(s'|s,a)b(s)}$$ (1)

Solving a POMDP consists in finding a policy function $\pi: \Delta \rightarrow A$ that maps to each belief state an optimal action that maximizes a performance criterion. The expected discounted reward from any initial belief $V^*(b) = E_{\pi} \sum_{\tau=0} \gamma^\tau r(b_t, \pi(b_t)) \mid b_0 = b$ is usually optimized. The value of an optimal policy $\pi^*$ is defined by the optimal value function $V^*$ that satisfies the Bellman optimality equation:

$$V^*(b) = \max_{a \in A} \left[r(b,a) + \gamma \sum_{s \in \Omega} p(o|a,b)V^*(b_{o}^s) \right]$$ (2)

where, $\gamma$ is the discount factor. When $r(b,a)$ can be computed as an average gain $r(b,a) = \sum_s r(s,a)b(s)$, the optimal value of belief states is proven to be piecewise linear and convex and solution of the Bellman’s equation [20]. As it, at $n^{th}$ optimization stage, the value function $V_n$ can be parametrized as a set of hyperplanes over $\Delta$ known as $\alpha$-vectors. An $\alpha$-vector and the associated action $a(\alpha_n)$ define a region of the belief state space for which this vector maximizes $V_n$. Thus, the value of a belief $b$ can be defined as: $V_n(b) = \max_{\alpha_n \in \alpha_n} b \cdot \alpha_n$. The optimal policy at this step is then: $\pi_n(b) = a(\alpha_n)$.

II. BACKGROUND

A. POMDP overview

POMDPs model situations where the agent only has access to partial information about the state of the system. A POMDP is a Markov Decision Process where the agent does not have access to the state of the system: it has only a partial and imprecise observation [20]. In this context, the agent maintains a probability distribution over states, i.e. a belief state, which is updated after each action executed and observation perceived.

A POMDP is a tuple $(S, A, \Omega, T, O, R, b_0, \gamma)$ where:
- $S$ is a bounded set of states;
- $A$ is a bounded set of actions;
- $\Omega$ is a bounded set of observations;
- $T: S \times A \times S \rightarrow [0; 1]$ is a transition function such that $T(s_{t+1}, a, s_t) = p(s_{t+1} \mid a, s_t)$;
- $O: \Omega \times S \rightarrow [0; 1]$ is an observation function such that $O(a_t, s_t) = p(o_t \mid s_t)$;
- $R: S \times A \rightarrow \mathbb{R}$ is a reward function associated with a state-action pair, and;
- $b_0$ is the initial probability distribution over states.

We note $\gamma \in [0, 1]$ is the discount factor.

![Figure 1. Transition models of a POMDP and a MOMDP.](image-url)
Recent offline solving algorithms HSVI2 [21] or SARSOP [22], for instance, approximate the value function with a bounded set of belief states \( B \), where \( B \subset \Delta \). These algorithms implement different heuristics to explore the belief state space using probabilistic trials reaching in this way only relevant belief states, and updating the value of \( V \) for them, instead of computing the value function for all the belief state space, which is a continuous space.

B. MOMDP

The Mixed Observability Markov Decision Process (MOMDP) is an extension recently proposed for the POMDP model [19], which explores the particular structure where certain state variables are fully observable. This factored model leads to a very significant time gain in policy computation, improving the efficiency of a point-based algorithms. According to [23] the completely observable state is represented by \( x \) and the partially observable state by \( y \).

In this way, the couple \((x, y)\) specifies the complete state with \(|S| = |\mathcal{X}| \times |\mathcal{Y}|\), where \( \mathcal{X} \) represents the space with all the possible values of the variable \( x \) (resp. \( \mathcal{Y} \) to \( y \)).

A MOMDP is a tuple \((\mathcal{X}, \mathcal{Y}, A, \Omega, T_X, T_Y, \Omega, R, b_0, \gamma)\), where:

- \( \mathcal{X} \) is the bounded set of fully observable state variables;
- \( \mathcal{Y} \) is the bounded set of partially observable state variables;
- \( A \) is a bounded set of actions;
- \( \Omega \) is a bounded set of observations;
- \( T_X : \mathcal{X} \times A \times \mathcal{X} \times \mathcal{Y} \to [0; 1] \) is a transition function such that \( T_X(x, y, a, x') = p(x'|x, y, a) \);
- \( T_Y : \mathcal{Y} \times \mathcal{X} \times A \times \mathcal{X} \times \mathcal{Y} \to [0; 1] \) is a transition function such that \( T_Y(y, x, y', a', x') = p(y'|y, x, y', a) \);
- \( O : \Omega \times \mathcal{X} \to [0; 1] \) is an observation function such that \( O(\alpha, a, x', y') = p(o|x', y', a) \);
- \( R : \mathcal{X} \times \mathcal{Y} \times A \to \mathbb{R} \) is a reward function associated with a state-action pair; and:
- \( b_0 = (x_0, b_{y_0}) \) is the initial probability distribution over states,
- \( \gamma \in [0, 1[ \) is the discount factor.

Note that, as the probability distribution over states concerns only the \( \mathcal{Y} \) set, the belief state update is redefined as:

\[
\begin{align*}
\pi_x^\alpha \leftarrow \arg \max_{\pi_x \in \Pi} & \sum_{x_b \in B} \sum_{y \in \mathcal{Y}} \sum_{y' \in \mathcal{Y}} \pi(y|x, y, a) p(y'|x, y, a, x') p(x'|x, y, a) b(y) \\
& \times \sum_{t=0}^{\infty} \gamma^t r((x_t, b_y^t), \pi((x_t, b_y^t))) | b_0 = (x_0, b_{y_0})
\end{align*}
\]  

(4)

As for the POMDP, the value function at a time step \( n < \infty \) can be also represented by a set of \( \alpha \)-vectors:

\[
V_n(x, b_y) = \max_{\alpha \in \Gamma(x)} (\alpha \cdot b_y)
\]  

(5)

where \( \alpha \) is the hyperplan over the space \( \mathcal{B}_y(x) \). In this way, the value function over the complete state space is parametrized by the set \( \Gamma_y(x) \), i.e. \( \Gamma = \{\Gamma_y(x), x \in \mathcal{X}\} \).

So, given a belief state \((x, b_y)\) the optimal action is defined by the action associated with the \( \alpha \)-vector that maximizes \( \max_{\alpha \in \Gamma_y(x)} (\alpha \cdot b_y) \). For more details about MOMDP algorithm resolution, please see [19], [23].

Next, we present previous experiments which were used as base for statistical data in order to leaning the MOMDP model for the target search mission taking into account the operator’s cognitive state.

III. PREVIOUS EXPERIMENTS

A. Material

The experimental set-up was developed at ISAE-SUPAERO. It was composed of a robot and a ground station. The robot was equipped with different sensors such as a GPS for autonomous navigation, an Ultrasound sensor to detect and avoid obstacle, a video camera and an Xbee transmitter to communicate with the ground station. It had a 15 minutes autonomy thanks to electrical battery. The robot could be operated in \textit{manual mode} or in \textit{supervised mode}. In manual mode, the robot was operated with a joystick. In supervised mode, the robot performed waypoint navigation autonomously, but any action of the operator with the joystick let her/him take over until the joystick was released. The ground station was displayed on a 24-inch screen showing different kinds of information to control and supervise the robot such as a tactical map, a panoramic video scene screen; a mission synoptic ; an interactive panel sending the requests to the human operator; a status panel panel indicating the state of the GPS, the ultrasound sensor and the battery; and a guidance mode state (i.e. supervised or manual). Note that the operator could not see the robot and only gathered information through the screen. Fig. 2 shows the interface of the ground station to operate the robot during the mission.

B. Assessing probability of failures of the robot

A first experiment was conducted to assess the probability of failure of the different sensors and devices embedded in the robot. Thirty tests were run and consisted of a 10 minutes navigation task with the robot evolving among four waypoints in a field at ISAE-SUPAERO campus. The results are summarized on Table I, where \( \text{FP} \) design the failure probability.
C. Assessing human’s performance

Data collected from a previous experiment [10] were used to assess the probability of the human operator to perceive robot failures. The scenario of this experiment consisted of a target localization and identification task involving a ground robot and a human operator. The target to be identified had two short messages written in white on each side (front side OK, back side KO). The mission had four main segments: S1 - Reach the area, S2 - Scan for target, S3 - Identify target, and S4 - Battery-Failure. At the beginning of the mission, the robot evolved in an autonomous supervised mode to reach the target area (S1) and then start scanning the area to detect the target (S2). When the robot was close to the target, the operator had to take over in manual mode and to identify the target (S3). While the operator was performing the identification task, a low battery event was triggered (S4). In turn this event yield to a safety procedure that made the robot to go back to base autonomously. As this event occurred while the operator was excessively focused on his target identification task, it was expected that he would miss the alerts and thus persist in achieving the target detection task.

1) Assessing failure perception: 12 subjects participated to the experiment and were equipped with an electrocardiogram (ECG) and a 25 Hz Pertech head mounted eye tracker. This latter device was used to collect participants’ eye gaze on the user interface. More specifically we focused our eye tracking analysis on eight areas of interest (AOI) of the user interface: 1) tactical map, 2) message panel, 3) guidance mode (supervised vs manual), 4) synaptic, 5) “back to base” warning, 6) GPS and ultrasound status, 7) battery status, 8) panoramic video. The collected ocular data were used to set the probability of the operator to perceive the sensor’s status (sensors statuses are summarized in Table I). This sensor status perception probability is based on the normalized sum of the averaged fixation time ($\Delta T$) on the related AOIs. For instance, when the GPS or the ultrasound are lost, the icons turns to red (area 6) and the robot is stopped (i.e it can be seen through the panoramic video - area 8). When the low-battery event occurs, three changes can be observed in the user interface: (i) the battery icon (area 7) turns to orange with the associated low battery message, (ii) the mode changes automatically from manual to supervised, and area 3 blinks twice and (iii) the segment status became back to base (area 5).

Thus we introduced the $spGpsUs$ boolean state variable that can be used to model perception about a GPS or ultrasound status by the operator:

$$p(spGpsUs = Y \mid auto \& \& (gpsKO \mid usKO)) = \frac{\Delta T_{Area\ 6} + \Delta T_{Area\ 8}}{\Delta T_{all\ areas}} = 0.70$$

$$p(spGpsUs = Y \mid manual \& \& (gpsKO \mid usKO)) = 0.86$$

With the same reasoning, for the $spBatt$ (Battery status perception) boolean state variable, the transition probability was defined by the normalized sum of the averaged time that the participants expended looking to areas 3, 5 and 7 during the manual and autonomous operations:

$$p(spBatt = Y \mid manual \& \& battKO) = \frac{\Delta T_{Area\ 3} + \Delta T_{Area\ 5} + \Delta T_{Area\ 7}}{\Delta T_{all\ areas}} = 0.021$$

$$p(spBatt = Y \mid auto \& \& battKO) = 0.033$$

2) Assessing cognitive availability: the result of the experiment revealed that 8 participants out of 12 did not understand the robot behavior, though some of them glanced at the battery failure indicator. These 8 participant persevered to achieve the no-longer-relevant identification task [10]. This typical behavior is known as "attentional tunneling" and is defined as "the allocation of attention to a particular channel of information, diagnostic hypothesis or task goal, for a duration that is longer than optimal, given the expected costs of neglecting events on other channels, failing to consider other hypotheses, or failing to perform other tasks" [25]. Therefore, the inference of such impaired attentional state is of great importance to design a mixed initiative system. We implemented an Adaptive Neuro-Fuzzy Inference System (ANFIS) to detect attentional tunneling that is associated with higher cardiac activity, decreased

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Ok</th>
<th>Not Ok</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery</td>
<td>batOK</td>
<td>batKO</td>
<td>2/30</td>
</tr>
<tr>
<td>GPS$^1$</td>
<td>gpsOK</td>
<td>gpsKO</td>
<td>3/30</td>
</tr>
<tr>
<td>Ultrasound$^2$</td>
<td>usOK</td>
<td>usKO</td>
<td>3/30</td>
</tr>
<tr>
<td>Camera$^3$</td>
<td>camOK</td>
<td>camKO</td>
<td>3/30</td>
</tr>
<tr>
<td>Ground station$^4$</td>
<td>stOK</td>
<td>stKO</td>
<td>1/30</td>
</tr>
<tr>
<td>Joystick$^2$</td>
<td>jOK</td>
<td>jKO</td>
<td>1/30</td>
</tr>
<tr>
<td>Xbee$^2$</td>
<td>xOK</td>
<td>xKO</td>
<td>2/30</td>
</tr>
</tbody>
</table>

$^1$ indicates the essential sensors for autonomous operation.
$^2$ indicates the essential devices for manual operation.

Figure 2. Operator’s interface and areas of interest [24].
saccadic activity and long concentrated eye fixations [24]. The ANFIS classifier had a probability of 91.1% to detect Attention Tunneling (please report to [24] for more details). This detection probability was used in this study to define the observation function of the Cognitive Availability state variable as shown further. Cognitive Availability is defined here as the capability of the human operator to be available and aware of the robot’s status during all mission tasks.

Table II
ATTENTION TUNNELING PROBABILITY FUNCTION

<table>
<thead>
<tr>
<th>available_N</th>
<th>available_Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>oAvailable_N</td>
<td>91.1</td>
</tr>
<tr>
<td>oAvailable_Y</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

3) Cognitive countermeasure to assist the human operator: many studies revealed that alarms are inefficient to warn human operator during high workload situations such as performing manual control and identifying target [26]. Rather than adding alarms during stressful situations, an optimal solution to warn the human operator consists of temporarily removing the information the human operator is focusing on, and replacing it by an explicit visual stimulus designed to change the attentional focus. The principle of this cognitive countermeasure was tested in second experiment with 11 participants facing the same scenario (i.e. target identification task and battery failure). The results revealed that these cognitive countermeasures helped 10 participants out of 11 to perceive the battery failure and to let the robot go back to base [9].

IV. MODELING THE COLLABORATIVE TARGET IDENTIFICATION MISSION

Using all those previous experimental data, a MOMDP model was defined in order to drive the adaptive interaction between the human operator and the UV. The choice for a Mixed Observability model comes from the nature of our problem: the robot and mission states can be considered as fully observable, while the operator’s cognitive ability, here considered as Cognitive Availability is a partially observable state variable by definition.

The mission can be decomposed in six high level phases: going to zone, searching the target, handling the target, returning to base, on base and failed. The robot states can be defined by the cross product of the phases, the embedded sensors statuses, the statuses of the ground station, the on board camera, the Xbee and the joystick devices, and the Cognitive Availability as cognitive state of the operator. Next, we present the fully and partially observable state variables considered in the model.

Fully observable state variables (X): The section III-C and tables I and III, present the fully observable state variables considered in the mission modeling. As discussed before, mission phases were classified according to the operation mode. The sensors statuses were discretized in two possibilities: OK and KO (not OK). It was also assumed that after a sensor failed, it switches to KO and remained KO until the end of the mission. The sensors’ failure probabilities were shown in Table I.

Fig. 3 summarizes the transition function for the mission phase state variable. A manual mode was associated with each autonomous mode (except the on base and failed mission phases). One can argue that, in a human operator’s point of view, there is only one manual mode, but for modeling propose, the manual mode was factored in four phases (see Table III) to prevent the planner from selecting a supervised mode already held when returning to the autonomous operation mode.

Figure 3. Mission phases and summarized transitions. The loop transitions (red lines) indicate the transitions observed after getAttention or cMeasure actions.

Table III
MISSION PHASES

<table>
<thead>
<tr>
<th>mission phase</th>
<th>autonomous mode</th>
<th>manual mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>going to zone</td>
<td>e0</td>
<td>m0</td>
</tr>
<tr>
<td>target searching</td>
<td>e1</td>
<td>m1</td>
</tr>
<tr>
<td>target handling</td>
<td>e2</td>
<td>m2</td>
</tr>
<tr>
<td>returning to base</td>
<td>e3</td>
<td>m3</td>
</tr>
<tr>
<td>on base (final)</td>
<td>e4</td>
<td>-</td>
</tr>
<tr>
<td>failed (final)</td>
<td>e5</td>
<td>-</td>
</tr>
</tbody>
</table>

As show in previous experiments presented before, it is relevant to define two fully observable variables that model the operator’s sensor’s status perception (see section III-C). Note that, operator’s perception about sensor’s status (GPS/ultrasound and battery) state variables are assumed as fully observable, because it is not possible to observe if the human operator perceived (i.e. his cognitive process) only with the eye-tracker device. In this case, for example, if the operator looked to the areas (6 and 8) between two decision time steps, he should detect, with a probability of 0.70, for supervised mode, or 0.86, for manual mode, if there was a GPS or a ultrasound breakdown because the related icon turns to orange (see Fig. 2 and Section III-C).

Partially observable state variable (Y): The operator’s Cognitive Availability is considered in this study as the opposite of Attentional Tunneling [9], [24], [27]. The measure
of the allocation of attention if not a straightforward task [24]. Therefore, we consider the Cognitive Availability of
the human operator as a partially observable variable. Hence: available_Y models that the human operator is cognitively
available (resp. available_N, not cognitively available). Associated with this partially observable state variable we have
two possible observations: oAvailable_Y meaning that the
operator is observed as cognitively available and potentially aware of the situation and oAvailable_N modeling that he is
observed as not cognitively available. Table II summarizes
the observation probability function for this observation
variable.

Actions: Discrete actions were defined as: goToZone, tgtSearch, tgtHandle, retBase, onBase, getAttention and
cMeasure. Action result depends on the aleatory sensors
behavior (cf. Table I). For instance, in a nominal case and
based on previous works [9], [24], [27], the robot is able to
autonomously navigate and avoid obstacles, but if the robot
chosen goToZone and the ultrasound sensor fails, the mission
phase turns to manual mode (m0) (see Fig. 3) because the
robot is no more able to avoid obstacles autonomously. If
a low battery event arrives, the robot can return to the base
automatically if it was in a supervised mode. When it was in a
manual mode, it can switch to returning to base (with any
action) automatically only if the human operator is aware of
the failure, i.e. if he was observed as cognitively available
being aware of the situation and by consequence leaving the
joystick.

The getAttention is considered as a non deterministic
action, since it should be used when the robot needed help and the operator’s Cognitive Availability was estimated as “not available” (oAvailable_N). The same occurs with
the action cMeasure (countermeasure), which should be
executed when a low battery event arrives during a manual
operation and the operator was considered as “not available”
(e.g his attention was focused on handling the robot and he
would not notice the alerts on the user interface). In such
case, when a cMeasure action is launched the robot wait the
human operator leaves the joystick (see Fig. 3).

Rewards: The reward function (R) was designed in order to
pay off suitable actions, for instance, goToZone in
the phase e0 when the navigation sensors are OK, and to
punish otherwise. The same occurs with the manual modes
and its essential devices (cf. Table I). Note that, we have
chosen to associate a increasing reward with sequential
phases, i.e reward associated with the action tgtSearch in
e1 considering essential sensors are OK (R=15) is more
important than the action goToZone in the e0 phase (R=10).
We have considered that processing the target in manual
mode is more dependable than in autonomous mode, since
the interpretation done by the human operator is more
reliable. In this case, the reward for the choice of tgtHandle
in manual mode (R=30) is more important than tgtHandle
in supervised mode (R=20).

The action getAttention only has a positive payoff (R=30)
if the operator’s Cognitive Availability is estimated as ”not
available” (oAvailable_N) considering that at least one of
the essential devices are KO and that the human operator
did not see the alert, otherwise the reward is negative (−500).
Similarly, the action cMeasure has a positive payoff (R=50)
only if the operator is perceived as ”not available” in a
manual mode when a low battery event arrives.

A mission is considered as fully accomplished if the robot
had passed through the phases e2 or m2 (resp. processed
the target autonomously or manually) and arrived at base
e4. When the robot returns to the base (autonomously
or manually) before processing the target, the mission
is considered as aborted and when the robot is unable to
reach the base, the mission is considered as failed.

V. SIMULATION RESULTS

The APPL_0.96win SARSOP1 [23] was used as solver. The
grammar of this solver has a special format that differs
from the classical input POMDP file format2. Therefore, we
have developed a script written in Python 2.7.8 to produce
the MOMDP input file. For the MOMDP resolution, we have
set the precision ε to 0.01. We recall that this precision is
related to the difference between the upper and lower bound
of the value function for the initial belief state, which is
considered as a stop optimization criterion.

Statistical analysis was performed to process the results
over 5000 policy simulations. Fig. 4(a) provides an overview
of how many times the robot passed by each mission phase.
Note that, the robot passed through phase e1 exactly 5000
times and never crossed the phase m0, this can be explained
by the fact that the initial state, when all fully observable
state variable values are known, caused a deterministic cycle
in the first time stamp.

To sum up, the mission was fully accomplished 2998
times (59.96%) (cf. Fig. 4(b)), meaning that the robot has
passed through the phases e2 or m2, respectively processing
the target autonomously or manually, and arrived at base
e4. In such cases, the target was handled in autonomous
mode (e2) 1228 times, which represents 41% of successful
missions, and it was handled in manual mode (m2) 1770
times (resp. 59% of successful missions). Fig. 4(b) also
shows that the robot returned to the base in 68.72% of times

1http://bigbird.comp.nus.edu.sg/pmwiki/farm/appl/
2http://www.pomdp.org/code/pomdp-file-spec.shtml
including aborted missions). The mission completely failed, i.e. it reached \( e^5 \), 1,564 times (31.28%).

The Table IV presents an example of a *Fully accomplished mission* where the robot changed its mode to manual (m2) for the operator process the target (bigger reward), then the GPS failed. This was not a problem at that moment because the operator did not need the GPS to handle the target. Next, the robot remained in manual (m3) but the operator seemed not to be aware, so, the robot ask for his attention (action `getAttention`), and the operator led the robot to the base.

Finally, a *aborted mission* is shown in the table V. In this aborted mission a low battery event occurred while the operator was processing the target and the robot observed him or her as “not available” (not aware) of the failure. Consequently, the robot executed a countermeasure action (cMeasure) trying to show the situation to the operator. After, the robot changed its phase to returning to base (e3) and went home. Here, is interesting to observe that the operator never looked to the areas 3, 5 or 7 (cf. Fig. 2), where the low battery event could be identified without the countermeasure.

VI. CONCLUSIONS

This study has shown the effectiveness of the MOMDP model as basis for mixed-initiative actions planning. In such cases, agents must collaborate by bringing, according to their recognized skills, the relevant elements to reach together a shared goal. In our application case, the robot counts on the operator to process a target, since the operator’s interpretation is considered more reliable that the robot’s. Also, the robot may needs the intervention of a human operator in cases where an essential sensor for autonomous navigation breaks down. Our principal contribution in this mixed-initiative problem is that we have considered that the human operator is not a providential agent, i.e. he can be unaware of the situation. To model the problem, we have used data collected from previous experiments with an heterogeneous human-robot system. Based on it, the probability functions were assigned for the sensors failure, operator’s perception about sensor’s status and for the operator’s cognitive availability. With the MOMDP model and a simulated environment, we checked that the collaborative system was in general able to successfully complete or terminate the mission, even when the simulated environment caused many simultaneous sensors/devices/operator failures.

Future work shall to take into account in the model more than one partially observable state variable. For the human factor community, the *estimation of the operator state* is obviously more complex and composed by more state variables than the one considered in this study (workload, stress, engagement, etc). In the future, we hope to take into account more cognitive states as, for instance, the operator’s workload or stress level, and evaluating the policy in a real manner set-up with human operator participants.

ACKNOWLEDGMENT

This work was supported by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil.

REFERENCES


### Fully accomplished mission EXAMPLE

<table>
<thead>
<tr>
<th>Step</th>
<th>oAvailable</th>
<th>Batt</th>
<th>GPS</th>
<th>US</th>
<th>Cam</th>
<th>Joystick</th>
<th>Station</th>
<th>Xbee</th>
<th>Phase</th>
<th>spBatt</th>
<th>spGpsUs</th>
<th>Y available</th>
<th>Action</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e0</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>goToZone</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e1</td>
<td>N</td>
<td>N</td>
<td></td>
<td>Y</td>
<td>tgtSearch</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>OK</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>m2</td>
<td>N</td>
<td>N</td>
<td></td>
<td>Y</td>
<td>tgtHandle</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>OK</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>m3</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>getAttention</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>Y</td>
<td>OK</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e4</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>retBase</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>OK</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e1</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>onBase</td>
<td>35</td>
</tr>
</tbody>
</table>

### Aborted Mission EXAMPLE

<table>
<thead>
<tr>
<th>Step</th>
<th>oAvailable</th>
<th>Batt</th>
<th>GPS</th>
<th>US</th>
<th>Cam</th>
<th>Joystick</th>
<th>Station</th>
<th>Xbee</th>
<th>Phase</th>
<th>spBatt</th>
<th>spGpsUs</th>
<th>Y available</th>
<th>Action</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e0</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>goToZone</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e1</td>
<td>N</td>
<td>N</td>
<td></td>
<td>Y</td>
<td>tgtSearch</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>m2</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td>tgtHandle</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e3</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>N</td>
<td>retBase</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>KO</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>e4</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td>onBase</td>
<td>35</td>
</tr>
</tbody>
</table>


