Decentralized Multi-agent POMDPs Framework for Humans-Robots Teamwork Coordination in Search and Rescue

Hend Al Tair
College of Engineering
Khalifa University
Abu Dhabi, UAE
Email: hend.altair@kustar.ac.ae

Tarek Taha
College of Engineering
Khalifa University
Abu Dhabi, UAE
Email: tarek.taha@kustar.ac.ae

Mahmoud Al-Qutayri
College of Engineering
Khalifa University
Abu Dhabi, UAE
Email: mqutayri@kustar.ac.ae

Jorge Dias
College of Engineering
Khalifa University
Abu Dhabi, UAE
Email: jorge.dias@kustar.ac.ae

***Abstract*—Despite the fact that robots have reached a high level of autonomy in recent years, the need for human presence in certain situations is still essential, especially in search and rescue operations. The human extends the robots capabilities beyond of what they are capable of with current technologies. While current robotic devices are able to navigate, locate, and map search and rescue areas, some interventions require high degree of dexterity and information exchange that implies cooperation between the agents intervening in the area - human and/or robots. This paper presents a framework for modelling the coordination between human responders and robots in search and rescue scenarios using Decentralised Multi-agent Partially Observable Markov Decision Processes (Dec-MPOMDP). In this framework the human is treated as an intelligent agent with separate observations and actions that are communicated with the remaining team (human and robots) to reach the level of synergy required to accomplish joined tasks.**

I. Introduction

Utilising robotic systems in hazardous situations continues to expand and hence research effort is being put to enhance the efficiency and performance of these systems. Robots involvement in hazardous situations not only saves lives by reducing the human exposure to dangerous environments, but also increases the efficiency of responding to these incidents. Examples of those incidents might vary from small fires [1], oil spillage incidents [2], to major nuclear incidents [3]. Most current robotic systems tend to either be autonomous or work under the control of a human operator. However, to deal with emergencies, particularity the large scale ones, systems that allows seamless collaboration between teams of robots and humans are highly desirable. Robust decision making and coordination among agents are required so that they can interact, collaborate and form one team based on shared observations and actions. However, in search and rescue missions it is normal that information can be uncertain which affects the decisions taken. The work proposed in this paper focuses on developing a framework that facilitates the deployment of a group of robots and humans in a search and rescue scenario.

II. Motivation and Related Work

Direct physical human intervention within search and rescue scenarios is very risky and dangerous, and the environmental conditions in these scenarios are beyond what humans can physically tolerate (heat, radiation, narrow openings, visibility, smoke ...). Modern robotic systems are being deployed in such domain. However, it is challenging to inject high degree on intelligence in the robotic system in order to be able to work safely and naturally along side humans.

Recently, researchers proposed various solutions to achieve collaboration between robots and humans. Some of this work is based on sharing map, environments, and locations. In some of these cases, however, a human supervisor is needed to update the global map based on the information from robots, and allocate tasks and resources [5] [6]. Alternatively, others suggested that robots obtain assistance from the human team during task execution through a dialogue such as [7]. Dialogues are used to share specific information such as data symbols and context regarding navigation and obstacle avoidance, in order to make human team aware of the situation. Robots take the initiative to perform a task when communication with the human team has been lost for a specific period of time. However, in these cases, communication channel are expected to be reliable in order to maintain the exchanging of messages. Other work related to human-robot collaboration proposed to use specialized software agents for every robot. For instance, Nourbakhsh et al. proposed to have four types of software agents; interface, task, information, and middle agents [8]. The interface agents are linked to the user interface while the task agent is in charge of ensuring accomplishing users goals. The information agent can get outside resources such as maps. Finally, the middle agent is in charge of providing infrastructure for dynamic run-time discovery of roles and agents.

The decision making in human-robot teams becomes critical when there is uncertainty in the information shared between the team, which is the expectations when dealing with multi-agent collaboration in search and rescue scenarios [9]. The uncertainty and partial observability nature of data gathered during multi-robot collaboration motivated the usage of frameworks that implicitly handle these conditions such as Partially Observable Markov Decision Processes (POMDPs). To handle multi-agent collaboration, POMDPs have been extended using two main models: Decentralized POMDP (Dec-POMDP) and centralized POMDP (MPOMDP). The difference between them is that in case of decentralized each agent observes its local observation only o_i while in the centralized each

978-1-4799-8966-9/15/$31.00 ©2015 IEEE

agent observe the full observation (local and global from other agents) which requires high communication bandwidth o [10]. Free communication decentralized POMDP [11], [12] has been introduced to tackle the heavy communication requirement in multi-agent POMDPs. Generally in search and rescue scenarios communication is limited. Therefore, an efficient approach should based on reducing the MPOMDP as proposed in this paper by adapting the Decentralized Multi-agent Partially Observable Markov Decision Processes (Dec-MPOMDP) framework.

The focus of the work in this paper is to utilize Dec-MPOMDP framework to plan tasks and execute actions among a human-robot teams in uncertain search and rescue environment in order to locate and rescue victims. The remainder of this paper is organized as follows. The paper begins with a brief review of some related work using POMDP with multi-agents and more specifically in search and rescue scenarios, followed by description of the Dec-MPOMDP model. The proposed multi-agent HRI collaborative framework is then presented. Finally, the paper is concluded with brief description of how the solving the model plan.

III. DEC-MPOMDP

The Dec-MPOMDP is a Dec-POMDP with reduced MPOMDP. In order to define the formal model Dec-MPOMDP, a definition of Dec-POMDP will be presented.

A. Model Parameters

Referring to Oliehoek et al. work [13] standard Dec-POMDP for n agents is defined as a tuple $\langle n, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, O, h, b^0 \rangle$ consists of:

- a finite set of n agents;
- $\mathcal{S}$ is a finite set of world states;
- $\mathcal{A} = \times_i \mathcal{A}_i$ is the set $\{a^1, ..., a^j\}$of J joint actions. $\mathcal{A}_i$ is the set of actions available for agent i. At each time step one a=$\langle a_1$,...,$a_n \rangle$ is taken.
- $\mathcal{T}$ the transition function which defined the probability of going to state s' when in state s under action a ,$p(s'|s, a)$.
- $\mathcal{R}$ is the reward function. A reward is given for taking an action a when in state s, $\mathcal{R}(s, a)$.
- $\mathcal{O} = \times_i \mathcal{O}_i$ is the set $\{o^1, ..., o^k\}$ of K joint observations. Every agent i has a set of observations o=$\langle o_1$,...,$o_n \rangle$ at each time step.
- O is the observation function, represents the probability of an observation o after an action a and ending up in state s' experienced, $p(o|a, s')$
- h is the horizon, the number of time steps that agents will interact with their environment.
- $b^0 \in \mathcal{P}(\mathcal{S})$, is the initial state distribution at time t. $\mathcal{P}(\mathcal{S})$ donates the set probability distributions over $\mathcal{S}$.

The planning problem aims at finding a policy π for every agent that is optimal for a particular number of steps h. Joint policies is a tuple of policies $\pi = \langle \pi_1, ..., \pi_n \rangle$. Each policy π_i is a mapping from histories of observations, that are received by agent i, to an action $\pi_i(o_i^1, ..., o_i^t) = a_i$. The joint belief b is a probability distribution over states. It is computed after the action-observation history $\vec{\theta}_i^t$ which consists of all observations received and actions taken up to time step t: $\vec{\theta}_i^t = (a_i^0, o_i^1, a_i^1, ..., a_i^{t-1}, o_i^t)$. The joint belief is calculated:

$$b^{\vec{\theta}^{t+1}}(s') = \frac{P(\mathbf{o}|\mathbf{a}, s')}{P(\mathbf{o}|b^{\vec{\theta}^t}, \mathbf{a})} \sum P(s'|s, \mathbf{a}) b^{\vec{\theta}^t}(s) \tag{1}$$

where $P(\mathbf{o}|b^{\vec{\theta}^t}, \mathbf{a})$ is used for normalization.

IV. DYNAMICS OF THE PROPOSED MULTI-AGENT FRAMEWORK

As illustrated in Figure 1 each agent (robots and humans) receives its local observation from the environment and communicates it to the other agents. A central agent receives the joint observations and within State Estimator (SE) unit it combines the joint observations with previous joint actions and joint belief to update the joint belief and compute joint policies. Joint actions then sent to the agents (robots and humans). Each one of the agents will perform its action.

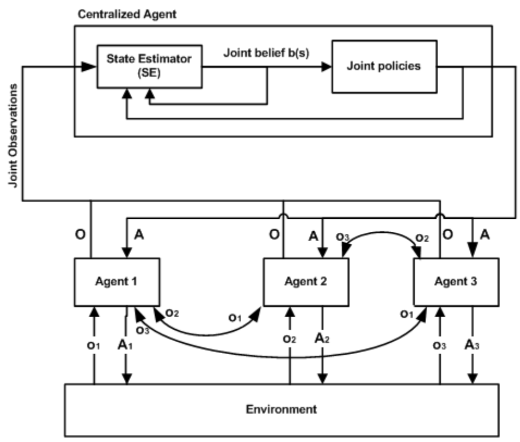

Fig. 1. The figure shows the dynamics of the propose multi-agent HRI in search and rescue

V. SCENARIO DESCRIPTION

In response to a search and rescue mission, a team has been dispatched in order to locate and extract a victim present in the rescue area. This team consists of 2 robots and a human, each with different capabilities, but collectively are capable of dealing with the rescue situation. The two robots are capable of scanning the area, locating the victim, locating any source of danger, and clearing the danger. The robots, however, do not have the capability to safely extract the victim once found. On the other hand, the human has the capability to locate the victim, and extracting him/her to the evacuation area, but doesn't have the capability to deal with dangerous situations that he/she might encounter.

The problem has been modelled with a known map topology as depicted in Figure 2. The topology of this map represents the connectivity between major intersection areas "nodes", and it's assumed to be known ahead. Only one victim is expected to be in the search and rescue area, but it can be at any of the nodes. Also, only one source of danger is expected to be present at any of the map nodes. The multi-agent team will start from different locations in an attempt to locate the victim, and extract him/her to the safe evacuation location.

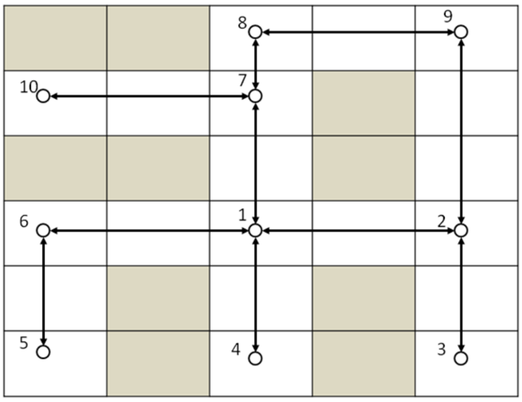

Fig. 2. The figure shows a 6 × 5 domain which shows a discretization of environment through a topological map with 10 nodes and their connections.

VI. POMDP MODEL

The problem is modelled in a two slice Dynamic Bayesian network (DBN) shown in Figure 3.

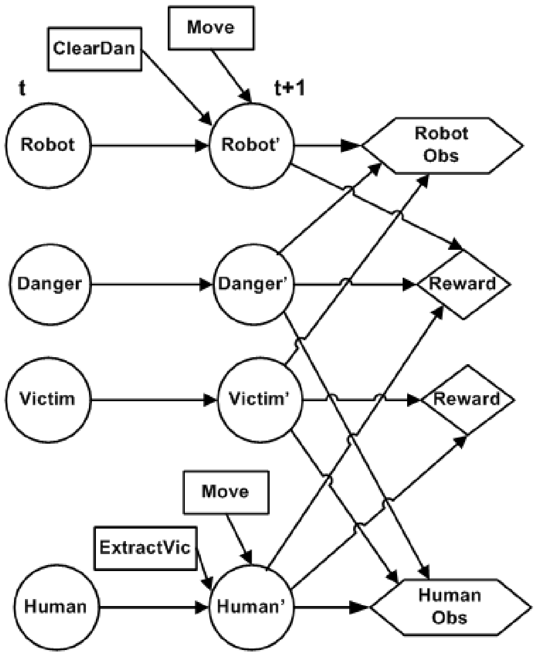

Fig. 3. POMDP representation using DBN

A. States

The state space for this problem is the cross product between the location of each of the 3 agents (2 robots and one human), as depicted earlier in the topological map in Figure 2, and the possible locations of the victim and the danger. The state space is described in Table I.

TABLE I. STATE SPACE OF THE PROBLEM MODEL

Variable Type	$(Quantity)$ Name	Domain
State	(2) Robot	10 nodes
State	(1) Human	10 nodes
State	(1) Victim	10 nodes, victimExt
State	(1) Danger	10 nodes,dangerClea
Observation	(1) RobotObs	vicDan, vicNoDan, noVicDan, noVicNoDan
Observation	(1) HumanObs	vicDan, vicNoDan, noVicDan, noVicNoDan

B. Actions

This represent the set of joint actions that the agents can perform to achieve the common objectives. In our model, the robotic agents can perform the following set of actions: $A_r = \{Up, Down, Right, Left, Stop, ClearDanger\}$. The first five actions in this set A_r are related to the robot's navigational abilities, while the last action can be performed to eliminate the source of danger in a map node/location once identified/observed. The human can perform the following actions: $A_h = \{Up, Down, Right, Left, Stop, ExtractVictim\}$, the first five actions are similar to the robot's action and are liked to the navigational actions. The last action will be performed by the human once the victim has been localized, and when the human is in that particular location.

C. Observations

Once in a certain state/node, the agents can observe the presence of either a victim or a danger in the current node/location. This set is the cross product of seeing a victim or a danger in a certain location: $O = \{vicDan, vicNoDan, noVicDan, noVicNoDan\}$, where $vicDan$ encapsulates observing a victim and a danger in the same state, $vicNoDan$ observing victim and no danger, $noVicDan$ observing no victim but danger, and $noVicNoDan$ not observing danger nor a victim.

D. Transition

To simplify the problem, it's reasonable to assume that the navigational actions are deterministic (they lead to only one state). It's also reasonable to assume that the robots are equipped with the right tools that enable them to deal with dangerous situations very effectively. The human in our case is capable of extracting the victim safely once it's been localized.

E. Reward Function

As shown in Table II small steps are penalized. Human is penalized for entering a node containing danger -50. The robots performing a $ClearDanger$ action that yields a state to transition from danger to no-danger are rewarded +50. The human executing the extract victim action is rewarded +100.

TABLE II. REWARD FUNCTION

Action	Reward
Small steps	-1
Human enter node that contain danger	-50
Robot clear danger	+50
Human extract victim	+100

VII. Solving the Model

The proposed scenario has been modelled to represent a search and rescue environment involving two robots and one human. The problem has been modelled with a known 6x5 topological map with 10 nodes. The model is then solved using Multi-agent Decision Process (MADP) toolbox version 0.3 [14]. The toolbox provides a number of solvers such as Joint Equilibrium based Search for Policies (JESP) [15], Generalized Multi-Agent A* and Incremental Clustering and Expansion (GMAA-ICE)[16]. The generated policy is then used online to determine the individual actions that each agent should perform given a sequence of observation histories.

VIII. Conclusions and Future Work

In order to handle the decision making with uncertainty problem in the context of multi-agent human-robot collaboration, this paper presented a human-robot collaboration framework for a search and rescue scenario. Robots and humans are treated as intelligent agents that share observations. Joint actions are taken based on Decentralised Multi-agent Partially Observable Markov Decision Processes (Dec-MPOMDP).

The proposed collaboration framework will be further developed and evaluated in the future in simulated and real environments. Simplifying the modelling process and allowing a domain expert to handle this task instead of a robotics expert will be closely explored. A survey that focuses on evaluating and analysing the success of the collaboration from the human responder's perspective will be also conducted to better enhance our modelling and provide a more natural interaction.

References

[1] G. Randelli, L. Iocchi, D. Nardi, User-friendly security robots, in Proc. 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2011, pp. 308-313.

[2] J. Clark and R. Fierro, Cooperative hybrid control of robotic sensors for perimeter detection and tracking, in American Control Conference, (Portland, OR), June 2005, pp. 35003505.

[3] E. Guizzo, Robot Vacuum Sucks Up Radiation at Fukushima Plant, Internet: http://spectrum.ieee.org/automaton/robotics/industrial-robots/robot-vacuum-sucks-up-radiation-at-fukushima, July 7, 2011 [March 13, 2013].

[4] From Internet to Robotics - A Roadmap for US Robotics 2013 edition, http://www.robotics-vo.us, March 20, 2013 [April, 1. 2013].

[5] F. Driewer, M. Sauer, and K. Schilling, Discussion of challenges for user interfaces in human robot teams, in Proc. The Third European Conference on Mobile Robots, 2007.

[6] G. M. Krujff, V. Tretyakov, T. Linder, F. Pirri, M. Gianni, E. Pianse, S. Carro, P. Papadakis, M. Pizzoli, and A. Sinha,,Rescue Robots at earthquake-hit mirandola, Italy: a Field Report, in Proc. IEEE International Symposium on Safety, Security, and Rescue Robotics, 2012.

[7] T. Fong, C. Thorpe, C. Baur, Multi-robot remote driving with collaborative control, IEEE Transactions on Industrial Electronics, vol. 50(4), pp. 699-704, 2003.

[8] I. R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, Human-robot teaming for search and rescue, in Proc. IEEE Pervasive Computing, 2005, pp. 72-79.

[9] J. Renoux, A. Mouaddib, and S. LeGloannec. "Distributed decision-theoretic active perception for multi-robot active information gathering." In Modeling Decisions for Artificial Intelligence, 2014, pp. 60-71.

[10] J. Capitan, M. T.J. Spaan, L. Merino and A. Ollero. "Decentralized multi-robot cooperation with auctioned POMDPs". The international Journal of Robotics Research, vol.32, pp. 650-671, 2013.

[11] D. V. Pynadath and M. Tambe. "The Communicative Multiagent Team Decision Problem: Analyzing Teamwork Theories and Models". Journal of Artificial Intelligence Research, vol 16, pp. 389-423, 2002.

[12] M. T.J. Spaan, F. A. Oliehoek and N. Valssis. "Multiagent planning under uncertainty with stochastic communication delays". In Proceedings of he Eighteenth International Conference on Automated Planning and Scheduling (ICAPS), vol 8, 2008, pp.338-345.

[13] F. A. Oliehoek, M. T.J. Spaan and N. Valssis. "Optimal and approximate Q-value functions for decentralized POMDPs". Journal of Artificial Intelligence Research, vol 32, pp. 289-353, 2008.

[14] F. A. Oliehoek and M. T.J. Spaan. "MADP Toolbox 0.2." (2009).

[15] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella. "Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings." In IJCAI, pp. 705-711. 2003.

[16] M. TJ Spaan, F. A. Oliehoek, and C. Amato. "Scaling up optimal heuristic search in Dec-POMDPs via incremental expansion." In IJCAI Proceedings of International Joint Conference on Artificial Intelligence, vol. 22, no. 3, p. 20-27. 2011.