Where do you believe you are going, human? Toward active intention recognition in unknown environments for joint robot-human search and rescue missions*

Dimitri Ognibene and Lorenzo Mirante

***Abstract*— Search and rescue missions are particularly dangerous tasks for first responders. Fully automating even the riskiest and earliest phases is still hindered by limited autonomous exploration algorithms.**

This work proposes a first step toward extending robot use in search and rescue missions by allowing them to anticipate human first responder actions, thus enabling exploration focused on areas more relevant for the responder and thus for the task.

The robot controller implements an active intention recognition paradigm to perceive, even under sensory constraints, not only the target position but also the first responders movements which can inform on her intentions: reaching the position where she expects the target to be. This is implemented using an extension of MonteCarlo based planning where reward is extended with an entropy reduction bonus.

I. INTRODUCTION

Search and rescue missions are particularly dangerous tasks for first responders. Fully automating even the riskiest and earliest phases is still hindered by limited autonomous exploration algorithms.

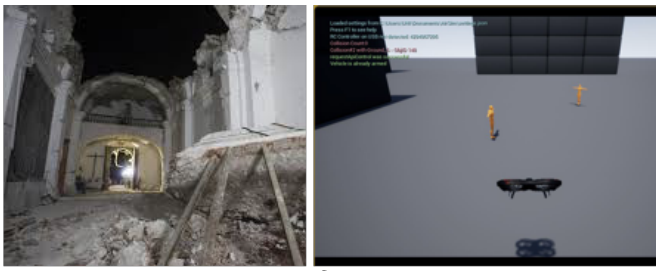
Human first responders rely not only on the flexibility of their perceptual skills but on non-verbal task knowledge accumulated over exploration of multiple structured and unstructured disaster environments which is difficult to transfer to a robot.

In this task we propose that using active intention recognition paradigm, robots can indirectly and naturally exploit first responders knowledge to improve its exploration performance. The intuition is that the responder may have better guess about where the target can be and start moving to reach it. Their initial movements, together with some partial information on the map (e.g. information on regions of interests), can be enough to guess where the responder is directed and thus anticipate her and test the location and secure the path of interest.

II. BACKGROUND

A. Search and rescue

Robot disaster responders are an important research target [1]. Disaster responders aim to quickly locate and extract survivors from dangerous environments. This has supported the study of the non verbal interaction strategies adopted by human responders [2] and the development of collaboration algorithms for group of robots [3], [4], [5]. Other works try to cope with the perceptual complexities of such unstructured environments where even detecting the targets and other responders is difficult. In this line of research we try to bridge these problems while finding computationally affordable solutions.



a b

Fig. 1: **a)** A real disaster condition **b)** Asimulated environment view generated by AIRSIM for testing our algorithm. The figure shows the drone, the responder and the target (e.g. a survivor in a disaster). The responder is supposed to be able easily guess a good location where the target can be and slowing approaching it. The drone enters the disaster from another position, has initial knowledge of the entrance location of the responder and a set of candidate areas where the target can be located. The drone, due to the complexities of the environment e.g. occlusions, can see the responder or the target only when it is close to it. It must employ a policy that can take advantage of the presence of the responder to find the target as fast as possible.

B. Planning in partial observable conditions

In this work we adopt the Markov Decision Processes (MDP) formulation for planning problems which allows a neat formulation that account for uncertainty and noise. The most common formulation describes the problem as finding a policy that maximizes the expected cumulative discounted reward given a representation of the environment in terms of a set of states and actions, a transition function describing the probability of reaching a certain state once an action is performed in a state, and reward function describing the reward expected once an action is performed in a state. More

[1]DO and LM are with the Department of Computer Science and Electronic Engineering, University of Essex, Colchester, UK dimitri.ognibene@essex.ac.uk

Workshop on Human-Aiding Robotics
IROS 2018 - International Conference on Intelligent Robots 2018, Madrid, Spain

realistic setups require to model lack of sensory information about the environment state, partially observable environments. This are particularly suitable to describe search and rescue environments where the positions of the targets or the responder are unknown and they are detectable only when the robot is in their proximity. These problems are described extending MDP with an observation model that describes the probability of obtaining a certain observation when the environment and the sensors are in certain configurations. To solve these problems we follow the Monte-Carlo (MC) approach proposed in [6].

C. Intention Recognition

Intention recognition supports natural, proactive collaboration between agents. Several algorithms have been proposed, some based on the idea of using pre-baked libraries of actions that suit a specific condition, others rely on the idea of inverse planning and allow to recognise others' intentions given a model of the environment that permits to evaluate the effectiveness of a sequence of actions to achieve any possible goal[7], [8]. This allows more precise and flexible recognition but has higher computational cost. Other approaches, particularly suitable for robotics application use the idea of intention recognition as inverse control, were multiple parametrizable controllers are compared to explain others' behaviours [9], [10].

D. Active Perception and Vision

Is the problem of actively controlling own sensors to improve speed and accuracy of behaviourally relevant variables [11], [12]. It has relevance not only for robotics but also for neuroscience and cognitive science [13]. Different approaches have been proposed, e.g. based on neural models often trained by RL [11], [12] or evolutionary algorithms [14], or information theoretic approaches [15], [9]. It is applied in different fields, such as inspection, localization, object recognition and autonomous driving.

E. Active intention recognition

Active intention recognition has been recently introduced [9], [16] as the problem of recognising other agents intentions when the robot sensor can be actively controlled to improve speed and accuracy of activity, intention or action recognition. This can be important as perceiving the other agents and their targets at the same time with fixed sensors can be impossible. It is a particularly challenging problem because it requires to perceive the possible interactions of multiple elements (e.g. affordances and effectors) which may not be known or observed simultaneously (e.g. watching a mouse running we may not see the cheese or not know that a cat is running after it). It is formally defined as selecting the policy of sensor control that minimize the final expected entropy on the state of the observed agents. Ognibene & Demiris [2013] presented an implementation on a humanoid robot based on the concept of intention recognition as inverse control. The formulation used a mixture of kalman filters to represent the observed agent possible movements, assuming they were generated by one of multiple parametrizable linear controllers, and using a single step myopic strategy based on a novel approximation of the information gain on the mixture selector variable. In Lee et a.l [2015] a similar concept was used to simultaneously recognize the complex activities of multiple actors which were represented using Probabilistic Context Free Grammars.

In this work we will extend Ognibene & Demiris [2013] relaxing the myopic constraint to consider multistep policies in discrete environments.

III. Methods

A. Planner

The controller implements a version of Monte Carlo Tree Search (MCTS) based on UCT (Upper Confidence Bounds applied to trees) [17] exploration strategy extended to partially observable environments. The algorithm is inspired from POMCP (Partially Observable Monte-Carlo Planning) [6] however it does not use a particle filter, instead it uses a complete discrete Bayesian filter. We expect that this helps dealing with the limited amount of information provided for most of the task that may strongly affect the particle filter representation. Also the selected representations for the robot belief state could be more effective for entropy estimation, which usage is the second difference with POMCP (however check [18], [19]). Inspired by the work in Ognibene & Demiris [2013] we consider as an additional component of the reward the entropy on the state of the environment (which in this case is a proxy of the uncertainty on the target). This should strongly improve performance in large environments despite reduced maximum exploration depth. Still it requires some additional computation (e.g. the update of the belief state during the rollout phase).

The POMCP algorithm has been described in [6] and already applied in human robot interaction in [20]. For space reasons we refer to these publications for a detailed description of the algorithm. The only implementation difference is that the observation is used to update the filter state and that the updated filter state is used to compute the entropy used to augment the reward. Given its computational cost, entropy is stored as part of the reached node.

A final difference is that the state of the drone is considered to be known exactly.

B. Environment

The problem is described according to the POMDP formalism, thus containing a Transition, Observation and Reward functions and also an end function was used as adopted in other POMDP problems with MC methods.

1) Transition function: The environment is static a part from the responder movements. The transition function is thus defined using a simple, deterministic, model of the drone movements and a probabilistic model of the movement of

the responder when searching and approaching the target. The drone is supposed to not be affected by obstacles while the responder is. The state is composed by a 6 dimensional tuple, containing the position of the drone, the position of the responder and that of the target. The drone moves in the direction of the target according to the following stochastic rule:Where c_h^t is the cell of the grid where the responder

c_h^{t+1}	$p\left(c_h^{t+1}\|c_h^t\right)$
c_h^t,	p_{still} or 1 if $c_h^t \equiv p_t$
$c_h^t + d$,	$\frac{\sqrt{\left(1+0.95\, sign(\Delta_x^t \cdot d_x)\right)\left(1+0.95\, sign(\Delta_y^t \cdot d_y)\right)}}{Z}$ and $0 < \|\|d\|\| < 1.5$
otherwise,	0

is located at time t. $p_{still} = 0.6$ is the probability of the responder to stay in the same cell while approaching the target. $d = c_h^{t+1} - c_h^t$ is the responder vector movement, i.e. how many cells have changed in x, y directions (d_x, d_y). $\Delta^t = c_t - c_h^t$ is the distance vector between the target and the responder. Z is the normalization factor.

2) Observation model: The observation model always provides the position of the drone while it would give the target or the responder position only if they were in the same position of the drone.

3) Reward function: In both [9], [16], the expected information gain on the mixing variable of the mixture model was derived and computed to drive active intention recognition. In this work we integrate more closely active intention recognition and proactive collaboration. The main reward for the system is thus for finding the target, the joint goal with the responder. This does not mean that the robot is blind to the responders actions, in fact seeing the responder in a certain position of the map, or even not finding her, still provides information that a POMDP formulation can exploit. Still practically exploiting this information can be difficult due to computational limits. E.g. if the planning process stops before finding the target, even after acquiring some useful information, no reward will be obtained and thus no preference for such action versus useless ones (e.g. staying still at the entrance) will be obtained. Thus in this work we also tested a source of intrinsic reward [21], [22] based on information acquisition, or uncertainty reduction, that could support the planning and exploration process.

The predicted next belief states sampled by the MC planning process are exploited to estimate the uncertainty reduction produced by a certain course of action. This is done by computing directly the entropy of the sampled belief state and adding it to the reward function. Many other ways to stimulate exploration can be devised but they may lead to suboptimal behaviours. E.g. giving a limited reward for visiting new states may lead to a policy that favours longer trajectories to reach the target. In this work we tested how to improve exploration performance by rewarding responder localization. The idea is that observing the responder may give a good guess on the target location, but focusing on tracking it may be suboptimal, e.g. if the drone is nearer to the possible target locations. Entropy instead values the information obtained even when the responder is not found.

4) End function: End function stopped both the real world simulation and the simulation inside the MC sampling algorithm when in the real state or the sampled one, respectively, the drone was near the target.

Current implementation uses Python 3.6.

TABLE I: Performance with 1000 MC samples.

1000 Samples				
	ER	RR	SER	STAR
Success	0.85	0.88	0.54	**0.95**
Steps	7.96	8.43	1.033	**7.91**
responder Obs	31	37	148	10

Average behavior over 100 trials with 1000 MC samples for each step. Success is the average number of times the drone found the target before 16 actions. Steps is the average number of steps necessary to the drone to find the target (or 16 if it fails). responder Obs total number of times the drone met the responder before finding the target.

IV. Results

A. Parametrizations and variations of the system

We present preliminary results on the following variations of the architecture:

- STAR reward was given only reaching the environment (1)
- RR reward was given also when observing the responder (0.1)
- ER reward contained an entropy bonus (-H*0.2)
- SER reward contained an entropy bonus only during the search not during rollout (-H*0.2)

We also varied the number of samples (100-500-1000) and the max-search-depth (6-10-14).

B. Environments

We considered a small and a big environment. The **small environment [SE]** was a 5x5 grid. The drone was initially positioned in the center and the responder was in one of 2 adjacent cells, but the drone was not informed of which of them. The target was positioned in one of the 4 corners. In total 8 equally probable initial conditions were possible, representing the initial belief state of the agent in this environment.

Another setup, **large environment [LE]** was a 10x10 grid. Again the drone was initially positioned in the center and the responder was in one of 4 adjacent cells, but the drone was not informed of which of them. The target was positioned in one of the 16 cells in the 4 corners, 4 cell in square per corner. In total 64 equally probable initial conditions were possible, representing the initial belief state of the agent in this environment.

The result for the small environment with 1000 samples and max-depth-search=14 (average on 100 trials) are reported in table I. When computational resources are available the STAR algorithm, not using the computation of

TABLE II: Performance with 100 MC samples.

100 Samples				
	ER	RR	SER	STAR
Success	**0.8**	0.63	0.25	0.63
Steps	**8.31**	9.12	11.58	9.74
responder Obs	20	24	29	13

Average behavior over 100 trials with 100 MC samples for each step. Success is the average number of times the drone found the target before 16 actions. Steps is the average number of steps necessary to the drone to find the target (or 16 if it fails). responder Obs total number of times the drone met the responder before finding the target.

entropy but just maximizing the search of the target gave the best performance. In this case we must note that it overcomes the trivial solution of navigating tRRough the for corners which would have an expected number of steps (2+7+11+15)/4=8.75. This is true also for the ER and RR.

ER and RR also observe the responder much more often than STAR, still STAR performs better than them and of the best uninformed algorithm, thus it makes good use of the observations of the responder (both when he finds it in a cell and when it doesn't).

We observed, without a statistical analysis, that the policy usually chosen by the STAR agent often consisted in (a) reaching a corner, (b) going back to the center if no target was observed, and (c) going to the next corner. This strategy may maximize the probability of encountering the responder or not meeting it which would decrease the probability of the target to be in that direction.

Unexpectedly the SER version of the system presented lower performance. Table II reports the results when only MC 100 samples per step were performed.

In this case the ER system was the more effective followed by RR and STAR. ER is the least affected by the reduced number of samples. This suggest that using and entropy based bonus may support exploration of environments when not enough time is available for extensive computations.

With the big environment and 100 samples per action the best performance were achieved by the ER who found the target 50% of the trials in less than 40 steps (maximum steps allowed). SER achieved (28%), RR 27 % and STAR 18%.

V. Conclusions and future work

Current result suggest that the idea of coupling an artificial agent with a more expert but constrained human partner in the task of joint search for a target in an unknown environment may be promising. Future work will focus on optimizing the algorithms, integrating with flexible SLAM algorithms, devising and testing alternative exploration bonus, employing realistic responder models, considering additional collaborative tasks, such as verifying that the surrounding of the responder are safe, and more.

References

[1] T. Kruijff, P. Linder, P. Gianni, S. Pizzoli, and C. Pianese, "Rescue robots at earthquake-hit mirandola, italy: a field report," in *Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on, IEEE*, 2012, pp. 1–8.

[2] F. Bacim, E. D. Ragan, C. Stinson, S. Scerbo, and D. A. Bowman, "Collaborative navigation in virtual search and rescue," in *3D User Interfaces (3DUI), 2012 IEEE Symposium on*. IEEE, 2012, pp. 187–188.

[3] Z. Beck, L. Teacy, A. Rogers, and N. R. Jennings, "Online planning for collaborative search and rescue by heterogeneous robot teams," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2016, pp. 1024–1033.

[4] O. Gigliotta, "Equal but different: Task allocation in homogeneous communicating robots," *Neurocomputing*, vol. 272, pp. 3–9, 2018.

[5] F. Broz, C. Nehaniv, T. Belpaeme, A. Bisio, K. Dautenhahn, L. Fadiga, T. Ferrauto, K. Fischer, F. Frster, O. Gigliotta, S. Griffiths, H. Lehmann, K. Lohan, C. Lyon, D. Marocco, G. Massera, G. Metta, V. Mohan, A. Morse, S. Nolfi, F. Nori, M. Peniak, K. Pitsch, K. Rohlfing, G. Sagerer, Y. Sato, J. Saunders, L. Schillingmann, A. Sciutti, V. Tikhanoff, B. Wrede, A. Zeschel, and A. Cangelosi, "The italk project: A developmental robotics approach to the study of individual, social, and linguistic learning," *Topics in Cognitive Science*, vol. 6, no. 3, pp. 534–544, 2014, cited By 4.

[6] D. Silver and J. Veness, "Monte-carlo planning in large pomdps," in *24th Advances in Neural Information Processing Systems,(NIPS 2010)*, 2010, pp. 2164–2172.

[7] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," *Nature Human Behaviour*, 2017.

[8] M. Ramirez and H. Geffner, "Plan recognition as planning," in *Proc. 21st Intl. Joint Conf. on Artificial Intelligence (IJCAI)*, 2009.

[9] D. Ognibene and Y. Demiris, "Towards active event recognition," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI Press, 2013, pp. 2495–2501.

[10] Y. Demiris, "Prediction of intent in robotics and multi-agent systems," *Cognitive Processing*, vol. 8, no. 3, pp. 151–158, 2007.

[11] D. Ognibene and G. Baldassare, "Ecological active vision: Four bioinspired principles to integrate bottom–up and adaptive top–down attention tested with a simple camera-arm robot," *Autonomous Mental Development, IEEE Transactions on*, vol. 7, no. 1, pp. 3–25, 2015.

[12] N. Sprague and D. Ballard, "Eye movements for reward maximization," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2004.

[13] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cognitive neuroscience*, vol. 6, pp. 1–28, 2015.

[14] G. de Croon, "Adaptive active vision," Ph.D. dissertation, Universiteit Maastricht, 2008.

[15] J. Denzler and C. Brown, "Information Theoretic Sensor Data Selection for Active Object Recognition and State Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 145–157, 2002.

[16] K. Lee, D. Ognibene, H. J. Chang, T.-K. Kim, and Y. Demiris, "Stare: Spatio-temporal attention relocation for multiple structured activities detection," *Image Processing, IEEE Transactions on*, vol. 24, no. 12, pp. 5916–5927, 2015.

[17] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning." Springer, 2006, pp. 282–293.

[18] M. Lauri and R. Ritala, "Planning for robotic exploration based on forward simulation," 2016.

[19] Y. Boers, H. Driessen, A. Bagchi, and P. Mandal, "Particle filter based entropy," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[20] A. Goldhoorn, A. Garrell, R. Alquezar, and A. Sanfeliu, "Continuous real time pomcp to find-and-follow people by a humanoid service robot," in *2014 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids) November 18-20, 2014. Madrid, Spain*. IEEE, 2014.

[21] G. B. . M. Mirolli, Ed., *Intrinsically Motivated Learning in Natural and Artificial Systems*. springer, 2014.

[22] M. G. Bellemare, S. Srinivasan, G. Ostrovski, T. Schaul, D. Saxton, and R. Munos, "Unifying count-based exploration and intrinsic motivation," *arXiv preprint arXiv:1606.01868*, Jun. 2016.