A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering
Jennifer Renoux, Abdel-Illah Mouaddib, Simon Le Gloannec

To cite this version:
Jennifer Renoux, Abdel-Illah Mouaddib, Simon Le Gloannec. A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering. Workshop on Multiagent Sequential Decision Making Under Uncertainty., May 2014, Paris, France. hal-01096027

HAL Id: hal-01096027
https://hal.archives-ouvertes.fr/hal-01096027
Submitted on 16 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering

Jennifer Renoux, Abdel-Ilah Mouaddib
GREYC, University of Caen
Caen, France
{jennifer.renoux, abdel-illah.mouaddib}@unicaen.fr

Simon Le Gloannec
Cassidian SAS
Val de Reuil, France
simon.legloannec@cassidian.com

ABSTRACT

Multirobot systems have made tremendous progress in exploration and surveillance. In that kind of problem, agents are not required to perform a given task but should gather as much information as possible. However, information-gathering tasks usually remain passive. In this paper, we present a multirobot model for active information gathering. In this model, robots explore, assess the relevance, update their beliefs and communicate the appropriate information to relevant robots. To do so, we propose a distributed decision process where a robot maintains a belief matrix representing its beliefs and beliefs about the beliefs of the other robots. This decision process uses entropy and Kullback-Leibler in a reward function to access the relevance of their beliefs and the divergence with each other. This model allows the derivation of a policy for gathering information to make the entropy low and a communication policy to reduce the divergence. An experimental scenario has been developed for an indoor information gathering mission.

Categories and Subject Descriptors
Computing Methodologies [Artificial Intelligence]: Planning and scheduling—Planning under uncertainty

General Terms
Design, Algorithm

Keywords
Active Information Gathering, Information Relevance, Decentralized Decision Process, Multiagent System

1. INTRODUCTION

Robotic systems are increasingly used in surveillance and exploration applications. In the future robots may assist humans and eventually replace them in dangerous areas. In these particular research fields the main goal of the system is not to reach a given goal but to gather information. The system needs to create a complete and accurate view of the situation. This built view may be used afterwards by some agents - human or artificial - to make some decisions and perform some actions. Therefore the information gathering system must be able to identify lacking information and take the necessary steps to collect it. However it is obviously not productive that all the robots in the system try to collect all possible information, just as it is not possible for the robots to communicate all the information they have all the time. They should select pieces of information to collect or to communicate depending on what they already know and what other agents already know. Developing methods to allow robots to decide how to act and what to communicate is a decision problem under uncertainty.

Partially Observable Markov Decision Processes (POMDPs) are traditionally used to deal with this kind of problem, more particularly Decentralized POMDPs (DEC-POMDPs) that are an extension of POMDPs for multiagent systems. However the classic POMDP framework is not designed to have information gathering as a target: information gathering is usually a means of reaching another goal. Some extensions have been developed for mono-agent systems but the plunge to multi-agent systems has not been taken. We suggest in this paper a formal definition of the relevance of a piece of information as well as a new model dedicated to multiagent information gathering that is able to explore actively its environment and communicate relevant information.

Section 2 presents some background knowledge and other studies that are relevant to our problem. Section 3 presents the proposed model to do active sensing with a multi-robot system. It defines a agent-oriented relevance degree and describes the Partially Observable Markov Decision Process used in the system. Finally, section 4 presents an implementation of the model on a simple indoor sensing problem.

2. BACKGROUND AND RELATED WORK

Relevance

Agents situated in an environment perceive a huge amount of data and they need to process those data to extract higher-level features. However the interest of a feature for an agent depends on several parameters such as the situation, the problem to be dealt etc. Since it is counterproductive to communicate non-interesting information, or to perform an action to collect it, agents need to quantify the importance of a piece of information according to the current situation. This degree of importance is the relevance of information. Borlund [4] defined two types of relevance: system-oriented relevance and agent-oriented relevance. System-oriented relevance analyzes relevance according to a request. The better the match between information and the request, the greater the degree of relevance. System-oriented relevance is used and studied in several Information Retrieval Systems.
Agent-oriented relevance defines a link between information and an agent’s needs. Information is relevant if it matches a specific need, if it is informative and useful for an agent which receives it. However the need may not be explicit. Floridi [6] suggested a base of epistemic relevance and Roussel and Cholvy [16] deepened this study in the case of BDI (Belief-Desire-Intention) agents and multimodal logic. However those studies are based on propositional logic and are not applicable for reasoning with uncertain knowledge. Therefore we need to define a relevance theory that may be used in uncertain reasoning.

**Active information gathering**

Using relevance, an agent is able to decide if a piece of information is interesting or not. Therefore it is able to perform active information gathering. Active information gathering defines the fact that an agent will act voluntarily to gather information and not just perceive passively its environment. In this context the agent has to make decisions in an environment that it cannot perceive completely. One of the best and commonly used models to deal with that kind of problem is the Partially Observable Markov Decision Process. Some studies have already been carried out to perform active perception using POMDPs. Ponzoni et al. [11] suggested a mixed criterion for active perception as a mission to recognize some vehicles. The criterion is based on an entropy function and a classical reward function to mix active perception and more classical goals of POMDPs. Meanwhile Araya-Lopez et al. [1] suggested a more general approach to use any reward function based on belief state in POMDPs. These two approaches proved the feasibility of a system where information gathering is the goal. However they are both monagent and are not applicable to a multiagent system. To our knowledge there is no model of a multiagent system for active information gathering.

It is obvious that information gathering would be more efficient if it is done by several agents instead of a single one. However, it is important that agents are able to coordinate themselves to make the gathering efficient.

**Multiagent active information gathering : a multiagent decision problem**

This problem of multiagent active information gathering relates to multiagent decision under uncertainty: a set of agents have to control together a decision process to collect information. However no agent is able to control the whole process. Different frameworks have been developed to deal with multiagent decision problems under uncertainty, such as DEC-POMDP [3], MTDP [14] and POIPSG [10]. All these frameworks are equivalent [20] and are extensions of POMDPs to multiagent systems. They share a common structure and can be described as a tuple 

\[
< I, \mathcal{S}, \{A_i\}, \{O_i\}, T, \omega, R > \text{ with :}
\]

- \(I\) is a finite set of agents
- \(\mathcal{S}\) is a finite set of states systems
- \(A_i\) is a finite set of actions available for agent \(i\). \(\bar{a} = \langle a_1,\ldots,a_n \rangle\) is a joint action and \(\bar{A} = \otimes_i A_i\) is the set of joint actions.
- \(O_i\) is a finite set of observations available for agent \(i\). \(\bar{\sigma} = \langle o_1,\ldots,o_n \rangle\) is a joint observation and \(\bar{O} = \otimes_i O_i\) is the set of joint observations.
- \(T: \mathcal{S} \times \bar{A} \rightarrow \mathcal{S}\) is the transition function
- \(\omega: \mathcal{S} \times \bar{A} \rightarrow \mathcal{O}\) is the observation function
- \(R: \mathcal{S} \times \bar{A} \rightarrow \mathcal{R}\) is the reward function

Solving a multiagent POMDP is a problem NEXP-complete [3]. Even if algorithms and heuristics have been suggested to overcome this complexity [7] [24] [19], those frameworks are usually not applicable to real problems. To overcome this issue, Spaan et al. [23] suggested a system based on POMDP to enable a Network Robot System to classify particular features by acting to get the best information possible. In this study, the authors decided to model the active information gathering by classifying actions. This avoids using a reward function based on entropy, that would increase the complexity of planning. However Araya-Lopez et al. [7] proved that it is possible to reuse techniques from the standard POMDP literature with a reward function based on belief states as would be a reward function using entropy. On top of that, the system suggested by Spaans et al. aims at improving a single agent’s beliefs and does not take into account the other agents of the system. That is to say that all the agents in the system have to make the classification steps and build a complete view of the environment. However, in usual active information gathering problems, the entire system should have a complete view of the environment but it not useful that each agent of the system has this complete view. Therefore, agents need to communicate with each other in order to avoid repetitive exploration. Communication in multiagent POMDP frameworks is usually assumed to be free and instantaneous: at each time step, all the agents communicate the observations they received. However such communication is not possible in real problems. Communication is an action that has a cost and must be decided by the agents. Roth et al. [15] presented an algorithm to take into account the communication cost in multiagent POMDPs. In this paper, the communication is considered only during execution. The multiagent POMDP is approximated using a heuristic, that makes its resolution more scalable. Then, at execution time, an agent compares the action that it should take according to the multiagent resolution and the action that it should take according to its local observation history. If the actions differ, then the agent communicates its history of observations. There is no decision concerning the observations to communicate: when an agent decides to communicate, it communicates all the observations it has received to all other agents. On top of that, communication is used only if it can improve the performance of the system. Information gathering is once again a means to reach a goal and not the goal in itself.

3. THE MODEL

3.1 Definition of an agent-based relevance

Let’s assume an agent \(a_i\) situated in an environment \(\mathcal{E}\). The environment is modeled as a set of features of interest. Each feature is described using a random variable \(X_k\) which can take values in its domain \(DOM(X_k)\).

\[
\mathcal{E} = \{X_k\}
\]

Let \(DOM(\mathcal{E})\) be the set of all possible values for all \(X_k \in \mathcal{E}\).

\[
DOM(\mathcal{E}) = \{DOM(X_k) \forall X_k \in \mathcal{E}\}
\]
The agent \( a_i \) has some beliefs \( B^i_k \) concerning the features of interest modeled as probability distributions over the \( X_k \in \mathcal{E} \).

\[ B^i_{k,t} = \{ b^i_{k,t} \mid X_k \in \mathcal{E} \} \]

with \( b^i_{k,t} \) being the probability distribution of agent \( a_i \) over the variable \( X_k \) at time \( t \). Let’s assume an agent receives observations concerning the features of interest. Possible observations are the possible values of the random variables \( X_k \).

\[ o_k \in \text{DOM}(X_k) \]

When receiving a new observation, agent \( a_i \) updates its beliefs according to it: \( B^i_{k,t+1} = \text{update}(B^i_{k,t}, o_k) \) [5, 22]

First of all we considered that observations received are true. As a matter of fact, an observation cannot be relevant if it is a false observation [6]. We discuss this assumption true. As a matter of fact, an observation cannot be relevant if it matches the following criteria:

1. agent \( a_i \) is interested in the subject of the observation \( o_k \), that is to say \( X_k \)
2. the observation \( o_k \) is new for agent \( a_i \) or
3. if the observation \( o_k \) is not new, it should render agent’s \( a_i \) beliefs more accurate

The first point is dealt with the way we represent agent’s beliefs : if agent \( a_i \) is interested in \( X_k \) then \( X_k \) is in agent’s \( a_i \) beliefs. We assume that an observation \( o_k \) is new for agent \( a_i \) if beliefs \( B^i_{k,t+1} \) and \( B^i_{k,t} \) are distant from each other. The distance between two probability distributions is measured by the Kullback-Leibler ratio.

**Definition 1.** An observation \( o_k \) is new for agent \( a_i \) at time \( t \) if and only if

\[ D_{KL}(B^i_{k,t}||B^i_{k,t+1}) > \epsilon \] (1)

\( \epsilon \) is a fixed threshold and \( D_{KL}(B^i_{k,t}||B^i_{k,t+1}) \) is the Kullback-Leibler ratio and defined by

\[ D_{KL}(B^i_{k,t}||B^i_{k,t+1}) = \sum_{X_k \in \mathcal{E}} \sum_{k=1}^n b_{k,t}^i(x_k) \log \left( \frac{b_{k,t}^i(x_k)}{b_{k,t+1}^i(x_k)} \right) \] (2)

where \( b_{k,t}^i(x_k) \) is the belief of agent \( a_i \) that the random variable \( X_k \) is equals to \( x_k \).

To model the accuracy of a belief \( B^i_{k,t} \), we use an entropy measure.

**Definition 2.** Belief \( B^i_{k,t+1} \) is more precise than belief \( B^i_{k,t} \) if and only if

\[ H(B^i_{k,t+1}) < H(B^i_{k,t}) \] (3)

with \( H(B^i_{k,t}) = -\sum_{X_k \in \mathcal{E}} \sum_{k=1}^n b_{k,t}^i(x_k) \log(b_{k,t}^i(x_k)) \).

Given the previous definitions we may define the degree of relevance as shown below :

**Definition 3.** The degree of relevance of an observation \( o_k \) concerning a random variable \( X_k \) for an agent \( a_i \), noted \( \text{rel}_i(o_k) \), is given by

\[ \text{rel}_i(o_k) = \alpha D_{KL}(B^i_{k,t}||B^i_{k,t+1}) + \beta (H(B^i_{k,t}) - H(B^i_{k,t+1})) + \delta \] (4)

with \( B^i_{k,t+1} = \text{update}(B^i_{k,t}, o_k) \), \( \alpha \) and \( \beta \) being weights and \( \delta \) being a translation factor to ensure the relevance is positive.

### 3.2 Decision Process for multiagent active information gathering

Let a multiagent system be defined as a tuple \( < \mathcal{E}, \mathcal{A}, B, D > \) with

- \( \mathcal{E} \) being the environment as defined previously
- \( \mathcal{A} \) being the set of agents
- \( D \) being the set of all agents’ decision functions

\( D = \{ D_i, \forall i \in \mathcal{A} \} \) is the set of all agents’ decision function. Each \( D_i \) is represented as a Factored Partially Observable Markov Decision Process (FPOMDP)[8].

A POMDP is a tuple \( < S, A, O, \mathcal{T}, \omega, R, b_0 > \) with

- \( S \) being a set of states, corresponding to the joint instantiations of the random variables \( X_k \in \mathcal{E} \).
- \( A \) being a set of epistemic actions [17]
- \( O \) being the set of observations
- \( \mathcal{T} \) being the transition function, defined as in [8]. For more information on how to represent it, see [21] and [13].
- \( \omega \) being the observation function, defined as in [13].
- \( R \) being the reward function
- \( b_0 \) being the initial belief state

**Set of actions**

We consider two type of actions : look for the value of a particular random variable (Explore-type actions) and communicate a set of observations to a set of agents (Communicate-type actions):

\[ A = \{ \text{Exp}(X_k), \forall X_k \in \mathcal{E} \} \cup \{ \text{Comm}(O, Ag), \forall O \subset \mathcal{O}, \forall Ag \subset \mathcal{A} \} \]

The size of the action set is :

\[ |A| = |A_{\text{Explore}}| + |A_{\text{Communicate}}| = |\mathcal{E}| + |A_{\text{Communicate}}| \] (5)

An agent may communicate a non empty subset \( O \subset \mathcal{O} \) of observations to a non empty subset \( Ag \subset \mathcal{A} \) of agents. Considering that the number of subsets of the observation set is \( 2^{|\mathcal{O}|} \) and the number of subsets of the agent set is \( 2^{|\mathcal{A}|} \), we obtain:

\[ |A_{\text{Communicate}}| = (2^{|\mathcal{O}|} - 1) \times (2^{|\mathcal{A}|} - 1) \] (6)

Equation 5 may be rewritten :

\[ |A| = |\mathcal{E}| + (2^{|\mathcal{O}|} - 1) \times (2^{|\mathcal{A}|} - 1) \] (7)

It is obvious that combinatorial explosion will occur even with a small number of observations and agents. It is however possible to reduce the size of the action set with a reasonable hypothesis. We may assume the simple case where an agent communicates only one observation to one other agent. In that case equation 7 becomes

\[ |A| = |\mathcal{E}| + |\mathcal{O}| \times |\mathcal{A}| \] (8)

In the remainder of this paper we will only consider this simplified Communication-type action. However one should note that the model is still valid if one needs to relax the hypothesis.
Set of observations

In a Partially Observable Markov Decision Process, an agent doesn’t know exactly the current state of the system. It only receives observations when performing actions, which are only indications about the current state. So the agent may estimate the current state from the history of observations it has received. When performing an Explore-type action, the agent receives an observation concerning the random variable it is trying to sense. Therefore the possible observations are the possible values the random variable may take, that is to say the domain of the random variable. Therefore the set of all possible observations from an Explore-type action is:

\[ O_{Explore} = \text{DOM}(E) = \bigcup_{\forall X_k \in E} \text{DOM}(X_k) \]

When performing a Communicate-type action the agent receives an observation stating whether the message has been properly sent or not. Therefore two observations are possible for any Communicate-type action:

\[ O_{Communicate} = \{ \text{okMsg}, \text{nokMsg} \} \]

So the entire set of possible observations is

\[ O = O_{Explore} \cup O_{Communicate} = \bigcup_{\forall X_k \in E} \text{DOM}(X_k) \cup \{ \text{okMsg}, \text{nokMsg} \} \]

Maintaining a belief state

The agent doesn’t know the exact current state of the system. It only has observations about it. Therefore it should maintain some beliefs concerning this current state. In the context of multiagent information gathering, an agent should not only have beliefs about the state of the environment but also about the other agents. To prevent agents from exploring the same areas and to enable them to choose the most relevant observation to communicate, they should model the knowledge of other agents in their own belief state. Thus we defined an extended belief state as following:

**Definition 4.** Let a extended belief state of an agent \( a_i \) at time \( t \) be defined as following:

\[ B_{i,t} = B_{i,t}^{\mathcal{E}} \cup \{ B_{i,t}^{\mathcal{E}} \}, j \in \{ \mathcal{A} \} \]

with \( B_{i,t}^{\mathcal{E}} = \{ b_{i,t,k}^{\mathcal{E}}, \forall X_k \in \mathcal{E} \} \) being the beliefs of agent \( a_i \) concerning the environment \( \mathcal{E} \) and \( B_{i,t}^{\mathcal{E}} = \{ b_{i,t,k}^{\mathcal{E}}, \forall X_k \in \mathcal{E} \} \) being the beliefs of agent \( a_i \) concerning the beliefs of agent \( a_j \) concerning the environment.

Let’s note that \( B_{i,t}^{\mathcal{E}} \) is an approximation of \( B_{i,t} \).

To keep an accurate representation of the current state of the system an agent has to update its beliefs regularly. An update will occur in three cases:

1. the agent receives a new observation from its sensors after an Explore action. It updates its own beliefs concerning the environment : \( B_{i,t+1}^{\mathcal{E}} \).
2. the agent receives a new observation from agent \( a_j \). It updates its own beliefs \( B_{i,t+1}^{\mathcal{E}} \) as well as its beliefs concerning agent \( a_j : B_{i,t+1}^{\mathcal{E}} \).
3. the agent sends an observation to agent \( a_j \). It updates its beliefs concerning agent \( a_j : B_{i,t+1}^{\mathcal{E}} \).

In all cases the update \( B_{i,t+1}^{\mathcal{E}} = \text{update}(B_{i,t}^{\mathcal{E}}, o_k, a_k \) being the observation received, is made as usual in Partially Observable Markov Decision Processes:

\[ B_{i,t+1}^{\mathcal{E}}(s') = \frac{\omega(o_k, s', a) \sum_{s \in S} p(s'|s, a)B_{i,t}(s)}{\sum_{s \in S} \sum_{s' \in S} \omega(o_k, s', a)p(s'|s, a)b_{i,t}^{\mathcal{E}}(s')} \]

**Reward function**

\( s \) a matter of fact, The best action to perform at a given time is set by a policy. The optimal policy is computed thanks to the reward function. The reward function defines the reward an agent may receive by performing action \( a \) in state \( s \). However in an information gathering context we are not interested in reaching some special state of the system but gathering and communicating relevant observations. Therefore the reward function is defined on the belief states of the agent and not on the real states of the system. An agent is rewarded if it collects observations that are relevant for itself and if it communicates observations that are relevant for other agents. As mentioned in section 3.1, an observation must be true to be relevant. Since agents only have beliefs concerning the world, they cannot ensure that an observation is true. However they should not exchange observations that reinforce their existing beliefs, regardless of their veracity. Therefore we need to find a compromise between the agent’s belief concerning the observation and the relevance of this observation. To do so we weight the relevance of a given observation by the agent’s belief concerning its truth. This belief is given by the probability of receiving the observation in the state \( s \) considered multiplied by the agent’s belief that the state \( s \) is the current state. The reward function is thus defined as follows:

\[ R(B_{i,t}, Exp(X_k)) = \sum_{s \in S} \sum_{o_k \in O} B_{i,t}(s)\omega(o_k, s, a)\text{rel}_i(o_k) - C_{Exp}(X_k) \]

\[ R(B_{i,t}, Comm(o_k, a_j)) = \sum_{s \in S} B_{i,t}(s)\omega(o_k, s, a)\text{rel}_j(o_k) - C_{Comm}(o_k, a_j) \]

with \( C_{Exp}(X_k) \) and \( C_{Comm}(o_k, a_j) \) the costs of taking the Explore or Communicate action and \( B_{i,t}(s) \) the belief of agent \( a_i \) that state \( s \) is the current state.

**Resolution**

In this POMDP actions are epistemic : they don’t modify the real state of the system. Therefore it is possible to transform our POMDP into a Belief MDP defined as a tuple \( < \Delta, \mathcal{A}, \tau > \) where:

- \( \Delta \) is the new state space. It corresponds directly to the belief state space in the initial POMDP. \( \Delta = \mathcal{B_i} \)
- \( \mathcal{A} \) is the same state of actions as previously defined
- \( \tau \) is the new transition function

**Theorem 1.** The transition function \( \tau \) of the Belief MDP is defined as follows:

\[ \tau(B_{i,t}, a, B_{i,t+1}) = \left\{ \begin{array}{ll} \sum_{s \in S} \sum_{a_k \in \mathcal{A}} \omega(a_k, s, a)B_{i,t}(s) & \text{if } U_t \neq \emptyset \\ 0 & \text{otherwise} \end{array} \right. \]
where \( U_t = \{ o_k \in O \mid B_{i,t+1} = \text{update}(B_{i,t}, o_k) \} \) is the set of all observations enabling the transition from state \( B_{i,t} \) to state \( B_{i,t+1} \), \( \omega(o_k, s_t, a) \) is the observation function of the POMDP and \( B_{i,t}(s_t) \) is the belief of agent \( a_i \) that the current state is \( s_t \).

Proof. If there is no observation such as \( B_{i,t+1} = \text{update}(B_{i,t}, o_k) \), it is not possible to transfer from belief state \( B_{i,t} \) to belief state \( B_{i,t+1} \). Therefore, \( \tau(B_{i,t}, a, B_{i,t+1}) = 0 \). If there exists at least one observation \( o_k \) such as \( B_{i,t+1} = \text{update}(B_{i,t}, o_k) \) we have the following equations:

\[
\tau(B_{i,t}, a, B_{i,t+1}) = P(B_{i,t+1} \mid B_{i,t}, a) = \sum_{o_k \in U_t} P(o_k \mid B_{i,t}, a) = \sum_{s \in S} \sum_{o_k \in U_t} P(o_k \mid s, a) B_{i,t}(s) = \sum_{s \in S} \sum_{o_k \in U_t} \omega(o_k, s, a) B_{i,t}(s)
\]

The value function corresponding to this Belief MDP is defined as following:

\[
V(B_{i,t}) = R(B_{i,t}) + \max_{a \in A} \int_{B_{i,t}'} \tau(B_{i,t}, a, B_{i,t}') V(B_{i,t}') \tag{12}
\]

Using discretization techniques on the probability distributions, we may transform equation 12:

\[
V(B_{i,t}) = R(B_{i,t}) + \max_{a \in A} \sum_{B_{i,t}' \in \text{Samples}} \tau(B_{i,t}, a, B_{i,t}') V(B_{i,t}') \tag{13}
\]

Then, any technique from the literature may be used to solve this belief-MDP [12] [9].

4. EXPERIMENTS

Simulated Robots

The suggested model was implemented on a simple scenario. Two robots have to explore an environment made of 4 different zones connected to each other. For each zone, two observations are possible: emptyRoom and notEmptyRoom. The optimal policy was computed using different ratios to make the compromise between the Kullback-Leibler ratio and the entropy measure, as well as different probabilities of obtaining a false observation. The system was run 50 times with each set of parameters. To evaluate the policy, we measured the average number of messages sent by the robots, the average time needed to get a stable belief state and the number of false belief states at the end of exploration. We compared those measures with a multirobot system without communication and with a system in which agents communicate each observation they received immediately. The results are presented in figure 1.

The linear part of the first two graphs represents the average time to reach a stable belief state, measured in number of iterations, an iteration being made up of the execution of an action, the reception of the associated observation and the reception of a communication message from another agent, if any. The bar part represents the average number of final states that were partially or totally incorrect. First of all, we notice that the communicating system (Figure 1a) takes around 18% longer than the explore-only system (Figure 1b) to reach a stable belief state. We did not measured the time
needed for the fully communicative (Figure 1c) system because it is not meaningful. Indeed agents are communicating each observation they receive and we consider communication as a separate action, so this system is much longer than the others. However, we notice that the number of false end states is reduced with partial and complete communication. In the worst case, that is to say a probability of 70% of receiving a correct observation after doing an explore action, the system with partial communication reduces the average number of false end states by 28%, and the system with complete communication reduces it by 8%. However it is possible to configure the Kullback-Leibler / Entropy ratio so that this amount increases to 50% for the system with partial communication (by considering Kullback-Leibler and Entropy with the same weights). In the average case, that is to say a probability of 80% of receiving a correct observation, the average number of false end states is reduced by 72% with our system and by 19% with a complete communication. On top of that, Figure 2a shows that the average number of messages sent remains almost constant and much lower than the system with complete communication (Figure 2b).

Those experiments seem to validate the hypothesis that choosing relevant information to communicate may improve system performances while reducing the number of communications.

Real Robots

We implemented the model on two \( \mu \)-troopers in a simple scenario where two rooms must be explored. In the figure 3, robot 1 decided to explore room 1 and robot 2 decided to explore room 2. Since beliefs of robot 2 about the environment are very accurate and it believes that robot 1 has incorrect beliefs, robot 2 decides to communicate the observation 0 to robot 1. Robot 1 receives this observation and updates its beliefs accordingly. This figure presents a case where an agent has approximated beliefs concerning the beliefs of the other agents. However this approximation does not prevent the robots from completing the mission and reaching a stable belief state, as presented on Figure 4.

5. CONCLUSION AND PROSPECTS

We have introduced a new model of agent-based relevance as well as a decision process to perform active information gathering with a multiagent system. Each agent computes the relevance of an observation regarding itself or another agent to decide whether it should explore a particular zone or communicate this observation. The relevance of an observation is a compromise between the novelty, modeled by Kullback-Leibler ratio, and the certainty of an observation, modeled by Entropy measure. Therefore it may be tuned depending on the environment considered. In a static environment, as presented in the experiments, the certainty of an observation is more important than its novelty. However, in a highly dynamic environment, the novelty of an information may be the most important. The system has been implemented and tested on real robots. Results show that this approach is more efficient than a fully-communicating system.

The decision process we described focuses on relevance and reasoning on belief states to perform active information gathering. In the system presented in this paper, an agent is able to communicate any observation from the observation set if it is relevant. Future works would maintain a history of observations received and allow
an agent to communicate only observations it has previously received. Moreover, the beliefs about the beliefs of other agents are updated only when there is an explicit communication. We plan to work on a less naive method: since the same policy is used by all agents, we may update those beliefs more often by assuming the action taken by other agents. Finally, future works would consider the integration of the system presented in non-epistemic POMDPs.

6. REFERENCES