Adaptive Action Selection of Body Expansion Behavior in Multi-Robot System using Communication

Tomohisa Fujiki [a,1], Kuniaki Kawabata [b] and Hajime Asama [c]
[a] *School of Engineering, The University of Tokyo*
[b] *RIKEN*
[c] *RACE, The University of Tokyo*

Abstract. In multi-robot system, cooperation is needed to execute tasks efficiently. The purpose of this study is to realize cooperation among multiple robots using interactive communication. An important role of communication in multi-robot system is to make it possible to control other robots by intention transmission. We consider that multi-robot system can be more and more adaptive by treating communication as action. In this report, we adopt action adjustment function to achieve cooperation between robots. We also run some computer simulations of collision avoidance as an example of cooperative task, and discuss the results.

Keywords. Q-learning, Multi-Robot System, Communication, Cooperation, Mobile Robot

1. INTRODUCTION

In multi-robot systems, communication is thought as a necessary skill for robotsto cooperate, and a number of schemes have been proposed for it [1,2]. However, these studies may not be useful to adapt in a dynamic and complex environment as they set rules to communicate. To achieve cooperation effectively in such environments, we have to discuss the adaptable cooperation using communication. Yanco et al. tried to develop a method to acquire an adaptive communication for cooperation of two robots[3]. Billard et al. proposed a learning method of communication through imitation [4]. This is an interesting approach but the system needs a teacher robot. In these methods and most of robotics resesarch, the communication is treated as special function for the robotic systems.

On the other hand, in developmental psychology, communication is considered as interaction between individuals [5]. Moreover, communication is the transmission of intention, and those who received have to comprehend the intention. In conventional studies on cooperation of robots based on communication as signal transmission, action is taken as a motion of its own body and they focused on decision making using sensory

[1]Correspondence to: Tomohisa Fujiki, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, JAPAN . Tel.: +81-47-136-4260; Fax: +81-47-136-4242; E-mail: fujiki@race.u-tokyo.ac.jp.

information. Communication is signal transmission over wireless LAN or other devices, but it is not correct in developmental psychological sense. There should be a sort of protocol between robots to communicate, and the intention should be exchanged.

Consequently transmitting one's intention could be treated as an action and receiving other's intention could be treated as perception in multi-robot system. By introducing this concept to their control architecture, robots can make an attempt to control other robots. This means that a robot can make an action over constraint of its own D.O.F.(body-expansion behavior), and multi-robot system can be more flexible and adaptable.

In this study, we take in communication to robot's model both as perception and action. It means to achieve cooperation between robots, not only robot's own movement but also sending message to other robots is treated as an action. We have previously developed an action selection method [6] which treats communication as above, but there was a problem of how to adjust different type of actions; self generated action and a requested one by communication. It seems that most effective strategy for the whole system is to accept a request only when the situations for both robots seem to improve.

In this paper, we propose an action adjustment function to achieve cooperation between mobile robots. We also have some computer simulations of collision avoidance as an example of cooperative task, and discuss the results.

2. ACTION SELECTION METHOD INCLUDING INTERACTIVE COMMUNICATION

2.1. Reinforcement Learning

Reinforcement Learning(RL,[7] is widely used in robotic systems to emerge robots' actions from the interaction between the environment. However, in multi-robot system, there is a possibility that the same action causes different state transition which misleads the learning. To avoid this problem, Q-Learning for Semi Markov Decision Process (SMDP, [8]) which can handle discrete time series is utilized generally. Q-Learning algorithm for SMDP is as follows.

1. Observe state s_t at time t in the environment.
2. Execute action a_t selected by action selection node.
3. Receive reward r and calculate the sum of discounted reward R_{sum} until its state changes.

$$R_{sum} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots + \gamma^{N-1} r_{t+N-1} \tag{1}$$

   Here, r is a discount factor $(0 \leq r \leq 1)$.
4. Observe state s_{t+N} at time $t + N$ after the state change.
5. Renew Q value by equation (2).

$$Q(s_t, a_t) \leftarrow (1 - \alpha) Q(s_t, a_t) + \alpha [R_{sum} + \gamma^N \max_{a'} Q(s_{t+N}, a')] \tag{2}$$

   Here, α is a learning rate $(0 \leq \alpha \leq 1)$ and a' is possible actions in state s_{t+N}.
6. Clear r.
7. Renew time step t to $t + N$, and return to 1.

2.2. Basic Actions

There are a variety of tasks considered as cooperative tasks, but in this paper, we are going to discuss collision avoidance problem of mobile robots. This is because that although there are a lot of the rule based schemes proposed for it, it can still see the effect of the communication and the expansion in D.O.F. directly.

We suppose omni-directional mobile robots which are equipped with omni-directional visual sensors. Considering communication as robots' action, basic actions for robots are set as Table 1. Here, "Comminication" means intention transmission, which is a requesting action to other robot to make an asked action. This means that a robot can request any actions which the other robot can make. Robots acquire their state-action policy by RL. We also configure robot's state space as Table 2. Numbers in Table 2 shows the number of state space for each domains. Example for the visual sensory information is shown in Figure 1. The size of the other robot on image plane is determined by the distance threshold. Both the direction of the other robot and the goal are devided into six state spaces. Direction of the wall has five state spaces, which are front, left, right, and the back of the robot, or, no walls. In Figure 1, the white wall is placed above the distace threshold, so only the grey wall is considered as an obstacle in robot's state space. In this framework, a robot selects, evaluates and learns its action from sensory information and other robots' intentions.

Table 1. Actions of robot

Move Own Body	Communication
- No changes in speed or direction	- No changes in speed or direction
- Speed down (2[mm/sec])	- Speed down (2[mm/sec])
- Speed up (2[mm/sec])	- Speed up (2[mm/sec])
- Change direction (+45 [deg])	- Change direction (+45 [deg])
- Change direction (-45 [deg])	- Change direction (-45 [deg])

2.3. Action Selection and Reward

There are a lot of action selection models for Q-Learning like Max Selection or Random Selection. One of the methods to improve its adaptability gradually by RL is probabilistic

Table 2. Configuration of state space

Visual sensory information	
- Size of other robot on image plane	2
- Direction of other robot	6
- Direction of the goal	6
- Wall direction inside the sensing area	4+1(none)
Communication	
- Other robot's request	5+1(none)
Other Information	
- Own Speed	2
Number of the State Space	4320

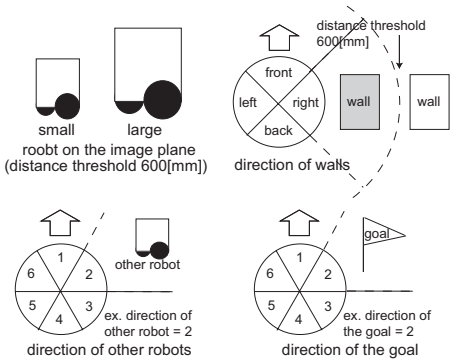


Figure 1. Visual Sensory Information

action selection using Boltzmann distribution (Boltzmann selection). It is used widely and is reported that probabilistic selection works better than deterministic policy in multi-agent systems [9].

In Boltzmann Selection model, probability $p(a \mid s)$ to make action a in state s is defined as equation (3).

$$p(a \mid s) = \frac{\exp^{Q(s,a)/T}}{\sum_{a_i \in A} \exp^{Q(s,a_i)/T}} \tag{3}$$

Here, T is temperature constant. If T is near zero, action selection will be deterministic, and if T becomes large, action selection will be more random and will do aggressive search for state-action policy. Evaluation of the selected action is done by using the distance from the goal $g(t)$ in time t. Reward r_t is defined by equation (4).

$$r_t = \mu(g(t) - g(t - \Delta t)) \tag{4}$$

Here, μ is a weight value and represents effectiveness of the reward. Δt is cycle time for decision making.

3. ACTION ADJUSTMENT FUNCTION

When communication is treated as an action for intention transmission, accepting all the requested actions will only to improve other robots' situations. However, for the whole system, it is seems that most effective way is to accept the request only when the situations of both robots can be improved. To accept such requests, there is a need for action adjustment function to compare the actions which are self determined action and a requested one by communication. It makes the robots to create better situations, and will be able to cooperate efficiently.

For this action adjustment, we introduce the algorithm which is illustrated in Figure 2. First, a robot decides whether to move itself or to make other robot move by communication. This is a selfish action selection which doesn't consider the state of other robot.

Of course there is a probability that the request will be refused, but whether to accept or reject the request is determined by the receiver. Next, a robot will determine which action to make; the selfish action that is decided at first step or a requested action by other robot. By those two steps, a robot can select an action considering a request from other robot.

This adjustment algorithm can be utilized generally by giving numeric values for each actions. In this paper, we use Q-Learning algorithm for SMDP and the Q values from the RL are used as numeric values for two step action selection. The implemented algorithm for the robot is shown in Figure 3.

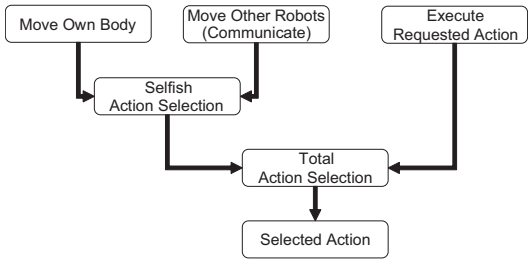


Figure 2. Action selection process

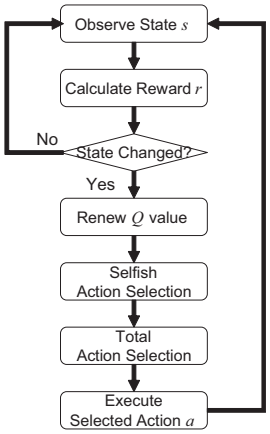


Figure 3. Algorithm for action selection including communication

4. COMPUTER SIMULATION OF COLLISION AVOIDANCE PROBLEM

In this section, we are going to have some computer simulations to test our approach and discuss the results.

4.1. Settings

There are two omni-directional mobile robots in simulation field, and the task is collision avoidance as an example of cooperative task. To compare our approach to general approach of communication in robotic research field, we set three conditions.

Case A is for our proposed approach and robots can use communication to move the other robot by intention transmission. In case B, robots can inform their adjacent action which they made by signal transmission. By this, we can eliminate an influence on the size of state space and robots in case B have same state number as in case A. Robots in case C cannot use communication.

Common settings are as follows. Robot is cylinder shaped and its diameter is 300[mm]. Start position of each robot is set 500[mm] from longitudinal sides of the environment, symmetry in ups and downs. Goal position is the starting point of the other robot, and is set face to face in initial condition. Maximum speed of the robots is 40[mm/sec], and minimum is 10[mm/sec]. It assumes that robots can output their speed without a time lag.

A trial is terminated under four conditions, which are the goal of the both robots, the collision of robots, the collision of either robot against walls or simulation area, or when the time step reached to 3000. The parameters for RL are set experimentally as $\Delta t = 1.0$[sec], $\mu = 0.1$, $\alpha = 0.04$, $\gamma = 0.9$ and $T = 0.2$. Reward for the robots are calculated by equation (4), but in case of any collisions, $r = -5$ is given as punishment value.

4.1.1. Simulation 1

In simulation 1, straightway environment in Figure 4 is utilized. Width $800 \leq x \leq 3000$[mm] is changed by 100[mm] and computer simulation is run for four times in every situation, and the learning is episodic for each simulation. Maximum trial number is 30000 for every experiment.

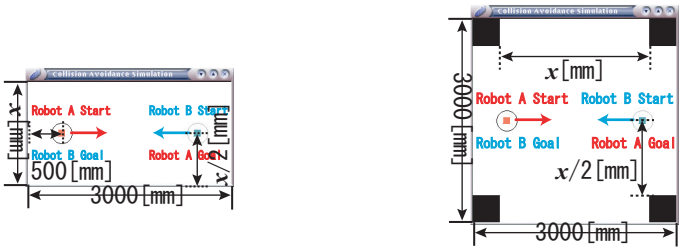
4.1.2. Simulation 2

In simulation 2, crossroad environment in Figure 4 is utilized. Simulation area is 3000[mm] square, and the width of both roads are x[mm], which changes $600 \leq x \leq 3000$[mm] by 200[mm]. Four black pieces in Figure 4 are walls (obstacles). Computer simulation is run for four times in every situation, and Maximum trial number is 100000 for every experiment. Learning is episodic for each simulation.

All settings has the same distance for goal, to make it easy to compare the results. In simulation 2, when a robot moves, physical relationship against the walls change, and it affects robot's state space. Consequently, robots' state change frequently when x is small, and the problem will be much difficult compared to the same x in Simulation 1.

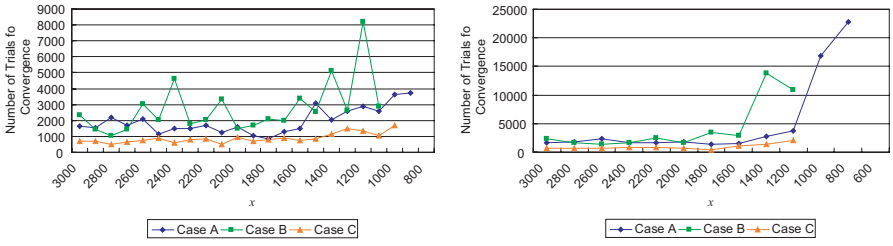
4.2. Results

Figure 5 shows the number of trials for convergence. In this report, "convergence" means 100 continuous goals. Horizontal axis shows the width of the road x. Data on those graphs are the average of four trials. It shows some oscillation, but the aptitude can been comprehended.



Straightway Crossroad

Figure 4. Over view of environment



Simulation 1 Simulation 2

Figure 5. Number of trials for convergence

4.2.1. Convergence Properties

When the width x is large enough for robots and the problem can be solved easily, Case C achieves convergence faster than other cases. We believe that this occurs because the state space of Case C is one fifth of other cases and therefore it is easy to acquire the state-action policy. The result of Case B shows large oscillation in both graphs. In this case, communication changes the state of the other robot, and it makes difficult to search state-action policies. Communication as signal transmission doesn't show its superiority in any case of our experiments. It only multiplies the number of states and prevents system from fast achieving of cooperation. Finally, Case A has superiority to other methods when x is small. This is the condition which the problem is hard to solve and is difficult to cooperate with others. Results show that our approach can solve the problem cooperatively even when the other approaches cannot solve it. It is a difficult situation for robots to cooperate without communication, and comparing Case A to Case B, our proposed system works better than usual usage of the communication such as information transmission.

4.2.2. Quality of the Solution

Figure 6 shows the number of steps to converge, which shows the quality of the solution achieved by the system. Data on those graphs are the average of four trials. Although there are many spikes, Case A apts to generate better solutions than the other methods. We consider that intention transmission worked effectively by affectiong the other robots, only when the communication is needed. This result supports our approach that it is not only in the fastness in finding solutions but also in the quality of the solution.

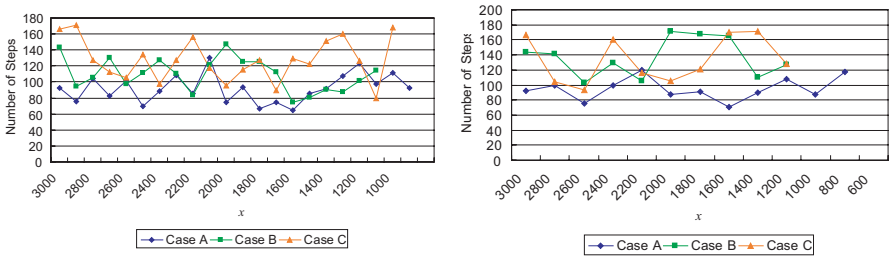


Figure 6. Number of steps

5. CONCLUSIONS

In this paper, we proposed a method to adjust different type of actions which include communication as intention transmission. By using this method, we enabled to treat communication as intention transmission action in multi-robot system and also examined its performance by computer simulations. The results show that our approach can find solution in difficult situation where cooperation is hardly achieved without communication, and is also excels in the quality of solution achieved by the system than ordinal way of communication or without using communication. In our future work, we will try our approach in more complex environments or other tasks.

References

[1] Y. Arai, S. Suzuki, S. Kotosaka, H. Asama, H. Kaetsu and I. Endo: Collision Avoidance among Multiple Autonomous Mobile Robots using LOCISS (LOcally Communicable Infrared Sensory System), *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2091–2096, 1996.

[2] N. Hutin, C. Pegard, E. Brassart, A Communication Strategy for Cooperative Robots, *Proc. of IEEE/RSJ Intl.Conference on Inteligent Robots and Systems*, pp. 114–119, 1998.

[3] H. Yanco, L. A. Stein, An Adaptive Communication Protocol for Cooperating Mobile Robots, *From Animals to Animats 2*, pp. 478–485, 1993.

[4] A. Billard, G. Hayes, Learning to Communicate Through Imitation in Autonomous Robots, *Artificial Neural Networks - ICANN'97*, pp. 763–768,1997.

[5] Walburga Von Raffler-Engel (Editor): Aspects of Nonverbal Communication, Loyola Pr, 1979.

[6] M. Hoshino, H. Asama, K. Kawabata, Y. Kunii and I.Endo: Communication Learning for Cooperation among Autonomous Robots, *Proceedings of the IEEE International Conference on Industrial Electronics, Control & Instrumentation*, pp. 2111–2116, 2000.

[7] Richard S. Sutton and Andrew G. Barto: Reinforcement Learning:An Introduction (Adaptive Computation and Machine Learning), The MIT Press, 1998.

[8] Steven J. Bradtke and Michael O. Duff: Reinforcement Learning Methods for Continuous-Time Markov Decision Problems, In G. Tesauro and D. Touretzky and T. Leen, editors, *Advances in Neural Information Processing Systems*, Vol.**7**, pp. 393–400, The MIT Press, 1995.

[9] Satinder P. Singh and Tommi Jaakkola and Michael I. Jordan: Learning Without State-Estimation in Partially Observable Markovian Decision Processes, *International Conference on Machine Learning*, pp. 284–292, 1994.