Paper:

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

Tomohisa Fujiki*, Kuniaki Kawabata, and Hajime Asama***

*RACE, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan
E-mail: {fujiki, asama}@race.u-tokyo.ac.jp
**Distributed Adaptive Robotics Research Unit, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
E-mail: kuniakik@riken.jp

[Received October 21, 2005; accepted December 21, 2006]

In a multi-robot system, cooperation within robots is essential in order to execute tasks efficiently. The purpose of this study is to investigate how robots cooperate with each other using interactive communication. A fundamental role of communication in a multi-robot system is to control other robots by an intension transmission. We believe that a multi-robot system can be more adaptive by treating communication as an action. In this paper, we implemented the action adjustment function to achieve cooperation between two mobile robots. Also we discuss the results of computer simulations of collision avoidance as an example of cooperative task.

Keywords: Q-learning, multi-robot system, communication, cooperation, mobile robot

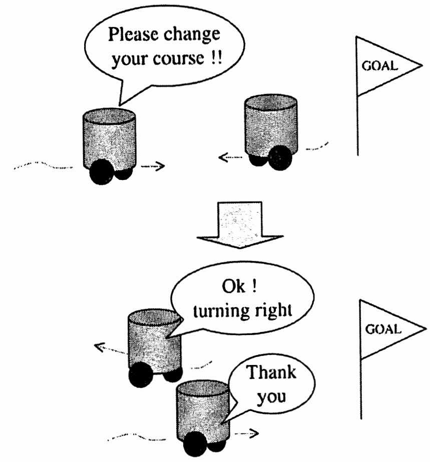


Fig. 1. Communication as an intention transmission.

1. Introduction

In a multi-robot system, robots can improve their capabilities by changing their configurations or by working in parallel according to the circumstances. In such a system, robots execute their tasks for their own purpose, but they have to share and consider the purpose of the whole system simultaneously. Therefore, communication is a necessary skill for robots to realize cooperative tasks [1]. Our previous works were on communication among two robots as a means of signal transmission to the other robot to achieve tasks cooperatively [2, 3]. However, these studies set rules to communicate, so these methods may not be adaptive in a dynamic and complex environment. Yanco et al. tried to develop a method to acquire an adaptive communication for cooperation between two robots [4]. In this system, un-interpreted vocabulary is given and the robots acquired the usage of the words. Billard et al. proposed a learning method of communication through imitation [5]. This is an interesting approach but the system needed a teacher robot. In these methods and in most of robotics researches, communication is treated as a special function for a robotic system.

On the other hand, in developmental psychology, communication is considered as an interaction between individuals, and nonverbal communication as its fundamental part [6]. Moreover, communication is a transmission of intention, and those who received the message have to comprehend the intention. In conventional studies on cooperation of robots, communication is taken as a signal transmission over wireless LAN or other devices, an action is taken as a motion of its own body, and they focused on decision making using sensory information. But it is not correct in developmental psychological sense. There should be a sort of protocol between robots to communicate, and an intention should be exchanged by communication.

Consequently transmitting one's intention could be treated as an action and receiving the other's intention could be treated as a perception in a multi-robot system (**Fig. 1**). By introducing this concept to their control architecture, robots can make an attempt to control the other robot. This means that a robot can make an action over constraint of its own degree of freedom, and a multi-robot system can be more flexible and adaptable. In this study, we take in communication to robot's model both as a perception and an action. It means that not only a robot's own movement but also sending message to the other robot is treated as an action.

We have previously developed an action selection method [7, 8] which treat communication as above, but there was a problem of how to adjust different type of actions: a self generated action and a requested action from the other robot by communication. It seems that the 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 two mobile robots. Also we will discuss the results for the computer simulations on collision avoidance as an example of cooperative task.

2. Intention Transmission

In most researches of multiple robots cooperation, only the physical actions(for example, just moving) were considered as the actions of a robot. However, by applying communication between robots, transmission of a robot's intention to the other robot could be treated as an action as well. Therefore, we have attempted to introduce this transmission of intention to the robot's control architecture.

In terms of control, one of the important role of communication is to make the robot control the other robot's behavior by intention transmission. If a robot's behavior is limited by the number of actuators, D.O.F. of the robot would also be limited by the number of actuators. However, if the robot can transmit its own intention and the receiver executes the order, the robot is capable of controlling the other's behavior. And this demonstrates the expansion of the robot's D.O.F., which we refer to as body-expansion behavior. We consider that a multi-robot system can improve flexibility and adaptability with body-expansion behavior.

In our previous work, we applied this concept to collision avoidance problem of mobile robots [7, 8]. We attempted computer simulations to examine the communication scheme among the robots and discussed the effect of our method. Also, we have focused on communication emergence for body-expansion behavior within two mobile robots. It was a task to acquire a common protocol for cooperation by exchanging meaningless symbols, and the robot system became more adaptive with our proposed method. However, the system had a problem that only one of the robot in the system was able to communicate to the other robot to request an order. Where as the other robot always obeyed to the orders made. When it comes to a bidirectional communication system, robots in the system have to decide which action to take: a self generated action or a requested action from the other robot by communication.

In this paper, we examine the action selection method with our discussed system for further effective adaptation.

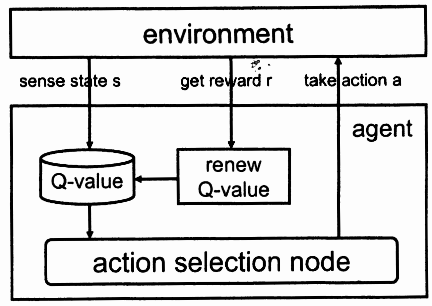


Fig. 2. Q-learning.

3. Action Selection Method Including Interactive Communication

To emerge functions of adaptation to a dynamic environment, we utilized reinforcement learning scheme. By utilizing this scheme, robots are able to acquire adaptive behaviors in a dynamic environment without a prior knowledge.

3.1. Reinforcement Learning

When a robot has a large state space, it is difficult to predict a solution, and to set all state-action rules beforehand. But in an autonomous robot system, action selection should be online and processed in real time. In such cases, robot must acquire behavior generation function to emerge its action from the interaction between the environment and the other robot. To treat this sort of acquisition problem, un-supervised learning approach is utilized, and Reinforcement Learning (RL, [9]) method is one of such approaches. In RL, trial and error approach is applied.

However, in multi-robot systems, there is a possibility that the same action causes a different state transition which can mislead the learning. To avoid this problem, Q-learning for Semi Markov Decision Process (SMDP, [10]) (**Fig. 2**) is utilized generally because it can handle discrete time series. Q-learning algorithm for SMDP are 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.

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

   Here, γ is a discount factor $(0 \leq \gamma \leq 1)$.

4. Observe state s_{t+N} at time $t+N$ after the state change.
5. Renew Q value by Eq. (2).

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

Table 1. Actions of robot.

Move own body
- No changes in speed or direction
- Speed down (2 mm/sec)
- Speed up (2 mm/sec)
- Change direction (+45°)
- Change direction (−45°)
Speech
- No changes in speed or direction
- Speed down (2 mm/sec)
- Speed up (2 mm/sec)
- Change direction (+45°)
- Change direction (−45°)

Table 2. Configuration of state space.

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

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.

This learning algorithm is utilized in each mobile robot.

3.2. Basic Actions

In this paper, we discuss collision avoidance problem of mobile robots as an example of cooperative task. We have assumed to use omni-directional mobile robots which are equipped with omni-directional visual sensors. By considering communication as one of robots' actions, basic actions for robots are set as **Table 1**. Here, "Speech" means communication through message transmission. Robots acquire their state-action policy by RL. We also configured a robot's state space as **Table 2**. An example of the visual sensory information is shown in **Fig. 3**. In this framework, a robot selects, evaluates and learns its action from sensory information and the other robot's intentions.

3.3. 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 robots' adaptability gradually by RL is probabilistic action selection using Boltzmann distribution (Boltzmann selection). Boltzmann selection is used widely and is reported that probabilistic selection is significant than deterministic policy in multi-robot systems [11].

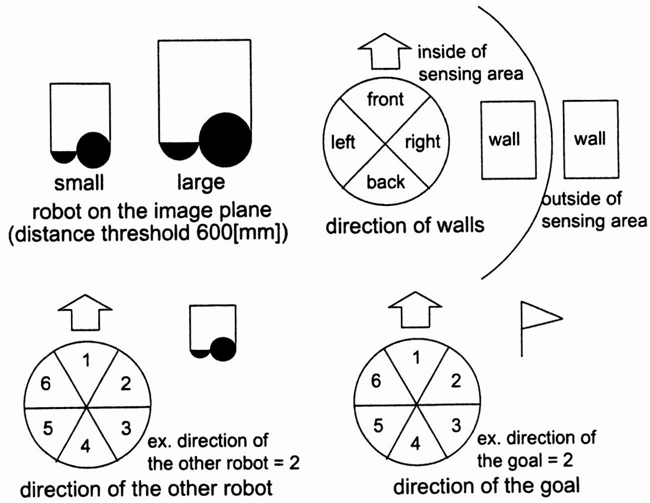


Fig. 3. Visual sensory information.

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

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

Here, T is a temperature constant. If T is near zero, action selection will be deterministic, and if T becomes large, action selection will be more random and carry out 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 Eq. (4).

$$r_t = \mu(g(t) - g(t - \Delta t)) \quad \ldots\ldots\ldots (4)$$

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

4. Action Adjustment Function

When communication is treated as an action for intension transmission, accepting all the requested actions will only improve the other robot's situation.

However, for the whole system, it seems that the most effective way is to accept requests only when the situations of both robots seem to improve. To accept such requests, action adjustment function is needed in order to compare actions which are self-determined action and a requested one by communication. This allows the robots to create favorable situations, and to cooperate efficiently.

For this action adjustment, we introduce an algorithm which is illustrated in **Fig. 4**. First, a robot decides whether to move itself or to make the other robot move by communication. This is a selfish action selection which does not consider the state of the other robot. Of course there is a possibility that the request will be refused. Next, a robot will determine which action to make: the selfish action that is decided at first step or a requested action by

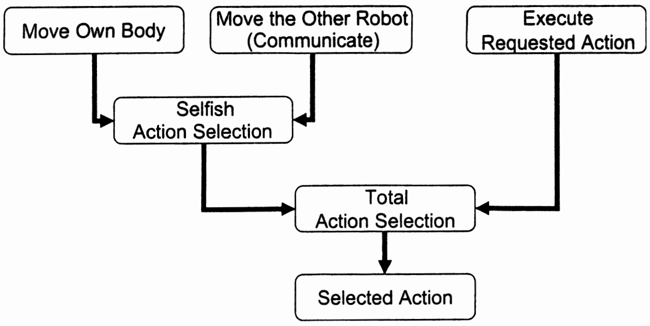


Fig. 4. Action selection process.

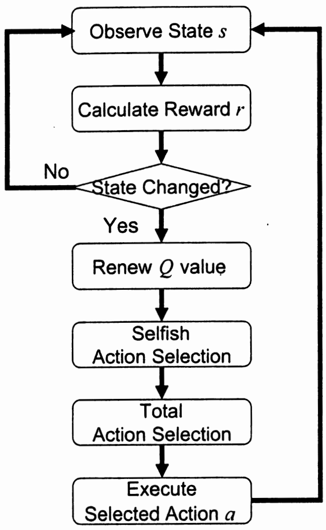


Fig. 5. Algorithm for action selection including communication.

the other robot. By those two steps, a robot can select an action considering a request from the other robot.

The Q value from RL is used for two-step action selection. The implemented algorithm is shown in **Fig. 5**.

5. Computer Simulation of Collision Avoidance Problem

In this section, we discuss the results of the computer simulations to test our approach.

5.1. Settings

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

Case A is for our proposed approach, and robots can move the other robot by intension transmission. In case B, robots can use communication which can send message of their conditions to the other robot. To eliminate an influence of the size of the state space, robots have the same state number as in case A, and robots can inform

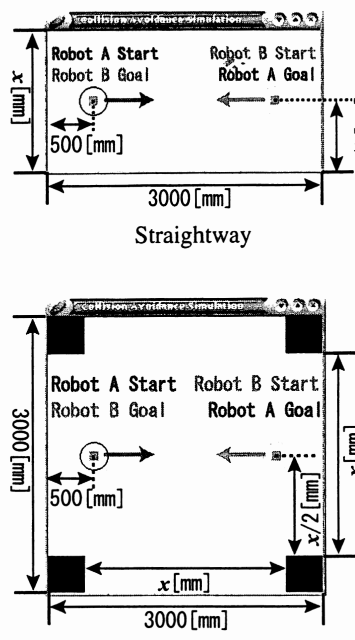


Fig. 6. Over view of environment.

their adjacent action. Robots in case C cannot use communication.

Common settings are as follows. A 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 robots are set face to face in the 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 goals of both robots, the collision of robots, the collision of either robot against walls or simulation area, or when the time step reach 3000. The parameters for RL are set 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 Eq. (4), but in case of any collisions, $r = -5$ is given as a punishment value.

5.1.1. Simulation 1

In simulation 1, straightway environment in **Fig. 6** is utilized. Width $800 \leq x \leq 3000$ mm is changed by 100 mm and computer simulation is run four times in each situation. Maximum trial number is 30,000 for every experiment.

5.1.2. Simulation 2

In simulation 2, crossroad environment in **Fig. 6** 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 **Fig. 6** are walls (obstacles). Computer simulation is run four times in each situation, and maximum trial number is 100,000 for every experiment.

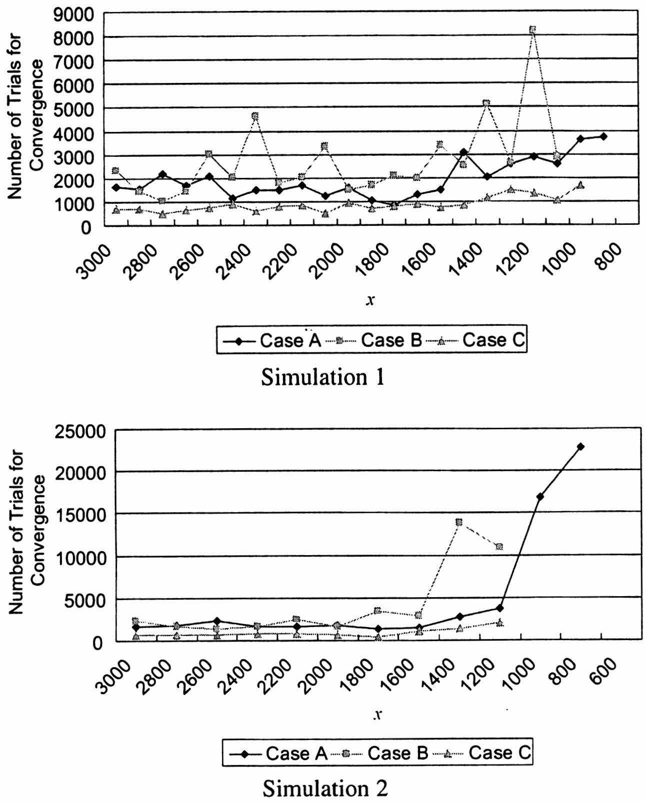


Fig. 7. Number of trials for convergence. Case B did not converge when $x = 800$, 900, and Case C did not converge when $x = 800$.

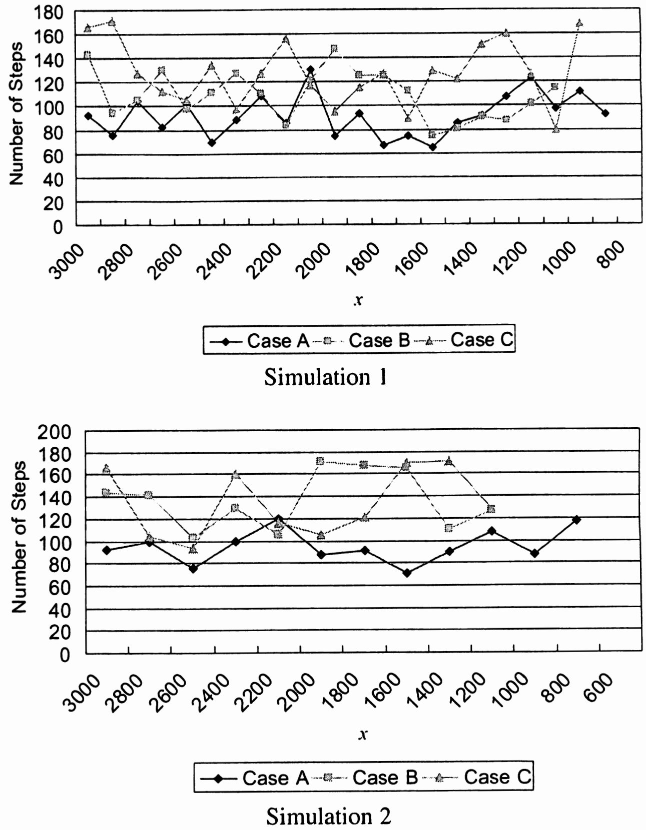


Fig. 8. Number of steps. Case B did not converge when $x = 800$, 900, and Case C did not converge when $x = 800$.

5.2. Results

Figure 7 shows the number of trials for convergence. In this report, "convergence" means 100 continuous goals. Data on those graphs are the average of four trials, and only the points which converged more than 3 out of 4 trials are depicted in the graphs. Horizontal axis shows the width of the road x.

5.2.1. Convergence Properties

When the width x was large enough for robots and the problem could be solved easily, Case C achieved convergence faster than other cases. We believe that this occurs because the state space of Case C was one fifth of other cases and therefore it was easy for robots to acquire the state-action policy. The result of Case B shows large oscillation in both graphs. In this case, communication changed the state of the other robot, and it made robots difficult to search state-action policies. Communication as signal transmission does not show its superiority in any case of our experiments. We believe that in this case, communication only multiplied the number of states and prevented system from achieving of cooperation immediately. Finally, Case C had superiority to other methods when x was small, the condition which the problem was hard to solve.

Results show that our approach solved the problem cooperatively even when the other approaches could not solve it. Those cases were the difficult situations for robots to cooperate without communication, and comparing Case C to Case B, our proposed system worked better than usual usage of the communication such as information transmission.

5.2.2. Quality of the Solution

Figure 8 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, and only the points which converged more than 3 out of 4 trials are depicted in the graphs. From these graphs, Case A tends to generate better solutions than other methods, and our approach is supported not only in the fastness in finding solutions but also in the quality of the solution.

6. Conclusion

In this study, we have discussed the adaptive cooperation behavior between mobile robots using communication as intention transmission. We have proposed to treat transmitting one's intention as an action and also receiving the other's intention as a perception. By introducing communication to control architecture, robots can control the other robot which implies to make an action over constraint of its body (body-expansion behavior). We have already discussed cooperative behavior acquisition 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.

In this paper, we proposed a method to adjust different type of actions which include communication as intention transmission. By using this method, we allowed to treat communication as intention transmission action in a multi-robot system and examined its performance by computer simulations. The results show that our approach can find solution in difficult situations where cooperation is hardly achieved without communication. Also, our approach showed significance in the quality of the solution achieved by the system, rather than the ordinal way of communication or without using communication.

In our future work, we will try our approach in more complex environments or other tasks.

Acknowledgements

This work has been partially supported by a Grant-in-Aid for Scientific Research on Priority Areas "Emergence of Adaptive Motor Function through Interaction between Body, Brain and Environment" from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References:

[1] N. Hutin, C. Pegard, and E. Brassart, "A Communication Strategy for Cooperative Robots," Proc. of IEEE/RSJ Intl. Conference on Intelligent Robots and Systems, pp. 114-119, 1998.

[2] Y. Ishida, H. Asama, K. Ozaki, A. Matsumoto, and I. Endo, "Design of Communication System and Development of a Simulator for an Autonomous and Decentralized Robot System," Journal of Robotics Society of Japan, 10(4), pp. 544-551, 1992 (in Japanese).

[3] 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.

[4] H. Yanco and L. A. Stein, "An Adaptive Communication Protocol for Cooperating Mobile Robots," From Animals to Animats 2, pp. 478-485, 1993.

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

[6] W. Von Raffler-Engel (Ed.), "Aspects of Nonverbal Communication," Loyola Pr, 1979.

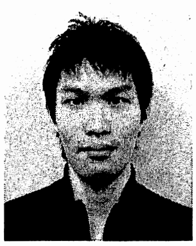
[7] M. Hoshino, H. Asama, K. Kawabata, Y. Kunii, and I. Endo, "Communication Learning for Cooperation among Autonomous Robots," Proceedings of the IEEE International Conference on Industrial Electronics, Control & Instrumentation, pp. 2111-2116, 2000.

[8] K. Kawabata, H. Asama, and M. Tanaka, "A Study of Communication Emergence among Mobile Robots: Simulations of Intention Transmission," Distributed Autonomous Robotic Systems 5, Springer-Verlag, pp. 71-80, 2002.

[9] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)," The MIT Press, 1998.

[10] S. J. Bradtke and M. O. Duff, "Reinforcement Learning Methods for Continuous-Time Markov Decision Problems," In G. Tesauro, D. Touretzky, and T. Leen (Eds.), Advances in Neural Information Processing Systems, Vol.7, pp. 393-400, 1995.

[11] S. P. Singh, T. Jaakkola, and M. I. Jordan, "Learning Without State-Estimation in Partially Observable Markovian Decision Processes," International Conference on Machine Learning, pp. 284-292, 1994.



Name:
Tomohisa Fujiki

Affiliation:
Graduate Student, Research into Artifacts, Center for Engineering (RACE), The University of Tokyo

Address:
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan

Brief Biographical History:
2005 Entered School of Engineering, The University of Tokyo

Main Works:
• distributed autonomous robotic systems, mobiligence

Membership in Academic Societies:
• The Japan Society of Mechanical Engineers (JSME)

Name:
Kuniaki Kawabata

Affiliation:
Unit Leader, Distributed Adaptive Robotics Research Unit, RIKEN (The Institute of Physical and Chemical Research)

Address:
2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Brief Biographical History:
1997- Special postdoctral researcher, Biochemical Systems Lab. at RIKEN
2000- Research scientist, Advanced Engineering Center at RIKEN
2002- Research scientist, Distributed Adaptive Robotics Research Unit, RIKEN
2005- Unit Leader of Distributed Adaptive Robotics Research Unit, RIKEN

Main Works:
• Distributed Autonomous Robotic Systems, Intelligent Ubiquitous Information Device Machine Learning and Visual Perception

Membership in Academic Societies:
• The Japan Society of Mechanical Engineering (JSME)
• The Robotics Society of Japan (RSJ)
• The Japanese Society of Instrumentation and Control Engineers (SICE)
• The Institute of Electrical Engineers of Japan (IEEJ)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Institute of Electrical and Electronics Engineers, Inc. (IEEE)

Name:
Hajime Asama

Affiliation:
Professor, Research into Artifacts, Center for Engineering (RACE), The University of Tokyo

Address:
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan

Brief Biographical History:

1986- Research Associate of RIKEN (The Institute of Physical and Chemical Research)

1998- Senior Scientist of RIKEN (The Institute of Physical and Chemical Research)

2002- Professor of Research into Artifacts, Center for Engineering (RACE), The University of Tokyo

Main Works:

• "Distributed Task Processing by a multiple Autonomous Robot System Using an Intelligent Data Carrier System," Intelligent Automation and Soft Computing, An International Journal, Vol.6, No.3, pp. 215-224, 2000.

Membership in Academic Societies:

• Institute of Electrical and Electronics Engineers, Inc. (IEEE)
• The Japan Society of Mechanical Engineers (JSME)
• The Robotics Society of Japan (RSJ)
• The Japanese Society of Instrumentation and Control Engineers (SICE)

an Article from

Journal of Advanced Computational Intelligence and Intelligent Informatics

4F Toranomon Sangyo Bldg., 2-29, Toranomon 1-chome, Minatoku, Tokyo 105-0001, Japan
Tel. +813-3508-0051, Fax: +813-3592-0648, E-mail: jaciii@fujipress.jp
homepage URL: http://www.fujipress.jp/