Cooperative Transportation by Humanoid Robots
— Learning to Correct Positioning —

Yutaka Inoue, Takahiro Tohge, Hitoshi Iba

Department of Frontier Informatics,
Graduate School of Frontier Sciences,
The University of Tokyo, 7-3-1 Hongo,
Bunkyo-ku, Tokyo, 113-8656, Japan
{inoue, tohge, iba}@miv.t.u-tokyo.ac.jp

Abstract. In this paper, we describe a cooperative transportation problem with two humanoid robots and introduce a machine learning approach to solving the problem. The difficulty of the task lies on the fact that each position shifts with the other's while they are moving. Therefore, it is necessary to correct the position in a real-time manner. However, it is difficult to generate such an action in consideration of the physical formula. We empirically show how successful the humanoid robot HOAP-1's cooperate with each other for the sake of the transportation as a result of Q-learning.

1 Introduction

In this paper, we first clarify the practical difficulties we face from the cooperative transportation task with two bodies of humanoid robots. Afterwards, we propose a solution to these difficulties and empirically show the effectiveness both by simulation and by real robots.

In recent years, many researches have been conducted upon various aspects of humanoid robots [1][2]. Since humanoid robots have physical features similar to us, it is very important to let them behave intelligently like humans. In addition, from the viewpoint of AI or DAI (Distributed AI), it is rewarding to study how cooperatively humanoid robots perform a task just as we humans can. However, there have been very few studies on the cooperative behaviors of multiple humanoid robots. Thus, in this paper, we describe the emergence of the cooperation between humanoid robots so as to achieve the same goal. The target task we have chosen is a cooperative transportation, in which two bodies of humanoids have to cooperate with each other to carry and transport an object to a certain goal position.

As for the transportation task, several researches have been reported on the cooperation between a human and a wheel robot [3][4] and the cooperation among multiple wheel robots [5][6]. However, in most of these studies, the goal was to let a robot perform a task instead of a human.

Research to realize collaboration with a legged robot includes lifting operations of an object with two robots [7] and box-pushing with two robots [8]. However, few studies have addressed cooperative work using similar legged robots. It is presumed that body swinging during walking renders cooperative work by a legged robot difficult. Therefore, it is more

<div style="text-align:center">
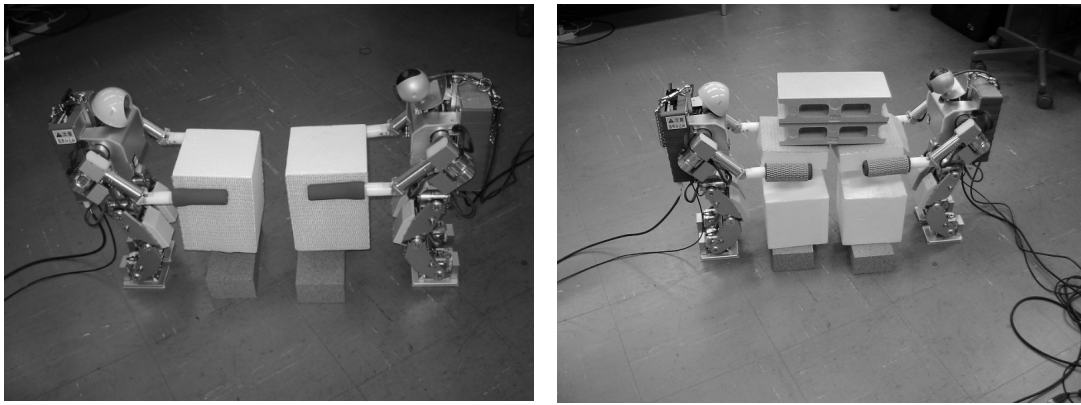
(a) Direct transfer   (b) Trunk-based transfer

Figure 1: Two kinds of tasks.

</div>

difficult for a humanoid robot to carry out a transportation task, because it is capable of more complicated motions and less stable than a usual legged robot.

In leader-follower type control [9][10], which is often used for cooperative movement, it is essential that a follower robot acquire information such as the position and velocity of an object fluctuated by the motion of a leader robot. This information is usually obtained by a force sensor or wireless communication. Such a method is considered to be effective for a robot with a stable center of gravity operating with less information for control. However, much information must be processed simultaneously to allow a humanoid robot to perform complicated actions, such as transporting an object cooperatively, with its difficulty to control caused by its unstable body balance. It would be expensive to build a system that carries out optimal operation using this information.

One hurdle in the case where multiple humanoid robots move carrying an object cooperatively is the disorder of cooperative motion by body swinging during walking. Therefore in this paper, learning is carried out to acquire behavior to correct a mutual position shift generated by this disorder of motion. Q-learning is used for this learning. We will show that behavior to correct a position shift can be acquired based on the simulation results of this study. Moreover, according to this result, the applicability to a real robot is investigated.

This paper is organized as follows. The next section explains the clarified problem difficulties with the cooperative transportation. After that, Section 3 proposes our method to solve the problem. Section 4 presents an experimental result in the simulation environment. Then section 5 shows an experimental result with real robots. Section 6 discusses these results and future researches. Finally, a conclusion is given in Section 7.

2   Operations and Problem in Practice

We conducted an experiment assuming tasks to transport a lightweight object all around, aiming to extract specific problems from using two humanoid robots: HOAP-1 (manufactured by Fujitsu Automation Limited). Dimensions of a HOAP-1 are 223 x 139 x 483 mm (width, depth, and height) with a weight of 5.9 kg. It has two arm joints with 4 degrees of freedom each, and two leg joints with 6 degrees of freedom each: 20 degrees of freedom in all for right and left.
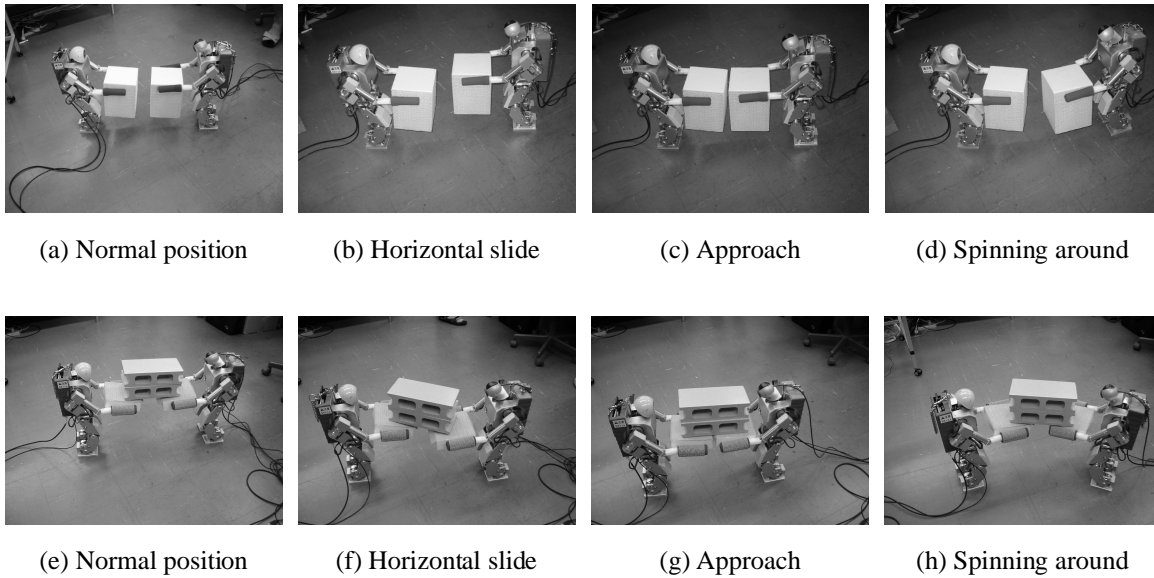
(a) Normal position	(b) Horizontal slide	(c) Approach	(d) Spinning around

(e) Normal position	(f) Horizontal slide	(g) Approach	(h) Spinning around

Figure 2: Normal positions and unexpected movements.

Transportation tasks are of the following two types: (i) each of the two robots transports an object of 150 x 150 x 200 mm and a weight of 80 g in the same direction (Figure 1(a)), and (ii) each of two robots transports a base of 150 x 150 x 50 mm and a weight of 20 g, with two objects of 120 x 240 x 6 mm and a weight of 8 g (Figure 1(b)) thereon.

Actually, it seems to be more practical for two robots to have a single object. However, unless both robots move synchronously in the desirable direction, too much load will be given to the arms of robots, which may often cause the mechanical trouble in the arm and the shoulder. Therefore, for the sake of safety, we divide a transporting object or a stretcher into two parts, each of which a humanoid robot carries and transports synchronously.

It is assumed in experiment (i) that the arm movement can cancel the position shift, and that the distance and angle that can be cancelled would be in the space between two objects. On the other hand, experiment (ii) assumes that two robots carry a stretcher with an object on it. A sponge grip is attached on each robot arm, so that an object would not slip off the arm during the experiment.

The two robots operate in Master-Slave mode. That is, the Master robot transmits data corresponding to each operation created in advance to the Slave robot; the two robots start a motion in the same direction simultaneously. The created basic motions consist of the following 12 patterns: forward, backward, rightward, leftward, half forward, half backward, half rightward, half leftward, right turn, left turn, pick up, and put down. These basic motions are combined to allow the two robots to transport an object.

Each experiment was conducted about 10 times. The results indicated that unintentional motions such as lateral movement (Figures 2(b), 2(f)) and back-and-forth movement (Figures 2(c), 2(g)) by sliding from the normal position (Figures 2(a), 2(e)), and rotation (Figures 2(d), 2(h)) occur frequently in basic transportation motions such as forward, backward, rightward, and leftward. This is considered mainly to result from swinging during walking and the weight of the object.

Such a position shift can be cancelled, if not much, by installing a force sensor on a wrist and moving arms in the load direction. However, a robot's position must be corrected in case

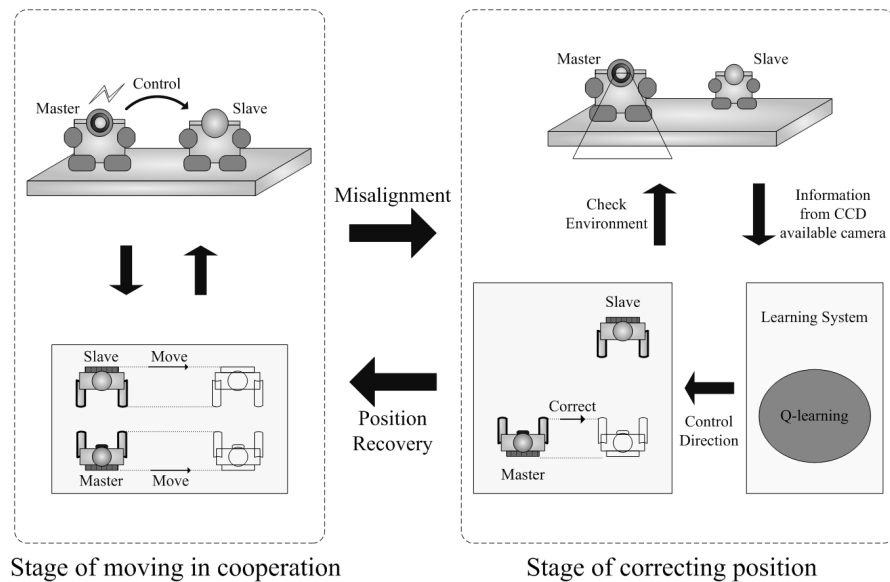Stage of moving in cooperation   Stage of correcting position

Figure 3: The operation overview.

of a shift beyond the limitation of an arm. Improper correction may cause failure of an arm or a shoulder joint and breakage of an object.

3   Learning Model of a Robot

The practical problem of transporting an object is the possibility that a robot falls during movement, due to loss of body balance in connection with a load on the arm by a mutual position shift after moving.

Therefore, it is important to acquire behavior for correcting the position shift generated from movement by learning. A situation is assumed in which two robots move face to face while maintaining the distance within a range to transport an object stably. This motion can be divided into two stages: one in which the two robots move simultaneously, and one in which one robot corrects its position. Simultaneous movement of two robots is controlled by wireless communication. A shift over a certain limit of distance or angle in this motion will be corrected by one robot according to behavior acquired by learning. Figure 3 shows the motion overview for conducting a transportation task. In the first stage, the Master robot performs a motion programmed in advance; simultaneously, it issues directions to perform the same motion to the Slave robot. If there is no position shift after movement, the process forwards to the next stage; otherwise, the position is corrected with the learning system. We have tried to realize a cooperative transportation task by repeating the series of this flow.

Learning for position correction is carried out with Q-learning. Q-learning guarantees that the state transition in the environment of a Markov decision process converges into the optimal direction. However, it requires much time until the optimal behavior obtains a reward in the early stage of learning. Thus, it takes time for the convergence of learning. Furthermore, because all combinations of a state and behavior are evaluated for a predetermined Q value, it is difficult to follow environmental change.

A CCD camera is attached to one robot to obtain information required for learning from the external environment. The external situation is evaluated with images from this CCD
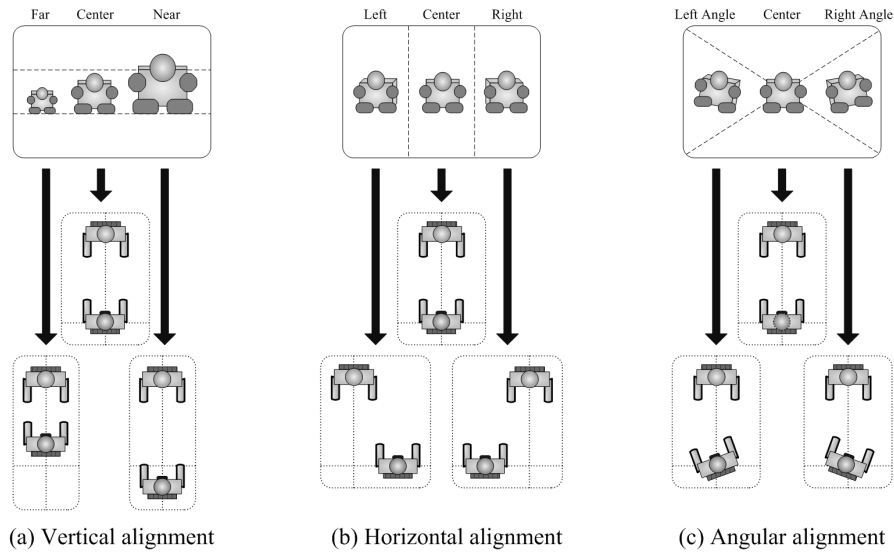
(a) Vertical alignment   (b) Horizontal alignment   (c) Angular alignment

Figure 4: Status segmentation.

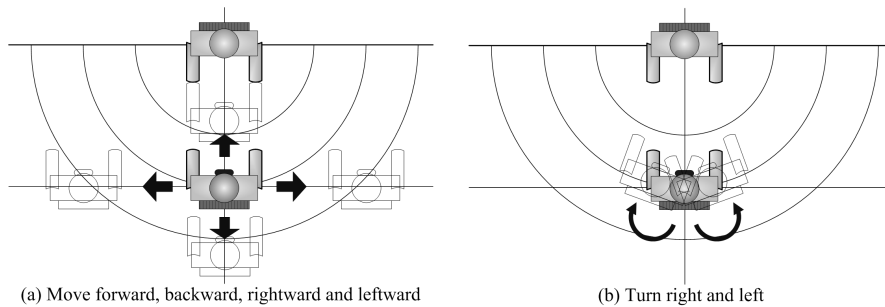(a) Move forward, backward, rightward and leftward   (b) Turn right and left

Figure 5: Image of selected actions.

camera. Based on the partner robot's position projected on the image acquired by the CCD camera, a state space is arranged as shown in Figure 4. It is divided into three states: vertical, horizontal, and angular. Hence, the total number of states of the environment is 27. If the vertical, horizontal, and angular positions are all centered, the goal will be attained.

We assumed 6 behaviors which a robot can choose among the 12 patterns mentioned in Section 2. They are the especially important motions of forward, backward, rightward, leftward, right turn, and left turn. Figure 5 depicts all these motions.

A given motion is not necessarily carried out ideally in practice. Moreover, a move error may arise depending on characteristics of each robot. Therefore, it is necessary to conduct learning including such an indefinite component.

4 Learning in Simulation Environment

The learning model stated in the preceding section has been realized in a simulation environment. This simulator sets a target position at a place of constant distance from the front of the partner robot, which is present in a plane. A task will be completed if the learning robot reaches the position and faces the partner robot. The target position here ranges in distance movable in one motion. In this experiment, back-and-forth and lateral distances and the
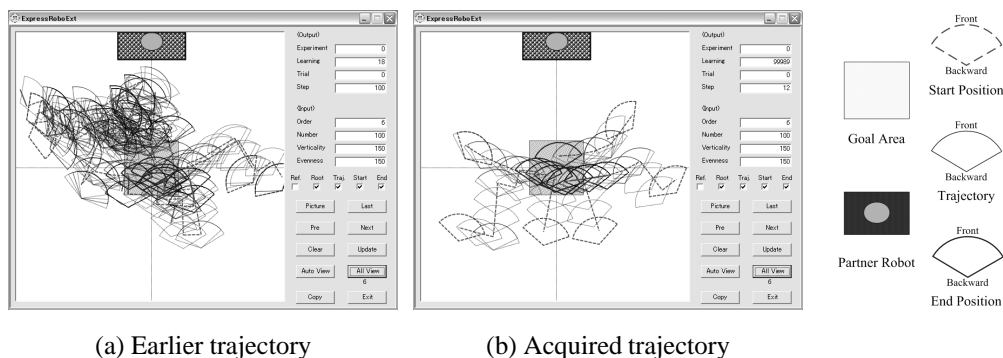
(a) Earlier trajectory   (b) Acquired trajectory

Figure 6: Learning results with Q-learning.

Table 1: Numbers of average movement.

Failuer	Recovery	Q-learning Iterations		
		1,000times	10,000times	100,000times
Horizontal	RL	8.8	8.0	7.4
side	LR	14.6	11.0	10.4
Approach	NF	15.0	7.4	5.6
and away	FN	4.8	4.6	2.4
Spinning	RLS	21.2	23.8	20.2
around	LRS	28.6	29.6	24.8

rotational angle movable in one motion are assumed to be constant. That is, if the movable distance in one step is about 10 cm back-and-forth and 5 cm laterally, the range of the target point will be 50 cm^2. In this range, the goal will be attained if the learning robot is in place where it can face the partner robot with one rotation motion.

The Q-learning parameters for simulation were as follows: the initial Q value, Q_0, was 0.0, the learning rate α was 0.01, the reduction ratio γ was 0.8 and the reward was 1.0 for the task achievement.

Behavior patterns obtained by simulation with the Q-learning approach in the early stage and acquired by learning are shown in Figures 6(a) and 6(b), respectively. In the early stage, motions are observed such as walking to the same place repeatedly and going to a direction different from the target position. Behavior approaching the target position is gradually observed as learning progresses; finally, behavior is acquired to move to the target position and turn to the front with relatively few motions.

5 Experiments with real robots

Following the simulation results described in the previous section, we conducted an experiment with real robots to confirm their applicability. For the recovery from the horizontal left (right) slide, a humanoid robot was initially shifted leftward (rightward) against the opponent robot by 7.5cm. On the other hand, it was initially moved forward (backward) from the correct position by 10cm for the recovery from front (back) position. In case of the rotation failure, the robot was shifted either leftward or rightward by 10cm and rotated toward the opponent by 30 degrees. The actions used for the recovery were of six kinds, i.e., half
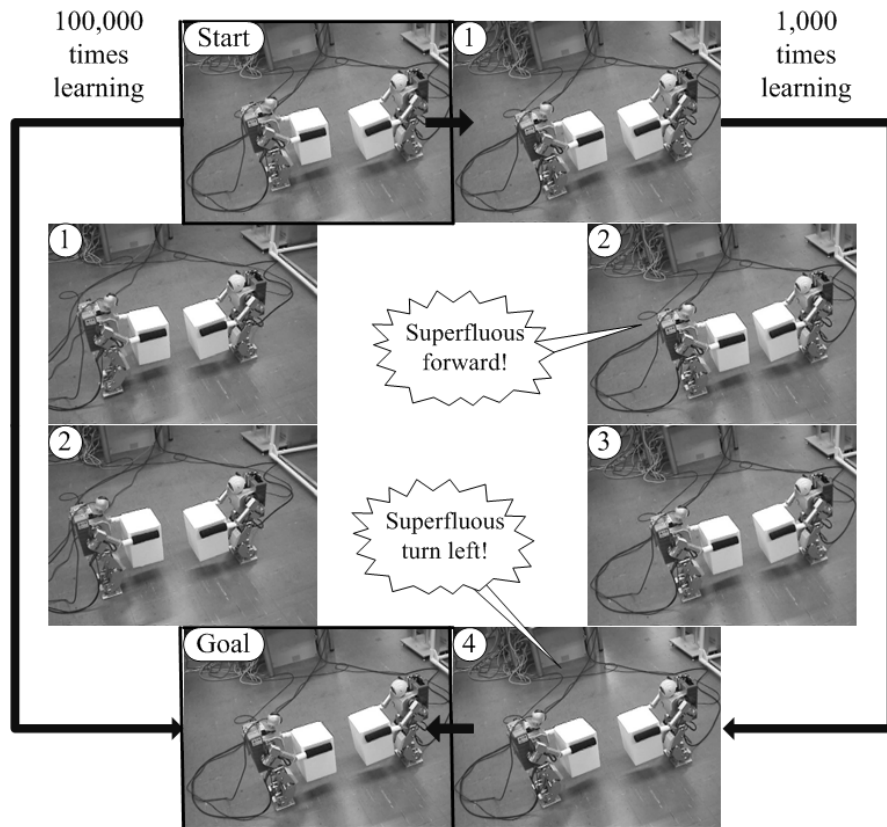
Figure 7: Behavior of LR with short-time learning and full learning.

forwardChalf backwardChalf rightwardChalf leftwardCright turn and left turn. In order to simplify the whole operation, the image data for a CCD camera was preprocessed manually to produce a sensor input.

For this experiment, robots started from one of the three patterns shown in Figures 2 (b), (c) and (d), which were classified as the failure of actions (see section 2). We employed two HOAP-1's, one of which used the learning results, i.e., the acquired Q-table, so as to generate actions for the sake of recovery from the failure. Q-learning was conducted by simulation with different numbers of iterations, i.e., 1,000, 10,000, and 100,000 iterations. The learning parameters were the same as in the previous section.

Table 1 shows the averaged numbers of actions for the sake of recovery from the above three failure patterns. In Table 1: RL represents the slide recovery from the right, LR is the slide recovery from the left, NF stands for the distance recovery from the front, FN is defined as the distance recovery from the back, RLS and LRS are respectively the angle recovery from the right and from the left. The averaged numbers of required actions were measured over 5 runs for each experimental condition, i.e., with different Q-learning iterations.

As can be seen in Figure 7, in case of the recovery from "horizontal slide", the robot often moved forward or made rotations in vain when using the Q-table acquired with 1,000 iterations. These actions were observed more often when sliding from left to right, which prevented effective recovery from the failure. With the increase of learning iterations, the required number of actions decreased so as to show the effectiveness of the learning.

With 1,000 iterations, more actions were needed to recover from the front position to the back. This is because the robot had acquired the wrong habit of moving leftward when the
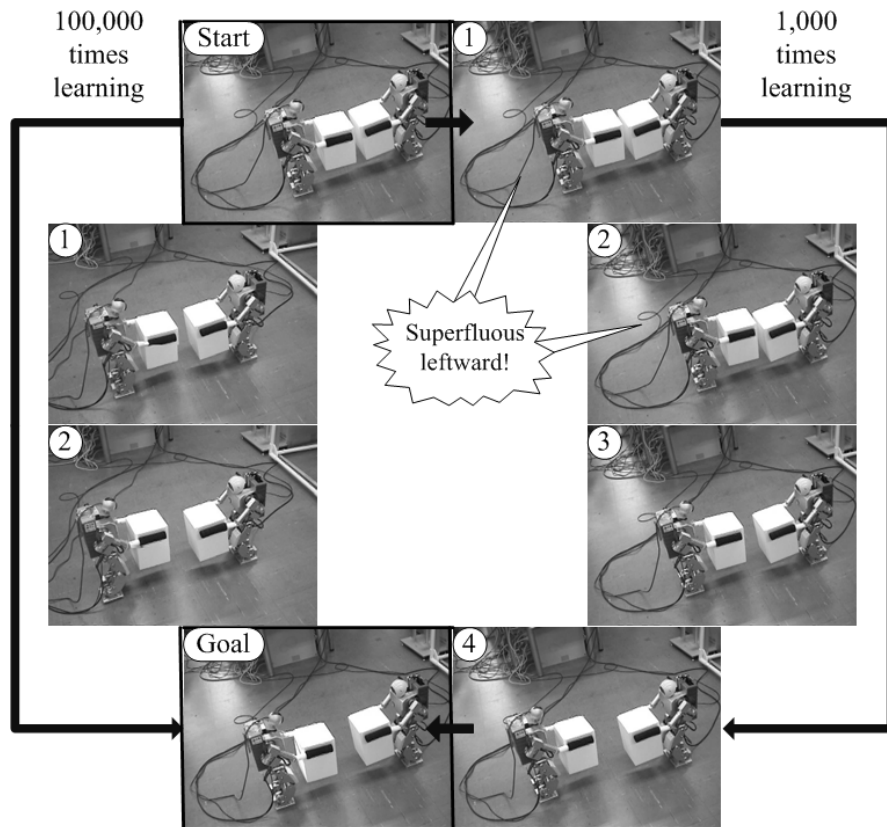
Figure 8: Behavior of NF with short-time learning and full learning.

opponent robot was approaching (see Figure 8). This habit has been corrected with 10,000 iterations, so that much fewer actions were required for the purpose of repositioning.

The recovery from "spinning around" seems to be the most difficult among the three patterns. For this task, the movement from the slant to the front (see Figure 9) was observed with 10,000 iterations, which resulted in the increase of required actions. This action sequence was not observed with 1,000 iterations. Figures 7, 8 and 9 shows a typical robot trajectory by means of the Q-table acquired with 100,000 iterations. As can be seen from the figure, the actions for the recovery were gradually optimized for each failure pattern. Note that with 1,000 iterations, the movement toward the wrong direction, other than the destination, had been observed. However, after the long term learning, this behavior disappeared and the robot learned to correct the failure positioning effectively.

6   Discussion

We have established a learning system for the cooperative transportation by simulation and confirmed its real-world applicability by means of real robots. The effective actions were acquired for the sake of recovery from the position failure as a result of simulation learning.

In a real environment, at the earlier stage of learning, we have often observed the unexpected movement to a wrong direction by real humanoid robots; which was also the case with the simulation. In the middle of learning, the forward movement was more often observed from the slant direction. These types of movements, in fact, had resulted in the better learning performance by simulation, whereas in a real environment they prevented the robot
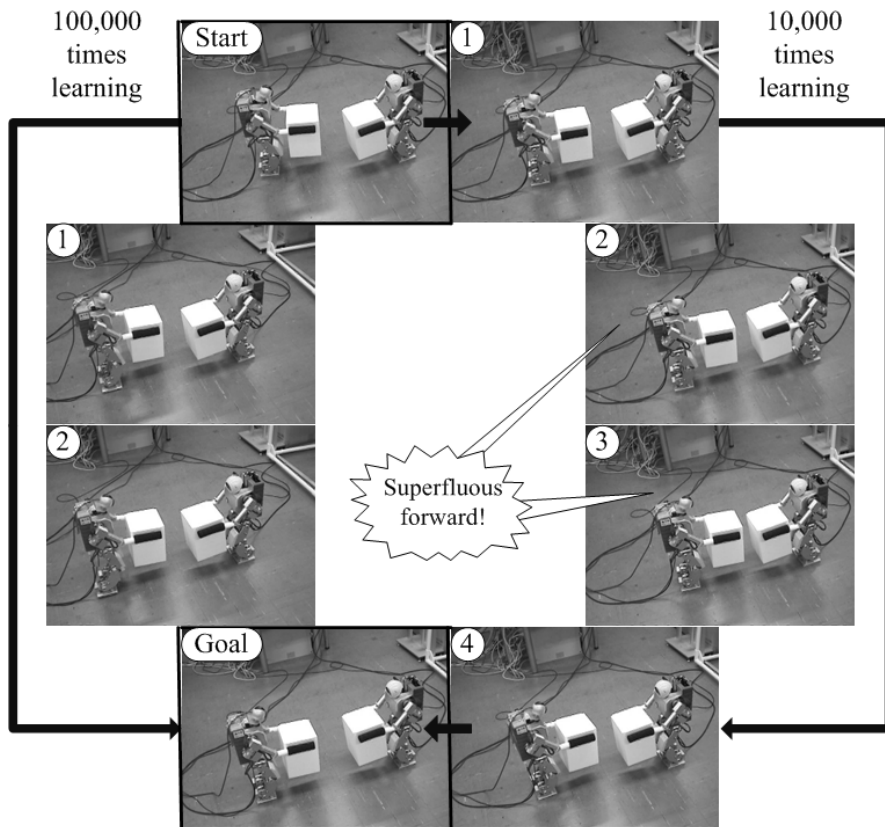
Figure 9: Behavior of LRS with short-time learning and full learning.

from moving effectively. This is considered to be the distinction between simulation and a real-world environment.

In order to solve the difficulty with the distinction, learning in the real world is essential. For this purpose, we are currently working on the integration of GP and Q-learning in a real robot environment [11]. This method does not need a precise simulator, because it is learned with a real robot. In other words, the precision requirement is met only if the task is expressed properly. As a result of this idea, we can greatly reduce the cost to make the simulator highly precise and acquire the optimal program by which a real robot can perform well. We especially showed the effectiveness of this approach with various types of real robots, e.g. SONY AIBO or HOAP-1.

In the above experiments, only one humanoid robot tried to make actions for the sake of recovery. However, more practical and effective will be both robots moving simultaneously to recover from the failure. For this purpose, two robots will have to learn to cooperate with each other to generate recovery actions. This task will require a more efficient learning scheme. We are in pursuit of this topic by using Classifier System [12].

7   Conclusion

Specific problems were extracted in an experiment using a practical system in an attempt to transport an object cooperatively with two humanoid robots. The result proved that both body swinging during movement and the shift in the center of gravity, by transporting an object, caused a shift in the position after movement. Therefore, the learning method for correcting a

position shift in a transportation task was modeled. The learning of behavior was performed in a simulation environment using Q-learning. Moreover, the applicability to a real robot has been verified by using the obtained results.

Learning in practice will be carried out in the future based on behavior learned in the simulation. Thereby, shortening of the learning time in practice is anticipated. Reduction of learning trials in a practical environment is an important subject because one learning elessonf requires much time in practice.

In addition, although only one robot corrects the shift at movement, it is more efficient that the partner robot also corrects its position to reach the target point. Therefore, study is in progress on a method whereby one robot, the one that learns about position correction by carrying a CCD camera, issues directions of motion to the partner robot according to the situation.

By resolving these subjects, the authors wish to realize efficient cooperative transportation to arbitrary places.

References

[1] K. Yokoi, et al., "Humanoid Robot's Application in HRP", In *Proc. of IARP International Workshop on Humanoid and Human Friendly Robotics*, pp.134-141, 2002.

[2] H. Inoue, et al., "Humanoid Robotics Project of MITI", *The 1st IEEE-RAS International Conference on Humanoid Robots*, Boston, 2000.

[3] O. M. AI-Jarrah and Y. F. Zheng, "Armmanipulator Coordination for Load Sharing using Variable Compliance Control", In *Proc. of the 1997 IEEE International Conference on Robotics and Automation*, pp.895-900, 1997.

[4] M. M. Rahman, R. Ikeura and K. Mizutani, "Investigating the Impedance Characteristics of Human Arm for Development of Robots to Cooperate with Human Operators", In *CD-ROM of the 1999 IEEE International Conference on Systems, Man and Cybernetics*, pp.676-681, 1999.

[5] N. Miyata, J. Ota, Y. Aiyama, J.Sasaki and T. Arai, "Cooperative Transport System with Regrasping Carlike Mobile Robots", In *Proc. of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1754-1761, 1997.

[6] H. Osumi, H.Nojiri, Y.Kuribayashi and T.Okazaki, "Cooperative Control of Three Mobile Robots for Transporting A Large Object", In *Proc. of International Conference on Machine Automation (ICMA2000)*, pp.421-426, 2000.

[7] M. J. Matarić, M. Nillson and K. T. Simsarian, "Cooperative Multi-robot Box-pushing", In *Proc. of the 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.556-561, 1995.

[8] H.Kimura and G.Kajiura, "Motion Recognition Based Cooperation between Human Operating Robot and Autonomous Assistant Robot", In *Proc. of the 1997 IEEE International Conference on Robotics and Automation*, pp.297-302, 1997.

[9] J. Ota, Y. Buei, T. Arai, H. Osumi and K. Suyama, "Transferring Control by Cooperation of Two Mobile Robots", *The Journal of the Robotics Society of Japan (JRSJ)*, vol.14, no.2, pp.263-270, 1996.

[10] K. Kosuge, T. Osumi and H. Seki, "Decentralized Control of Multiple Manipulators Handling an Object", In *Proc. of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.318-323, 1996.

[11] S. Kamio, H. Mitsuhashi and H. Iba, "Integration of Genetic Programming and Reinforcement Learning for Real Robots", In *Proc. of the Genetic Computation Conference (GECCO2003)*, 2003.

[12] L. B. Booker, D. E. Goldberg, and J. H. Holland, "Classifier Systems and Genetic Algorithms", In Machine Learning: Paradigms and Methods, MIT Pres, 1990.