Proceedings of the 2nd International Symposium
on Adaptive Motion of Animals and Machines,
Kyoto, March.4-8, 2003

Necessity of Body Image in Applying Reinforcement Learning to Redundant Robots

K. Ito[1], A. Gofuku[1], M. Takeshita[1] and F. Matsuno[2]

[1]Okayama University, 3-1-1, Tsushimanaka, Okayama-city, Okayama, Japan, kazuyuki@sys.okayama-u.ac.jp
[2]Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama, Japan, matsuno@dis.titech.ac.jp

Abstract

Reinforcement learning is very interesting for robot learning. However, there are some significant problems in applying conventional reinforcement learning algorithms to the robot with many degrees of freedom, because the size of exploration space increase exponentially with increase of degrees of freedom, and it makes it impossible to accomplish learning process. On the other hand, animals and humans can learn and accomplish various tasks using many redundant degrees of freedom of the body in spite of the exploration space is very huge.

In this paper, we summarize our previous works of QDSEGA, which is reinforcement learning algorism for redundant robot, and discuss how to apply reinforcement learning to redundant robots. We introduced the body image, which is inspired by studies of animals, and propose a hypothesis that the body image makes it possible to realize the ability to restrict the exploration space. To demonstrate the validity of the hypothesis, simulations and experiments have been carried out. As a result effective behaviors have been acquired.

1. Introduction

Reinforcement learning [1] is effective for robot learning [2, 3]. It does not need priori knowledge, and has higher capability of reactive and adaptive behaviors. By applying reinforcement learning to the robot with many redundant degrees of freedom (DOF), adaptive autonomous system can be realized. So applying reinforcement learning to the robot with many DOF is very attractive. However, there are some significant problems in applying conventional reinforcement learning algorithms to them. The most one is the large size of exploration space. The size of exploration space increase exponentially with increase of DOF, and it makes it impossible to accomplish learning process.

On the other hand, animals and humans can learn and accomplish various tasks using many redundant DOF of its body in spite of the exploration space is very huge. They can extract necessary behaviors for each task from the huge number of behaviors that can be realized by many redundant DOF.

Based on the above motivation, we have studied a reinforcement learning algorithm for the robot with many redundant DOF, and we had proposed new reinforcement learning algorithm "Q-learning with dynamic structuring of exploration space based on genetic algorithm (QDSEGA) [4]". The idea to cope with large exploration space in QDSEGA is to extract small closed-subset. A reinforcement learning algorithm is applied to the subset to acquire some knowledge, and the acquired knowledge is utilized to create new subset of exploration space. Repeating this cycle, an effective subset and policy in the subset is obtained. In the QDSEGA, the ability to restrict the exploration space is realized by layered structure, reinforcement learning is realized by Q-learning, and the subset of exploration space is restructured by GA.

The effectiveness of QDSEGA had been demonstrated by simulations of application to the task of obstacle avoidance of manipulator [4] and walking of multi-legged robot[5].

However, the reason why the QDSEGA can be applied to the redundant robots had not been discussed enough. Mainly, the way to realize the ability to restrict the exploration space is very important and interesting, because the frame problem is occurred in realizing the ability. In the QDSEGA, the frame problem has not solved completely, but as for complexity of the body, some part of the frame problem is solved.

In this paper, we summarize our previous works and discuss how to realize the ability to restrict the exploration space. We introduce the body image, which is inspired by studies of animals, into reinforcement learning, and proposed a hy-

pothesis that the body image makes it possible to extract closed-subset from the large exploration space without being bothered by the frame problem. To demonstrate the validity of the hypothesis, simulations and experiment is carried out.

2. Reinforcement learning and the frame problem

Reinforcement learning is a model of learning mechanism of animals and humans. In the reinforcement learning, worth of a behavior is reinforced using a reward that is given from the environment and effective policy to accomplish a task are obtained throw trials and errors.

To accomplish effective learning, composition of state space is very important, because if the state space is very large, the learning process does not converge, and if the state space is too small and some important states are ignored, the learning process does not converged either. Humans and advanced animals have function to choose necessary states and ignore unnecessary stares, so they can learn effectively even in the large exploration space. We consider that they accomplish the function by two ways, one is the unconscious way and the other is the conscious way. The unconscious way is carried out before learning. They can ignore unnecessary states and compose state space before learning, and then reinforcement learning is carried out in the state space. On the other hand, the conscious way is carried out in the learning process. If some states that are ignored before learning is appeared or discovered in the learning process, the state space is recomposed and learning process is continued. In the artificial intelligence, the problem of composition of state space is very difficult. In the conventional way, for example Q-learning, state space must consist of all different states to compose Markov decision process.

The state space contains many unnecessary states to accomplish a task, and the larger state space makes it impossible to complete the learning process. So disregard of unnecessary states are important. However, the way to ignore unnecessary states unconsciously is very difficult and is well known as the frame problem [6]. The frame problem has various means. In this paper we regard the frame problem as the problem which is how to ignore unnecessary condition without priori consideration.

The frame problem have not been solved completely yet. In this paper we focus on a part of the frame problem concerning complexity of a body, and discuss how to solve the problem in QDSEGA.

3. The Body Image

Recently, to explain learning mechanism of humans and animals, body image and body schema has much attention, and neurons that express body image had been discovered in the brain of a monkey by Iriki 1998 [7], and he showed that when some food is around the monkey, the monkey can know whether own hand can reach the food or not, without trial.

In this paper, we consider that the body image makes it possible to predict movements of own body and the animals and humans can move own body at will. In other words, they can express desired movement as series of desired states. When we humans want to move own body, we image a desired state, and the body can move desirably. We do not have to image a force of each muscle, they are adjusted unconsciously, and usually it does not move unexpectedly. So animals and humans can restrict the movable range and it makes it possible to restrict state space concerning the body. When they want to move own body, they have already known how the body moves, unconsciously. It means that they can ignore unrealizable state unconsciously. Therefore we consider that the part of the frame problem that concerning the body is solved by the body image.

In this paper, we define the body of the learning agent as a set of part which the agent can control directly. And we define the body image of learning agent as follows. "When an agent can control its body freely, the agent has a body image."

4. QDSEGA

4.1. Outline

Fig. 1 shows the outline of QDSEGA. At first, small subset of exploration space is extracted from the large exploration space which is composed of state space and action space. Next, reinforcement learning is applied to the subset and some knowledge of the task is obtained. And then new subset of the exploration space is created utilizing the acquired knowledge. The reinforcement learning is applied to the new subset, and by repeating this cycle, effective subset and effective policy in the

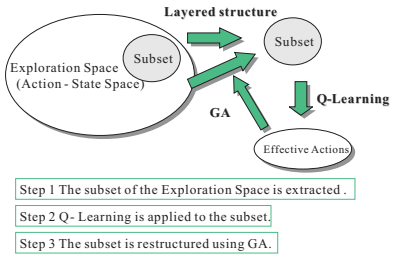

Figure 1: Outline of QDSEGA

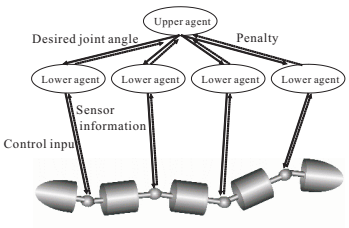

Figure 2: Hierarchal Structure

subset is acquired.

By extracting the closed-subset, it becomes possible to apply the reinforcement learning to the small extracted exploration space. And by utilizing the acquired knowledge to restructure the subset, the search becomes more efficient compare to trial and error only.

The function to extract the subset is realized by layered structure of learning architecture and the reinforcement learning is realized by Q-learning. The subset is restructured using genetic algorithm.

4.2. Interior State and exterior State

In this paper, we define an interior state and an exterior state as follows. The interior state is the set of states that the agent can control directly. And the exterior state is all the state except for the interior state.

4.3. Layered Structure

Proposed algorithm has 2 level layered structures. Fig. 2 shows an example of application to a snake-like robot. An upper agent plans all trajectories of interior state, and passes them to lower agent as a desired state. Each lower agent corresponds to an actuator of the snake-like robot by one to one, and controls each joint angle so that it becomes the desired state.

The lower agents return controlled results to the upper agent. If a lower agent can not realize a desired state that is given by the upper agent, the lower agent returns the information to the upper agent as a penalty.

If the upper agent catches a penalty from any lower agent, the upper agent withdraws the desired states that is given to lower agents, and plans new desired states. So by repeating the learning process, desired states that can not be realized by lower agents are rejected, and a trajectory that complete given task is composed of only realizable states.

By the two way communication between the lower agents and the upper agent, QDSEGA can be applied to the real systems that have dynamics and limitation of ability of actuators.

4.4. Extraction of Closed Subset

A set of desired states that are given by the upper agent to the lower agents at a step can be regarded as an action of reinforcement learning of the upper agent. In case that the lower agents accomplish the action, which means that each interior state converges to the desired state, a set of actions is equivalent to a set of desired interior state. So by restricting usable actions, the upper agent can restrict necessary interior states, and it becomes possible to extract a closed subset from the exploration space. The term "closed" means that the interior state that can be transited by any action in the subset is surely contained in the subset. By this nature, we can apply reinforcement learning to the small subset instead of the large exploration space.

If the lower agents cannot accomplish an action, a penalty is imposed to upper agent and new trial is started form the initial state. So the learning process is preceded in the restricted exploration space.

We can structure the subset of exploration space dynamically by structuring the action space dynamically. In QDSEGA, the actions are structured using genetic algorithm in the learning process of the upper agent.

4.5. Learning Process of Upper Agent

The learning algorithm of the upper agent has two dynamics. One is a learning dynamics based on

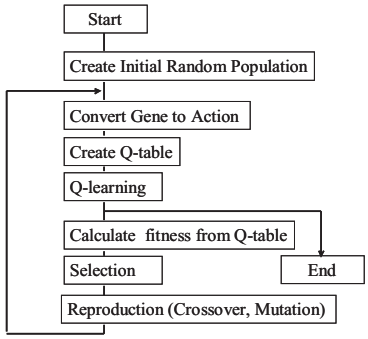

Figure 3: Learning process of the upper agent

Q-learning and the other is a structural dynamics based on Genetic Algorithm. Fig. 3 shows the flowchart of the learning process of the upper agent. Each action is expressed as a phenotype of genes and restructured by Genetic Algorithm. At first, an initial set of population is structured randomly, and the Q-table that consists of phenotype of the initial population is constructed. The Q-table is reinforced using learning dynamics and the finesses of genes are calculated based on the reinforced Q-table. Selection and reproduction are applied and new population is structured. Repeating this cycle, effective behaviors are acquired. Details are written in subsection 4.6.–4.9..

4.6. Encoding

In this algorithm, each individual expresses the selectable action on the learning dynamics. It means that subsets of actions are selected and learning dynamics is applied to the subset. The subset of action is evaluated and a new subset is restructured using Genetic Algorithm. The number of individuals means the size of the subset.

4.7. Create Q-table

To reduce the redundancy of actions, the genes that have a same phenotype are regarded as one action and the Q-table consists of all different actions. The interior states consist of states that can be transited by the generated actions. By repeating the structural dynamics using GA, actions that have a same phenotype are increased, and then the size of the Q-table is decreased.

4.8. Learning dynamics

In this paper, the conventional Q-learning[8] is employed as a learning dynamics. The dynamics of Q-learning are written as follows.

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\{r(s,a) + \gamma \max_{a'} Q(s',a')\} \quad (1)$$

where s is the state, a is the action, r is the reward, α is the learning rate and γ is the discount rate.

4.9. Fitness

4.9.1. Fitness of Q-table

The fitness of genes is calculated at two steps. The first step is regulation of the Q-table and the second step is calculation of the fitness from the regulated Q-table. At first, we calculate the maximum and minimum value of the state as follows.

$$V_{max}(s) = \max_{a'}(Q(s,a')), \quad V_{min}(s) = \min_{a'}(Q(s,a'))$$

Then Q' of the regulated Q-table is given as follows

$$\text{if } Q(s,a) \geq 0 \quad \text{then} \quad Q'(s,a) = \frac{1-p}{V_{max}(s)}Q(s,a) + p \quad (2)$$

$$\text{else} \qquad Q'(s,a) = -\frac{p}{V_{min}(s)}Q(s,a) + p \quad (3)$$

where p is a constant value which means the ratio of reward to penalty. Next, we fix the action to a_i and sort $Q'(s,a_i)$ according to their value from high to low for all states, and we define them as the $Q'_s(s,a_i)$ and repeating the operation for all actions. For example $Q'_s(1,a_i)$ means the maximum value of $Q'(s,a_i)$ and $Q'_s(N_s,a_i)$ means the minimum value of $Q'(s,a_i)$, where N_s is the size of state space. In the second step, we calculate the fitness. The fitness of the gene whose phenotype is a_i is given as follows

$$fit_Q(a_i) = \sum_{j=1}^{N_s} \left(w_j \frac{\sum_{k=1}^{j} Q'_s(k,a_i)}{j} \right) \quad (4)$$

where w_i is a weight which decides the ratio of special actions to general actions.

4.9.2. Fitness of frequency

We introduce the fitness of frequency of use to save efficient series of actions. We define the fitness of frequency of use as follows

$$fit_u(a_i) = \frac{N_u(a_i)}{\sum_{j=1}^{N_a} N_u(a_j)} \quad (5)$$

where N_a is a number of all actions of one generation and $N_u(a_i)$ is the number of times which a_i was used for in the Q-learning of this generation.

4.9.3. *Fitness*

Combining discussion in the subsections 4.9.1., 4.9.2. we define the fitness as follows

$$fit(a_i) = fit_Q(a_i) + k_f \cdot fit_u(a_i) \qquad (6)$$

where $k_f (k_f \geq 0)$ is a constant value to determine the rate of fit_Q and fit_u.

4.10. Selection and Reproduction

Various methods of selection and reproduction that have been studied can be applied to our proposed algorithm. The method of the selection and reproduction should be chosen for each given task. In this paper the method of the selection and reproduction is not main subject so the conventional method is used.

5. Solution for the frame problem in QDSEGA

In this section, we discuss a solution for the part of the frame problem concerning to the complexity of the body in QDSEGA.
In section 3., we have defined a body image as "When an agent can control its body freely, the agent has a body image".

In QDSEGA, the body image for the upper agent is realized by layered structure and lower agents. To move the body of the robot, the upper agent only has to order the lower agent to move the body to a desired state. The order is realized by the distributed control of the lower agents. In a viewpoint of reinforcement learning, this means that it becomes possible to restrict exploration space. Fig. 4 shows the outline of extraction of closed subset. The orders from the upper agent to the lower agents can be considered actions of the upper agent. And when an action is executed by the upper agent, the body moves to a desired state by the distributed control of lower agents, so the actions, that is desired interior state, is equivalent to real interior state. Therefore, by restricting usable actions (STEP 1), the upper agent can restrict movable interior states (STEP 2) and can extract closed-subset of exploration space. The

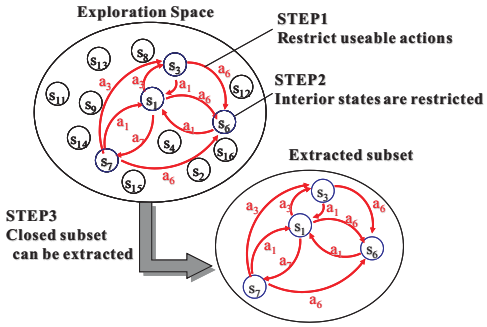

Figure 4: Extraction of Closed Subset

term "closed" means that all interior states that can be changed by any actions in the subset are surely contained in the subset. So reinforcement learning is applicable to the closed-subset of exploration space.

From the method mentioned above, we can find that no pre-consideration is needed to restrict the exploration space. The closed-subset of exploration space for the upper agent can be restricted automatically. So we can conclude that the part of the frame problem concerning to the complexity of the body is solved. Next, we compare the QDSEGA with the conventional reinforcement learning algorithms. In the conventional reinforcement learning, distinction between states of environment and states of the body is not cleared. For the learning agent, these states are confused and a framework of the problem is how to complete a given task using the uncontrollable robot. The agent has to learn two things simultaneously, one is how to control the robot and the other is how to complete the task using the robot. So increase of DOF of the robot makes it difficult to learn in spite of the increase of DOF generally means the increase of mechanical redundancy and it makes it easy to complete tasks at the mechanical viewpoint.

On the other hand, there are two important improvements in the QDSEGA. The first one is division of state space by the layered structure. The state space is divided into interior state, which is the state of the robot, and exterior state, which is the state of environment. It means that the robot becomes the body of the learning agent (upper agent), and a framework of the problem becomes how to complete a given task using own body. The second one is realization of the body image. In the QDSEGA, the upper agent can move the body at will by the distributed control of the lower agents,

and framework of the problem becomes how to complete a given task using own controllable body.

In the QDSEGA, the upper agent can decide a subset of exploration space before learning, and Q-learning is carried out in the subset, and acquires one of solutions for the task from many solutions that is realized by the mechanical redundancy of the body. To acquire more effective solution, new subset is created by GA based on the acquired solution, and Q-learning is carried out again, and repeating this cycle effective subset and policy in the subset is obtained.

Finally we discuss incompleteness of the body image. In the above discussion, we describe that the complete body image make it possible to extract closed-subset of exploration space, and it is important in applying reinforcement learning to the robot with many DOF. However, to realize complete lower agent is very difficult, because ability of actuators are limited and mutual interferences are exist. So realization of some part of desired interior state that is indicated by the upper agent is impossible. In real world, therefore the upper agent has to learn using incomplete body image.

In the QDSEGA, to cope with this problem, the communications between the upper agent and lower agent is two way. If a lower agent can not accomplish a desired state, the lower agent return a penalty to the upper agent. If the upper agent catches a penalty from any lower agent, the upper agent withdraws the desired states that is given to lower agents, and plans new desired states. So by repeating a learning process, desired states that can not be realized by lower agents are rejected, and a trajectory that complete given task is composed of only realizable states.
The function to compensate the incompleteness is very important because it makes it possible to apply the QDSEGA to the real world.

6. Simulation

In this section, we apply the QDSEGA to the task of moving up top of manipulator with non-powerful actuator, and consider the validity of the discussion in section 5.

6.1. Task

We consider a handstand task of a manipulator with non-powerful actuators. Let us define the

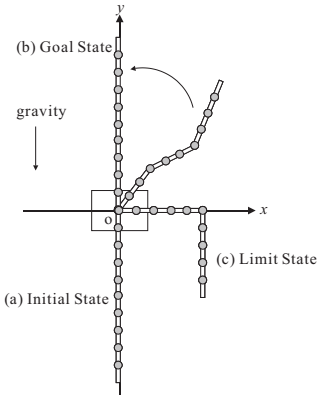

Figure 5: Handstand Task of Manipulator

origin and coordinate as shown in Fig. 5. We consider a vertical plane with gravity. The origin means the fixed end of the manipulator. The goal of the task is to move up the top of the manipulator from initial position (Fig. 5(a)) to goal position (Fig. 5(b)) using non-powerful actuator. We consider that actuators do not have enough torque, and we introduce limitation of the maximum torque of each actuator. Limitation of torque is set as an equivalent torque that the 1st joint keeps the manipulator configuration as shown in Fig. 5(c). We assume that the manipulator move enough slowly and we neglect dynamics of the manipulator except for the gravity.

6.2. Simulation model of the manipulator

The number of links is 10. We regard that all actuators are stepping motors and the angle and angular velocity can be controlled. We assume that all joints are moved to the desired angle with the same constant angular velocity. And when the joint reaches the desired angle, the joint is stopped. And when all joint angles reach the desired angles, the manipulator is stopped.

6.3. Acquired behavior

Fig. 6 shows an acquired behavior. We can find that the task is accomplished using actions that can be realized by lower agents. So we can conclude that the QDSEGA is applicable even if the body image is uncompleted because of limitation of ability of actuators.

Fig. 7 shows the transition of Gains and size of subset of the exploration space. The gain means

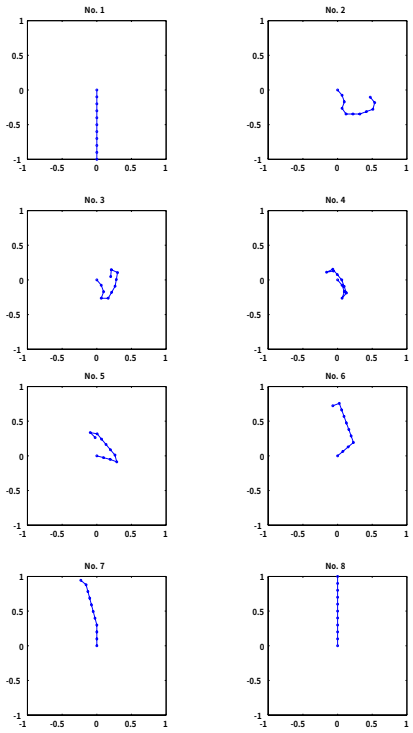

Figure 6: Acquired behavior (Handstand of manipulator with considering limitation of power of actuators)

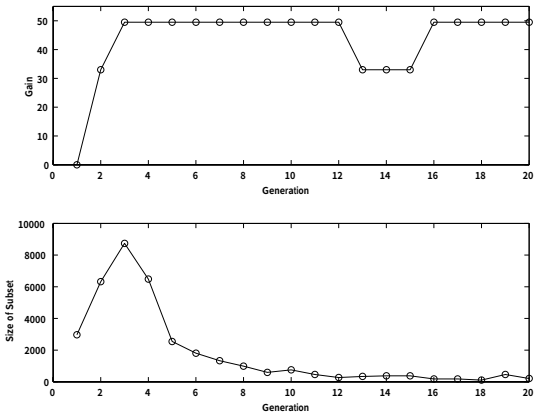

Figure 7: Gains and Number of Actions

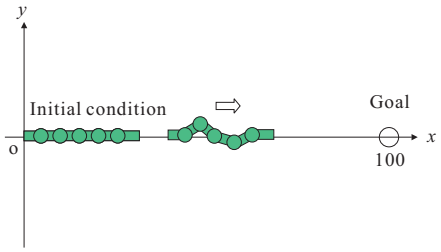

Figure 8: Locomotion Task

the average reward for each step and the size of subset means the size of Q-table at each generation.

From Fig. 7, we can find that the small subset is extracted. In this task, the total size of exploration space is 19^{20}, so even the maximum size of the simulation result, which is about 10000, is very small.

Next we focus the transition of gains and size of subset. At first the gain is 0, which means that necessary actions do not exist in the subset and the task has not been realized. Then the size of subset increase automatically to widen extracted expropriation space. In 2nd generation, task is realized and after 3rd generation, the size of subset is decreased to necessary size but gain is kept at high level. We can conclude that the method to extract subset from exploration space is effective for redundant robot.

7. Application to Real Robot

In this section we apply the QDSEGA to the real 5-link snake-like robot to demonstrate the validity of the body image. In this experiment, simulations are carried out using dynamic model [9] and acquired behaviors in the simulation are applied to the real robot.

7.1. Task

The task is how to get closer to the goal.
Fig. 8 shows the outline of the task. The goal is far enough from the start position and the reward is calculated using the distance from the goal.

7.2. Acquired Behavior

Fig. 9 shows the transient responses of each joint. The circle in the Fig. 9 means the desired state that is acquired by the learning process of the upper agent, and the dotted line means the desired joint angle that is realized by the lower agent, and the line means transient responses of real robot.

We can find that the joint angles of the real robot converge to the desired values that are acquired by the QDSEGA. It means that the robot can be controlled to the desired state and it shows

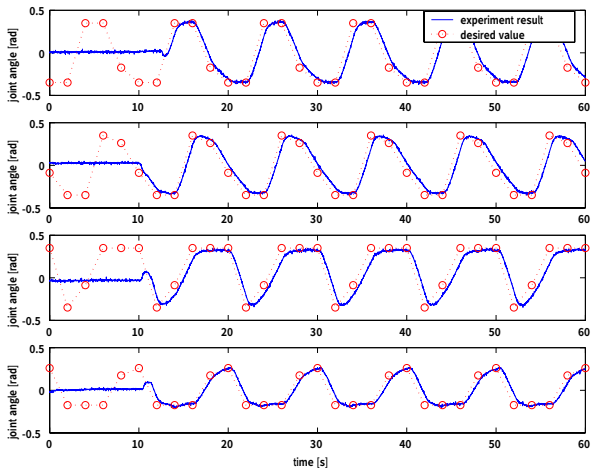

Figure 9: Transient Responses

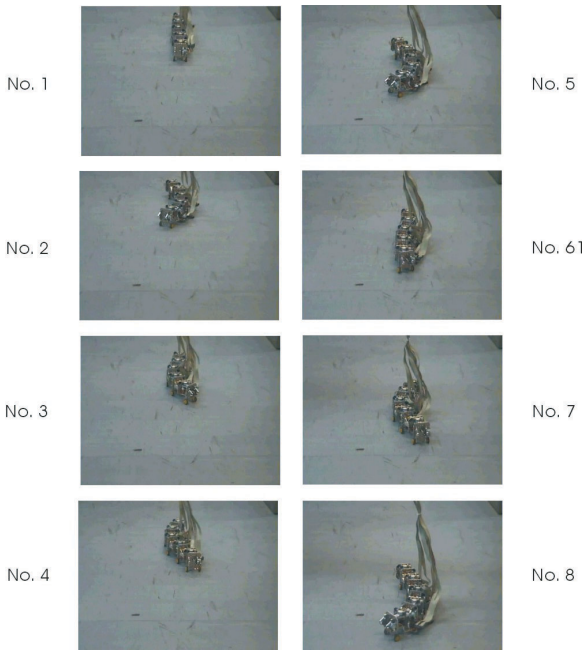

Figure 10: Realized Locomotion

that the body image for the upper agent is realized.

And we can find that the acquired behavior consist of only realizable actions, it means that the discussion of incompleteness of body image is valid and the QDSEGA is applicable for real robot with incompleteness body image.

Fig. 10 shows the realized locomotion by the real robot. We can find that the winding motion is realized and the task is accomplished.

We can conclude that the discussion of the body image in this paper is valid not only for ideal simulated robots but also for the real robots.

8. Conclusion

In this paper, we have summarized our previous works of the QDSEGA, and discussed how to apply reinforcement learning algorithms to redundant robots. We have introduced the body image, which is inspired by studies of animals, and proposed a hypothesis that the body image makes it possible to extract small closed-subset from the large exploration space without being bothered by the frame problem. The effectiveness of the body image in applying reinforcement learning to the robot with many redundant DOF has been discussed and to demonstrate the validity of the discussions, the simulation of a redundant manipulator and experiment of a snake-like robot have been carried out. As the result effective behaviors are obtained. We can conclude that the framework of the body image is effective in applying reinforcement learning to the redundant robots.

References

[1] R. S. Sutton. *Reinforcement Learning: An Introduction*. The MIT Press, 1998.

[2] K. Doya, H. Kimura, and M. Kawato. Neural mechanisms of learning and control. *IEEE Control Syatems Magazine*, 21(4):42–44, 2001.

[3] S. Ushio, M. Svinin, K. Ueda, and S. Hosoe. An evolutionary approach to decentralized reinforcement learning for walking robots. In *Proc. of the 6th Int. Symp. on Artificial life and Robotics*, pages 176–179, 2001.

[4] K. Ito and F. Matsuno. A study of Q-learning: Dynamic structuring of exploration space based on genetic algorithm. *Transactions of the Japanese Society for Artificial Intelligence*, 16(6):510–520(in Japanese), 2001.

[5] K. Ito and F. Matsuno. A study of reinforcement learning for the robot with many degrees of freedom -acquisition of locomotion patterns for multi legged robot-. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pages 3392–3397, 2002.

[6] Rolf Pfeifer and Christian Scheier. *Understanding Intelligence*. The MIT Press, 1999.

[7] A. Iriki. Monkey tool use and the body image. *Shinkei Kenkyu no Shinpo (Advances in Neurological Sciences)*, 42(1):98–105(in Japanese), 1998.

[8] C. J. C. H. Watkins and P. Dayan. Technical note Q-learning. *Machine Learning*, 8:279–292, 1992.

[9] M. Saito, M. Fukaya, and T. Iwasaki. Serpentine locomotion with robotic snakes. *IEEE Control Syatems Magazine*, 22(1):64–81, 2002.