Hand Reaching Movement Acquired through Reinforcement Learning

Katsunari Shibata†, Masanori Sugisaka†& Koji Ito‡

†Dept. of Electrical & Electronics Engineering, Faculty of Engineering, Oita University
700 Dannoharu, Oita 870-1192, JAPAN
‡Dept. of Computational Intelligence & Systems Science, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226-8502, JAPAN
{shibata, msugi}@cc.oita-u.ac.jp, ito@dis.titech.ac.jp

Abstract

This paper shows that a system with two-link arm can obtain hand reaching movement to a target object projected on a visual sensor by reinforcement learning using a layered neural network. The reinforcement signal, which is an only signal from the environment, is given to the system only when the hand reaches the target object. The neural network computes two joint torques from visual sensory signals, joint angles, and joint angular velocities considering the arm dynamics.

It is known that the trajectory of the voluntary movement of human hand reaching is almost straight, and the hand velocity changes like bell-shape. Although there are some exceptions, the properties of the trajectories obtained by the reinforcement learning are somewhat similar to the experimental result of the human hand reaching movement.

Key Words: Direct-Vision-Based Reinforcement Learning, Neural Network, Hand-Eye Coordination, Reaching Task, Trajectory Planning

1 Introduction

It has been proposed that the visual sensory signals are put into a neural network directly and the network is trained to output appropriate motion signals by reinforcement learning[1]. The continuous state space can be formed adaptively and purposively through the learning. It shows the possibility that the reinforcement learning is useful not only for the motion planning, but for the total functions from sensors to motors, including recognition, attention, and so on. This is called Direct-Vision-Based Reinforcement Learning.

It has been shown that hand-eye coordination can be obtained by the combination of reinforcement learning and neural network in a robot arm reaching task[2]. It has been realized only by adding the joint angles as input signals in Direct-Vision-Based Reinforcement Learning. However, the dynamics of the arm was not introduced, but the joint angular velocities were the output of the neural network.

In this paper, the dynamics of the arm is introduced. It means that the joint angular velocities are also the input signals to the neural network, and the output signals are the joint torques. Finally, it is shown that the trajectories and tangential velocity curve of the hand after learning have somewhat similar properties to the experimental result of the human hand reaching movement.

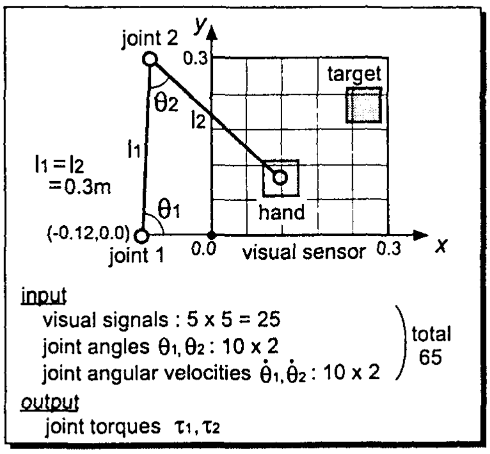

Figure 1: The robot hand-reaching task.

2 Task Setting

2.1 General Setting

Here the setting of the task as shown in Fig. 1 is described. The visual sensor has $5 \times 5 = 25$ cells and the output of each cell is the area ratio occupied by the target or the robot's hand against its receptive field. Below here, the left-bottom corner of the visual sensor is supposed to be the origin. The size of each visual cell is 0.06×0.06, and the size of the target and hand is also the same. The target and the hand cannot be distinguished with each other on this visual sensor.

The target is located randomly in the range where the whole target can be caught in the visual field, i.e., $0.03 \leq x, y \leq 0.27$. The initial hand location is also chosen randomly. In the early phase of the learning, it is chosen from only around the target, and according to the progress of the learning, the range becomes wider gradually until $-0.09 \leq x, y \leq 0.39$ under the condition of $((x-0.12)^2 + y^2) \leq (l_1 + l_2)$. So the hand sometimes cannot be caught by the visual sensor initially after some trials. The base of the arm (joint 1) is fixed at (-0.12, 0.0). The both joint angles are limited in the range of $0 \leq \theta_i \leq \pi$, and joint angular velocities are limited in the range of $-\pi \leq \dot{\theta}_i \leq \pi$. The target is fixed during one trial. But if the hand cannot reach after many time steps, the trial finishes with no reward, and in some following trials, the target is moved towards the hand gradually. The length of each link is the same as the side of the visual sensor. There is