

Short paper

Acquisition of a page turning skill for a multifingered hand using reinforcement learning

JUN UEDA^{1,*}, RYOJI NEGI² and TSUNEO YOSHIKAWA³

¹ Graduate School of Information Science, NARA Institute of Science and Technology (NAIST),
8916-5 Takayama-cho, Ikoma city, Nara, 630-0192, Japan

² Dome Ltd., 198-1 Hanajiri-cho, Yase, Kyoto 601-1251, Japan

³ Department of Mechanical Engineering, Kyoto University, Yoshida-Honmachi,
Kyoto 606-8501, Japan

Received 28 April 2003; accepted 26 May 2003

Abstract—This paper proposes a method of acquisition of a page turning skill for a multifingered robotic hand using reinforcement learning. The goal of this paper is generation of the manipulation skill of a flexible object without its explicit model and tactile sensation during its control. In this paper, a page turning task is considered as an example of such tasks, where a sheet of paper, generally used in books, is a flexible object. The fingertip trajectories are obtained by reinforcement learning based on simulation. The reward considering a friction condition is given, so that a page turning skill without slip between the finger and the paper is obtained. The validity is confirmed by an experimental system which consists of a 2-d.o.f. manipulator and two 2-d.o.f. fingers.

Keywords: Flexible materials; manipulation; reinforcement learning; page turning skill; multifingered robot hand.

1. INTRODUCTION

Dexterous manipulation for multifingered robotic hands has long been studied in the robotics field. Many studies have been presented concerning manipulation of rigid objects such as stability of grasping, calculation of internal and external force [1], grasping planning based on LMI [2], and envelope grasping [3]. Recently, the research interest has expanded to the manipulation of flexible/deformable objects [4, 5]. Many problems arise in the manipulation of flexible objects compared to rigid manipulation. They mainly come from the difficulties of modeling, estimating and sensing a great number of parameters of the object. Therefore, the

*To whom correspondence should be addressed. E-mail: uedajun@is.aist-nara.ac.jp

manipulation of flexible objects has not been established in comparison with that of rigid objects.

A page turning task is one example of manipulation of flexible objects. It seems that humans can access the information of books by handling the pages without any difficulties. However, from the viewpoint of robotic manipulation, this kind of task is very difficult to accomplish, and it is thought that an explicit model with its real-time calculation and sensitive force sensors is considered necessary for control.

This paper proposes a method of acquisition of a page turning skill for a multifingered robotic hand. The fingertip trajectories for turning a page are obtained through reinforcement learning [6]. The reward considering a friction condition (friction cone) is given, hence a page turning skill without slipping between the finger and the paper is obtained. The benefit of this approach is that both an explicit model of the flexible object and tactile sensation such as slip/force information are unnecessary for the control.

A model of a page is unnecessary either for the learning if the learning is performed all in the real environment. However, learning in the real environment consumes much time and seems inefficient. Therefore, the learning is firstly performed through a simulation to shorten the calculation time. Next, the validity is confirmed by implementing an experimental system which consists of a 2-d.o.f. manipulator and two 2-d.o.f. fingers.

2. ACQUISITION OF A PAGE-TURNING SKILL

2.1. Target page turning

As shown in Fig. 1, two fingers are used for pinching and manipulating a page. In this paper, we consider the page turning task shown in Fig. 1. As a matter of course, there exist several methods for handling flexible objects; however, we consider the method shown in Fig. 1 gives stable manipulation and is suitable for a robotic hand.

In order to achieve stable manipulation, slip between the paper and the fingers should be avoided. In this paper, the trajectories of the robot fingers are obtained by which a desired page turning task is achieved and the friction condition considering the friction cone of each finger is always satisfied. The trajectories of the fingers and the hand are calculated by reinforcement learning where the friction condition and the achievement of the page turning are set as the reward. In our approach, the obtained trajectories are executed fully based on position control; an explicit model of the object with its real-time calculation and force/tactile sensors are unnecessary for the control.

2.2. Reinforcement learning for a page turning skill

Reinforcement learning is briefly described in this section. For more detail, see Ref. [6]. Reinforcement learning is the problem faced by an agent that learns

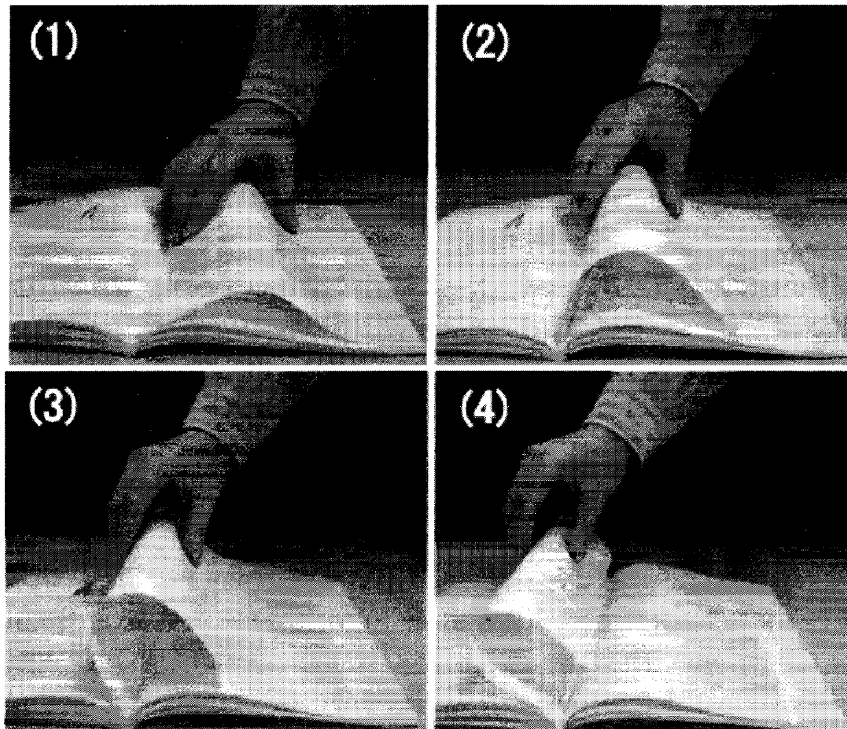


Figure 1. Page turning.

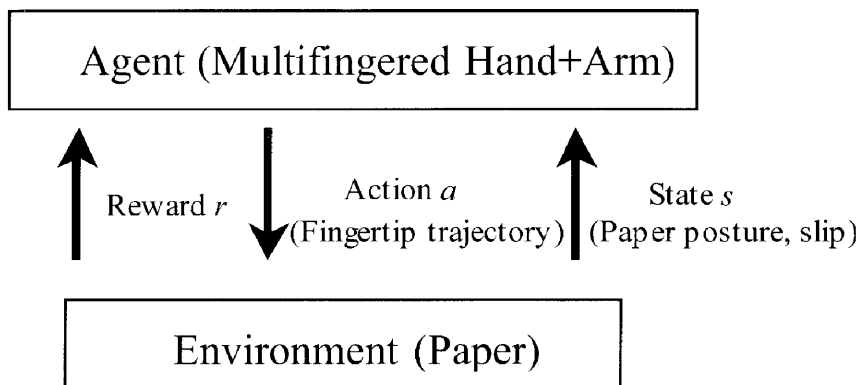


Figure 2. Reinforcement learning.

behavior through trial-and-error interactions with a dynamic environment. The agent acts in the environment so as to maximize rewards given by its designer. The advantage of this learning method is that the learning is done without *a priori* specific knowledge of the environment or the agent itself. Figure 2 shows the fundamental framework of reinforcement learning.

In this paper, a multifingered robotic hand corresponds to the agent, and a paper and a floor correspond to the environment. The agent learns the trajectories of the fingers without *a priori* knowledge of the model and the parameter of the paper. The reward is set larger if the page turning task is achieved by shorter steps. In contrast, a negative reward is given if slip occurs on the fingertips. As a result, a page turning skill without slip is obtained in as short a number of steps as possible.

2.3. Q-learning

A representative example of the reinforcement learning algorithm is Q-learning. This algorithm assumes that the environment is a discrete Markov process, thus both states and actions are discrete. An agent observes discrete states of the environment s and executes discrete actions a . The agent observes state s , takes action a , observes new state s' and receives immediate reward r , in each discrete time step. An agent estimates $Q^*(s, a)$ through this trial-and-error. $Q^*(s, a)$ denotes the quality value if the agent takes the optimal actions in every step after taking a under state s . If $Q^*(s, a)$ is given, an action which maximizes $Q^*(s, a)$ is called the optimal action. The set of the optimal actions is called the optimal policy. The Q value is updated as follows:

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

where α is a learning rate ($0 < \alpha \leq 1$), and γ is a discount rate ($0 \leq \gamma < 1$).

The goal of the reinforcement learning is to maximize the reward; however, the selection of the actions based on only the Q value does not always encounter the optimal policy. In order to check a wide range of actions interacting with the environment, a random action should be included based on ϵ -greedy selection. In this method, an agent takes a random action with probability ϵ , otherwise it takes an action which maximizes the value Q .

3. MODELING OF A PAPER

A model of a paper is unnecessary for the learning algorithm if the learning is all performed in the real environment. However, learning in the real environment consumes much time and seems inefficient. In order to reduce the learning time, it is favorable that the learning is first done under a simulation environment and then the result is implemented to the real system. Note that the model described in this section is used only for the learning and not used for control.

The paper and the solid floor as the environment are modeled as follows. To simplify the real environment as a distributed parameter system, a lumped parameter model is considered. Yoshikawa *et al.* proposed a modeling of flexible link robots and its control [7]. In this paper, a sheet of a paper is divided into three parts (beams) shown in Fig. 3.

A spring-mass-damper model of a beam shown in Fig. 4 is considered as one of the parts which is presented by Yoshida *et al.* [8] for paper feed mechanisms. Each beam is divided into n linear elements where m_i , l_i , ϕ_i , γ_i , F_{xi} , and F_{yi} are the mass of the i th element, the length of the i th link, the bending angle, the angle by shear deformation, and the forces along the x axis and z axis, respectively. m_i is obtained as:

$$m_i = \frac{1}{2}\rho(l_{i-1} + l_i), \quad (2)$$

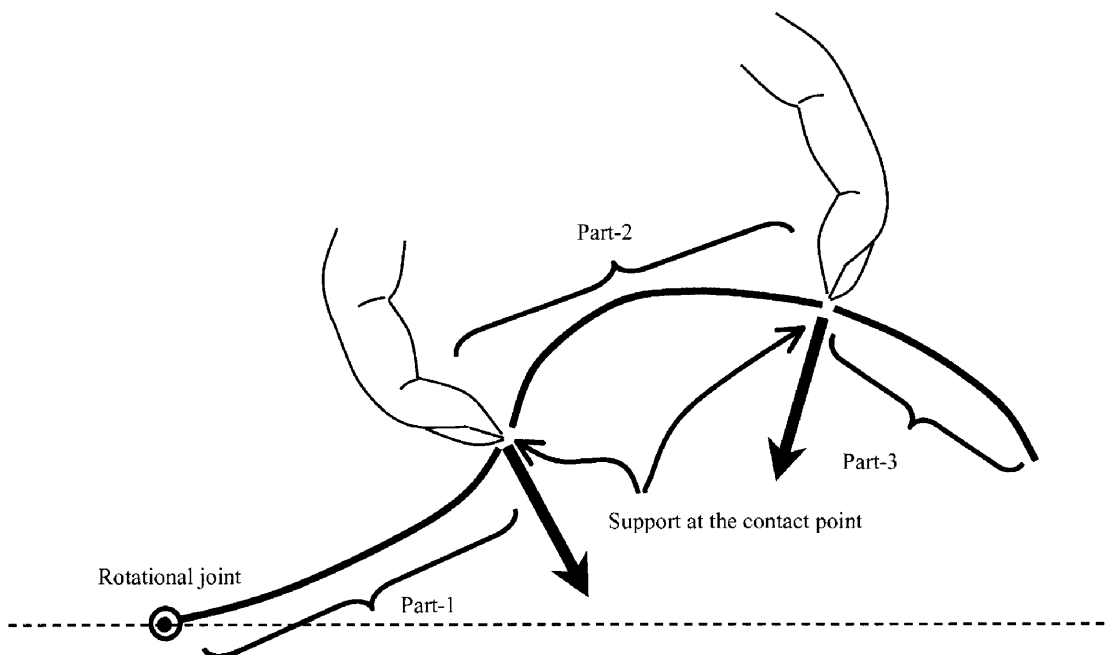


Figure 3. Modeling of a paper.

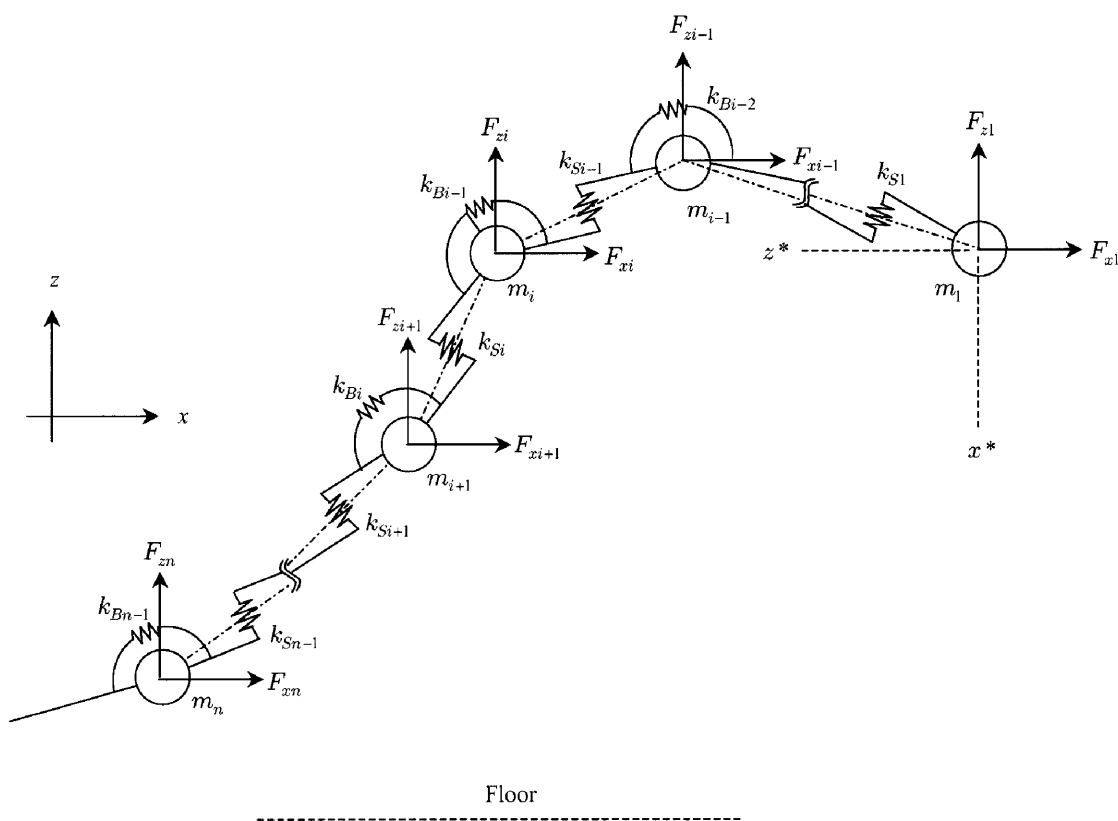


Figure 4. A spring-mass-damper model.

where ρ is density, k_{Bi} and k_{Si} are the stiffness of the rotating and linear springs which represent the bending and the shear stiffness. E , I , G and A denote Young's modulus, geometrical moment of inertia, shear modulus and the cross-section of the beam. χ is the ratio between the average shear stress and the maximum shear stress on the neutral axis.

Let x^* , z^* be the displacement of the constrained edge along the x - and z -axes. For example, the potential energy V of a beam is given by:

$$\begin{aligned}
 V = & \sum_{i=i_0}^N \frac{k_{Bi}}{2} (\phi_i - \phi_{i+1})^2 + \sum_{i=1}^n \frac{k_{Si}}{2} l_i^2 \gamma_i^2 + \sum_{i=1}^n m_i g \left\{ \sum_{j=n}^i l_j \sin(\phi_j + \gamma_j) \right\} \\
 & - \sum_{i=1}^n F_{xi} \left\{ \sum_{j=n}^i l_j \cos(\phi_j + \gamma_j) \right\} - \sum_{i=1}^n F_{zi} \left\{ \sum_{j=n}^i l_j \sin(\phi_j + \gamma_j) \right\} \\
 & + g_x \left[\sum_{i=1}^n l_i \{1 - \cos(\phi_i + \gamma_i)\} - L + x^* \right] \\
 & + g_z \left\{ \sum_{i=1}^n l_i \sin(\phi_i + \gamma_i) - z^* \right\}, \quad (3)
 \end{aligned}$$

where the forth and fifth terms represent the potential energy caused by the external force F_{xi} and F_{zi} , the sixth and seventh terms represent the constrained condition of the edge if necessary. L denotes the length of the beam.

Each beam is connected by a spring to other beams at the contact point where the position is constrained by the fingertip. As shown in Fig. 3, the boundary condition of part-1 and part-2 is simple support, and that of part-3 is simple support and free. Note that one edge of part-1 is supported by a rotational joint without a spring to the ground. In the free condition of the part-3, we can set as $g_x = 0$ and $g_z = 0$. The floor is modeled as a solid plane whose constraint is represented by a penalty method. When the two geometric constraints of the robotic fingers are given, the

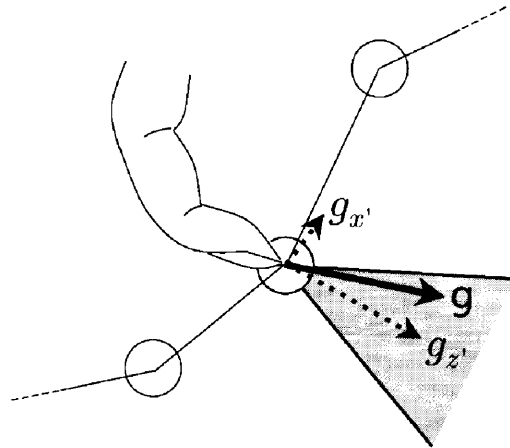


Figure 5. Friction condition by friction cone.

shape of the paper ϕ and γ , and the contact force at the fingertips are obtained by a convergence calculation in which the sum of the potential energy of three beams is minimized.

In order to include the friction condition between each finger and the paper, a slip detection base on the friction cone is added. Let $\mathbf{g} = [g_x, g_z]$ be the force which is applied to the element contacting the fingertip. As shown in Fig. 5, $g_{x'}$ and $g_{z'}$ are the tangential and the normal force of \mathbf{g} . The broken line of Fig. 5 shows the tangent at the contact point. The non-slipping condition is, given that \mathbf{g} is inside the friction cone:

$$\mu g_{z'} \geq |g_{x'}|, \quad (4)$$

where μ is a static friction coefficient.

4. SIMULATION-BASED LEARNING

4.1. Parameters of the paper

Table 1 shows the parameters of the paper [8]. Let the total length of the paper be 0.3 m, and the contact points be 0.12 and 0.24 m from the supported edge of the paper. The paper is modeled by 30 elements of lumped mass; $l_i = 0.01$ m ($i = 1, \dots, 30$). The trajectories of the fingers are obtained by the learning algorithm. The penalty method is iterated until the magnitude of denting of all the elements to the solid floor is less than 5×10^{-3} m.

4.2. Set of finger actions and states

A 2-d.o.f. planar manipulator with two 2-d.o.f. planer fingers is considered for the page turning task. The set of finger actions is shown in Fig. 6. The actions from 1 to 4 represent the movement of the left fingertip along the x - and y -axis, respectively. Similarly, the actions from 5 to 8 represent the movement of the right fingertip. The actions from 9 to 12 represent the movement of the hand part, i.e. both fingertips

Table 1.
Parameters of the paper

Parameter	Value
E	3430 MPa
G	350 kPa
χ	0.67
A	$8.8 \times 10^{-6} \text{ m}^2$
I	$4.7 \times 10^{-15} \text{ m}^4$
ρ	0.0083 kg/m
μ	1.1

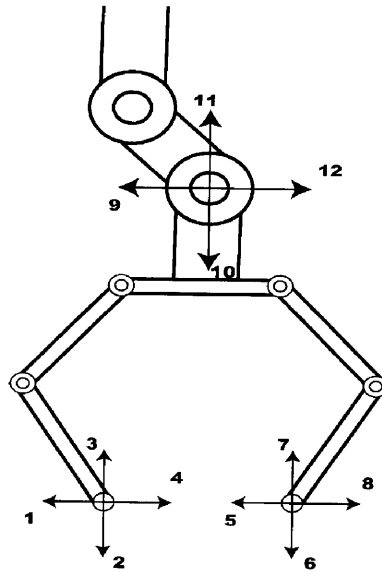


Figure 6. Set of finger actions.

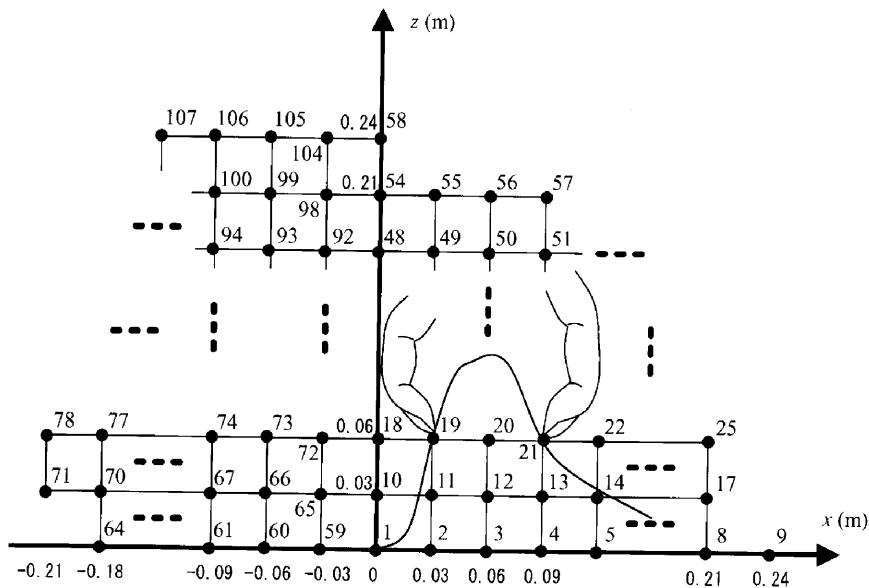


Figure 7. Set of states.

move simultaneously according to the hand movement. All the moving distance is set as a discrete 0.03 m in each step and the diagonal movement is not considered.

Figure 7 shows the set of states of the fingertips. The paper is supported by a rotational joint at $[0, 0]$. Grids are defined on the x - z plane and numbered as shown in Fig. 7. The number of the state totals 3103 including 29 grids for the contact point of 0.12 m and 107 grids for the contact point of 0.24 m. The fingertip moves from a grid to its neighborhood in each step. We assume the page turning task is completed if the contact point of 0.12 m crosses the line $x = -0.09$ m.

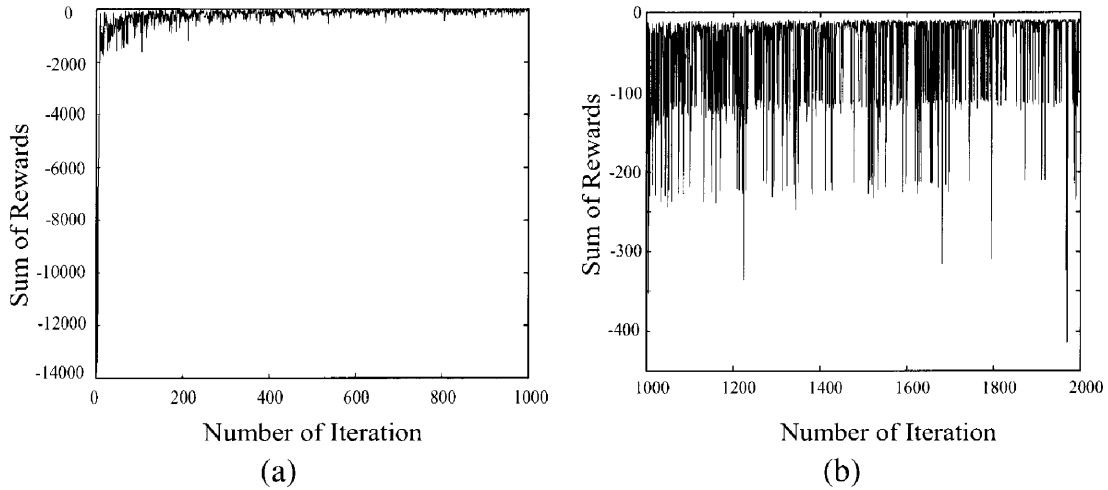


Figure 8. Change of the reward. (a) 0–1000 times. (b) 1000–2000 times.

4.3. Parameters and reward for reinforcement learning

Let the parameters of the reinforcement learning be $\epsilon = 0.1$, $\gamma = 0.9$ and $\alpha = 0.1$. The reward is set as $r = -1$ is given at each step, $r = -100$ is given if a slip occurs and $r = 1$ is given if page turning is completed. In addition, $r = -100$ is given if the bending angle exceeds 160° , since page turning without folding the paper is favorable. As a result, a page turning skill is obtained aiming at as short a number of steps as possible, avoiding slip and folding.

4.4. Result of simulation

Figure 8 shows the change of the reward versus the iteration time. Figure 8a shows the result from 0 to 1000 times and Fig. 8b shows that from 1001 to 2000 times. The range of the vertical axis is changed according to the magnitude of the reward. It can be observed that the learning is performed at the beginning of the iteration and almost converged in Fig. 8b. The optimal policy is obtained as [9, 5, 5, 5, 11, 3, 9, 9, 9, 9, 9, 9, 2, 9], according to the notation of Fig. 6.

Figure 9 shows the finger positions and the shape of the paper by the obtained optimal policy. In Fig. 9, the circles show the fingertip positions, the broken line shows the floor, the heavy line is the shape of the paper and t denotes the number of time steps. Note that the trajectories of the fingers are generated only by the learning algorithm and no intentional operations are performed in order to assimilate the human-like method shown in Fig. 1.

5. EXPERIMENT

5.1. Experimental set-up

Figure 10 shows a hand–arm robot system used in the experiment. The system consists of a PA-10 manipulator (Mitsubishi Heavy Industries) as an arm and a

hand, and two miniature robotic fingers (Daikin) are attached on the manipulator's tip. A rubber semi-sphere is attached to each fingertip in order to achieve soft

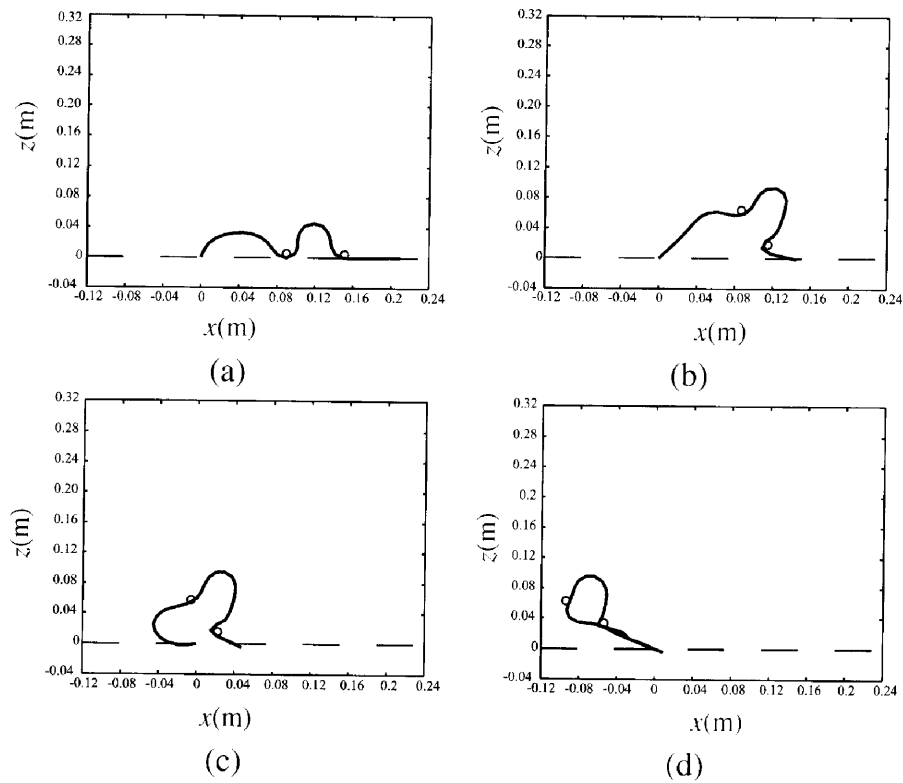


Figure 9. Result of the simulation. $t = 4$ (a), 6 (b), 9 (c) and 12 (d).

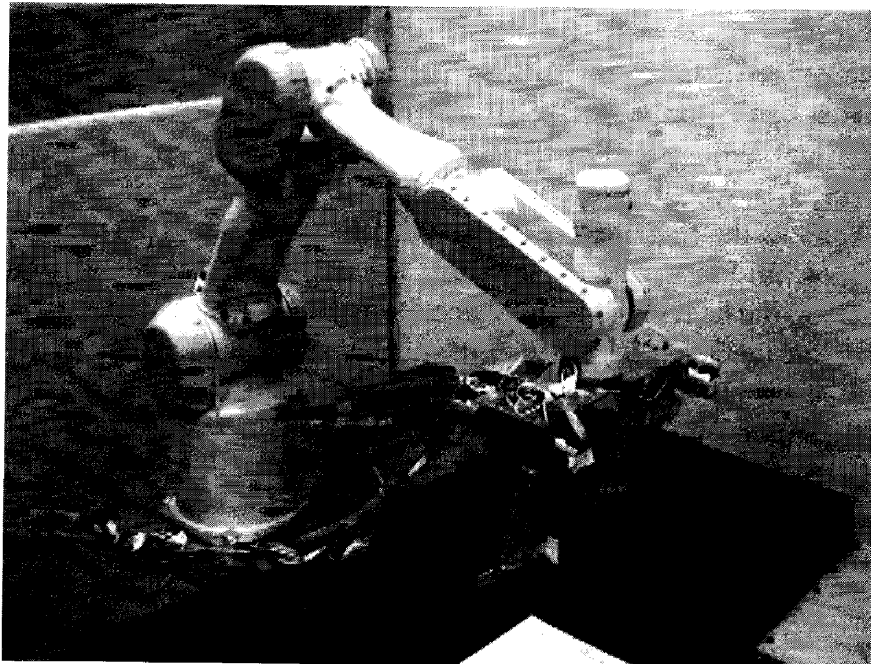


Figure 10. Experimental hand-arm system.

contact. The manipulator has 7 d.o.f. originally, but its motion is constrained to planar 2 d.o.f. in this experiment. Each finger has planar 2 d.o.f. thus the system has a total of 6 d.o.f. in order to achieve the actions shown in Fig. 6. The joint angles of the manipulator and fingers are measured and sent to the client PC (MMX Pentium 233 MHz, VxWorks) through bus interfaces. The commands based on the learning results are sent from a host PC to the controller via the client PC. The page turning task is achieved only based on position control, thus no force/tactile sensor or external camera is used in this system.

5.2. Result of experiment

The result obtained by the learning is directly implemented to the experimental system without relearning in the real system. Figure 11 shows the experimental results. Two markers are attached in the neighborhood of the contacting points. The experimental result corresponds well to the simulation result. Page turning without slip is successfully performed utilizing the deformation of the paper.

Figure 12 shows the result judged that the objective page turning fails since a slip occurs in the simulation. The policy is [9, 5, 5, 5, 11, 9, 9, 9, 9, 9, 2, 9] according

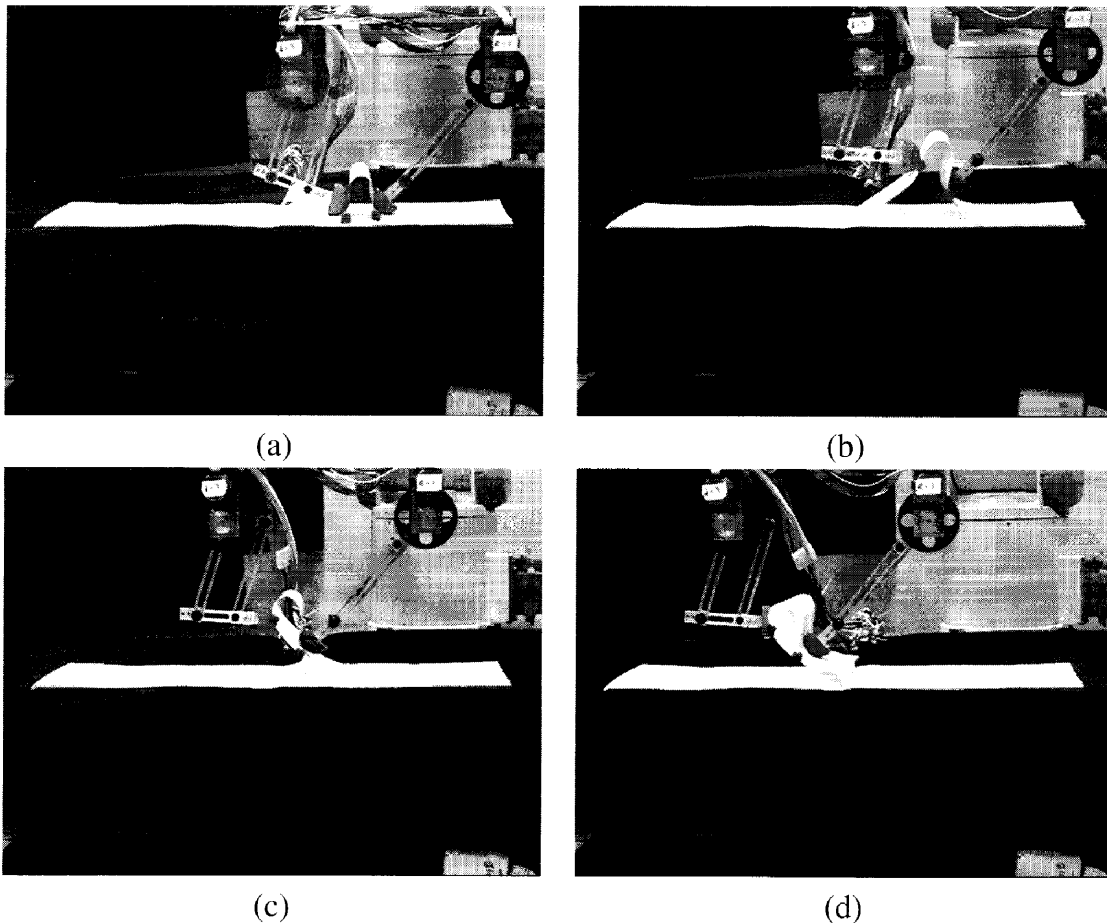


Figure 11. Experimental result (successful page turning). $t = 4$ (a), 6 (b), 9 (c) and 12 (d).

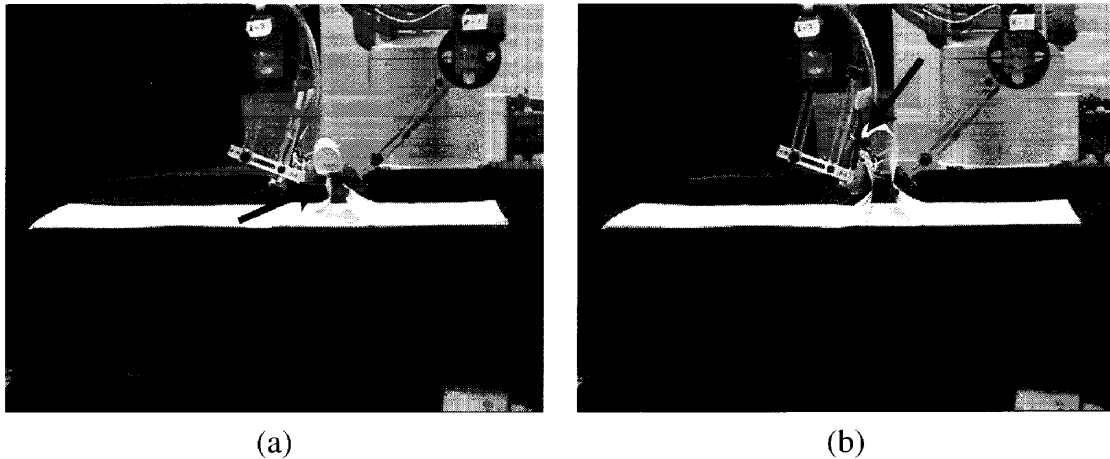


Figure 12. Experimental result (example of failure: the arrow indicates the slip occurring at the contact point). $t = 5$ (a) and 6 (b).

to the notation of Fig. 6. Compared to the result shown in Fig. 11, the actions are the same till the fifth step, but different actions are taken after the sixth step. As shown in Fig. 12, a slip occurs on the left fingertip just after the different action is taken. Note that the page turning task itself has been achieved even in Fig. 12. If the objective is to simply achieve page turning without considering slip, the reward concerning slip can be eliminated. However, we are seeking a narrower class of page turning skills in this paper, by which a stable manipulation without slip can be performed. The experimental result shows the possibility and validity of this approach.

6. CONCLUSIONS

In this paper, a method for acquisition of a page turning skill was proposed. It seems that a page turning task without slip is complex for multifingered robotic hands, and an explicit model and its real-time calculation, force or tactile sensors are needed. However, it has been shown that a simple approach, based on discrete positioning control, works sufficiently well for manipulating flexible objects.

First, a simulation environment has been developed to reduce the learning time where a lumped parameter model is used. The fingertip trajectories were generated by the reinforcement learning in which the friction condition of both fingertips were considered. Secondly, the result obtained by simulation was directly implemented on the experimental system without relearning in the real system. The obtained skill was performed in the experimental system consists of a 2-d.o.f. manipulator and two 2-d.o.f. fingers. We confirmed that page turning without slip had been accomplished in a practical system. The result suggests that the obtained skill is relatively robust against the modeling error between the simulation and the real system. In addition, the learning algorithm itself does not depend on the object, hence the method can be easily applied to the manipulation of other flexible objects.

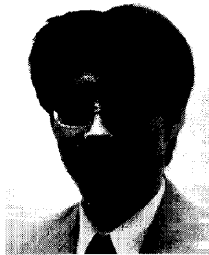
If the objective is to simply achieve page turning without caring about slip, the skill can be relatively easily obtained by eliminating the reward concerning slip. However, we are seeking a narrower class of page turning skills in this paper, by which stable manipulation without slip can be achieved. The results have shown the possibility and validity of this approach.

Future work includes (i) achieving smooth movement of the fingertips, (ii) on-line error compensation using visual feedback and (iii) expanding to three-dimensional manipulation.

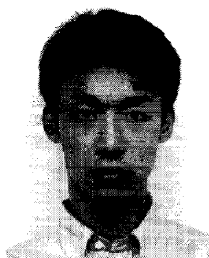
REFERENCES

1. Y. Nakamura, K. Nagai and T. Yoshikawa, Dynamics and stability in coordination of multiple robotic mechanisms, *Int. J. Robotics Res.* **8**, 44–61 (1989).
2. M. Buss, H. Hashimoto and J. B. Moore, Dexterous hand grasping force optimization, *IEEE Trans. Robotics Automat.* **12** (4), 507–512 (1996).
3. K. Harada and M. Kaneko, A sufficient condition for manipulation of envelope family, *IEEE Trans. Robotics Automat.* **18** (4), 597–607 (2002).
4. D. Henrich and H. Worn (Eds), *Robot Manipulation of Deformable Objects*. Springer, Berlin (2000).
5. H. Nakagaki, K. Kitagaki and H. Tsukune, Study of insertion task of a flexible beam into a hole, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Nagoya, pp. 330–335 (1995).
6. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA (1998).
7. T. Yoshikawa and K. Hosoda, Modeling of flexible manipulators using virtual rigid links and passive joints, *Int. J. Robotics Res.* **15** (3), 290–299 (1996).
8. K. Yoshida and M. Kawauchi, The analysis of deformation and behavior of flexible materials (1st report, study of spring-mass beam model of the sheet), *Trans. Jpn Soc. Mech. Eng.* **552** (58), 1474–1480 (1992) (in Japanese).

ABOUT THE AUTHORS



Jun Ueda received the BS, MS and PhD degrees from Kyoto University, Kyoto, Japan, in 1994, 1996 and 2002, respectively, all in Mechanical Engineering. During 1996–2000, he was a Research Engineer of the Advanced Technology R&D Center, Mitsubishi Electric Corp., Japan. In 2002, he joined the Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Japan, where he is currently a Research Associate of the Robotics Laboratory. His research interests include vibration control, robust control, mechanical design and manipulation.



Ryoji Negi received the BS degree from Kyoto University, Kyoto, Japan, in 2002 in Mechanical Engineering. From 2002, he has been an Engineer with Dome Co. Ltd., Japan, and has been engaged in the development of racing cars. His research interests are in multifingered robotic hands and skill learning.



Tsuneo Yoshikawa received the PhD degree in Applied Mathematics and Physics from Kyoto University, Kyoto, Japan, in 1969. Since 1969 he has been with Kyoto University, where he is currently Professor of Mechanical Engineering. From 1973 to 1975, he was an NRC Resident Research Associate at NASA Marshall Space Flight Center, Huntsville, AL. He is the author of *Foundations of Robotics* (MIT Press, 1990). His current research interests include robotics, haptic virtual reality and control of non-linear mechanical systems. He has received several awards including the 1995 JDSMC Best Paper Award from the Dynamic Systems and Control Division of the ASME, and the 1997 ICRA Best Conference Paper Award from the IEEE Robotics and Automation Society. He is a fellow of the IEEE.