Acquisition of Synergy for Low-dimensional Control of Multi-fingered Hands by Reinforcement Learning

Kazuki Higashi
Graduate School of Engineering Science, Osaka University, 560-8531 Osaka, Japan
E-mail: higashi@hlab.sys.es.osaka-u.ac.jp

Tomohiro Motoda
Graduate School of Engineering Science, Osaka University, 560-8531 Osaka, Japan

Akiyoshi Hara
Graduate School of Information Science and Technology, Osaka University, 565-0871 Osaka, Japan

Kensuke Harada
Graduate School of Engineering Science, Osaka University, 560-8531 Osaka, Japan

Abstract

Synergy is the method that reduces the control inputs of a multi-fingered hand and is utilized for designing underactuated robotic hands and efficient control. Calculating conventional synergies depends on the measured human grasping postures. Therefore, preparing synergies for the not-human-like multi-fingered hands is challenging. We propose a reinforcement learning platform for acquiring synergies of a multi-fingered robotic hand through learning a grasping task. The learning process automatically generates postures for creating synergies so that this system can prepare synergies for any robotic hand. Experiments show that this reinforcement learning platform improves learning tasks and acquires the synergy that is suitable for the learned task.

Keywords: Synergy, Reinforcement Learning, Dimensionality Reduction, Multi-fingered Hand

1. Introduction

Synergy is a practical approach to decrease the control inputs of a robotic system with high degrees of freedom (DoFs) [1]. This unique advantage of synergy makes the multi-fingered hand, which is too complicated for industrial use, easy to control, and more adaptive to various tasks than simple robotic hands. For reducing the control inputs, the synergy is calculated by dimensionality reduction methods such as the principal component analysis and the gaussian process latent variable model. Since these methods are data-driven, many postures of a robotic hand performing tasks must be gathered. Conventional studies generally use human hand postures and transfer the postures to robotic hands [2], [3]. These methods allow for creating a low-dimensional control space for synergy while maintaining the dexterity of human manipulation. Synergies made from human postures are generally applied to humanoid robotic hands. Besides, some methods for making synergies for non-humanoid multi-fingered robotic hands have been proposed. Ficuciello et al. [4] developed a mapping method from human hand posture to robotic hands. However, these methods cannot guarantee synergies to utilize non-humanoid robotic hands' dexterity fully. Because kinematics between the human hand and the non-humanoid hand, such as the number of joints and the dimension, are entirely different. Therefore, more suitable synergies must exist that can fully exploit the non-human robotic hand's kinematics than the synergies made by human postures.

This paper proposes a reinforcement learning (RL) platform to acquire synergies for every kind of robotic hand. First, a multi-fingered hand that acquires synergy through RL is trained on a specific task. As the learning progresses, the postures of successful tasks are accumulated in a database. The principal component analysis is applied to the postures registered in the database to compress the dimension of control inputs. We design a reward for the learning so that the compressed control space can express many postures in which the task can be performed. This allows us to obtain synergy with low dimensional controllability and high task performance. Simulation-based experiments confirm the effectiveness of the method. It is also shown to be effective in improving learning efficiency.

Section 2 describes the details of the method, Section 3 describes the experimental and evaluation methods, Section 4 discusses the experimental results, and Section 5 concludes this paper.

2. Proposed Method

This section explains the details of the proposed RL platform.

2.1 *Reinforcement Learning*

The RL algorithms used in this platform are the Deep Deterministic Policy Gradient (DDPG) method [5], [6], and the Hindsight Experience Replay (HER) [7]. An overview of the learning system is shown in Fig.1. DDPG consists of an actor network that outputs actions and a critic network that computes a Q-function for evaluating actions in a given state. The actor network takes only the state s as input, while the critic network takes the state s and action a as input. These networks consist of an input layer, three all-coupled layers with 256 units in each layer, and an output layer.

Because tasks performed by a multi-fingered hand are represented by continuous values for both actions and states, the number of possible states and actions is significant, and it is often the case that exploration is not rewarded. HER adds experience four times per experience $(s_t \mid\mid g, a_i, r_i, s_{t+1} \mid\mid g)$. The reward r, r' is recalculated for all recorded experiences, along with the calculation of the reward for the successful completion of the task, as described below. Mini-Batch Learning is performed on each network of DDPG using all the experiences thus obtained.

2.2 *Reward system*

Fig. 2 shows the rewards for successful and failed tasks. r_{failure} is a reward in the case of task failure. In the case of task success, we consider a reward r_{success} and a penalty e which is the approximation error of the successful grasping posture projected onto the synergy space. λ means a coefficient of feedback for the approximation error e. *Synergy space* is a principal component space from the dataset of successful grasping postures. We aim to learn a task while increasing the number of postures during task execution that can be accurately represented in the synergy space.

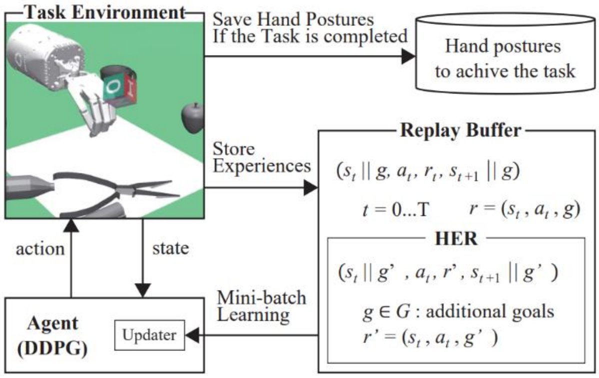

Fig. 1 Overview of the proposed method

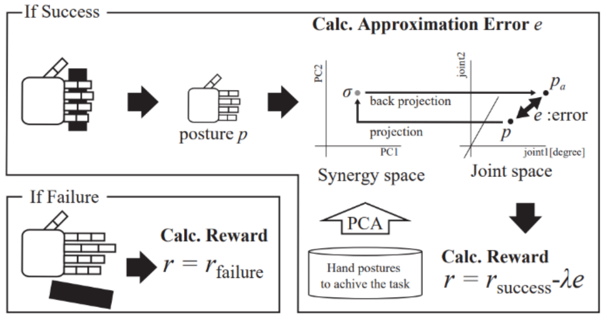

Fig. 2 Reward system

3. Experiments

3.1. *Task description*

A grasping task is trained on various objects to compare the results with synergies computed from the postures of humans grasping various objects. The tasks are executed on a physics simulator, Mujoco [8]. The robot hand running in the simulator is a kinematic model like the Shadow Dexterous Hand [9], with 20 actuators driving 24 joints (2 rotational wrist joints and 22 hand joints). In addition, one linear motion joint is added for the vertical motion of the hand, and the robot operated by this system includes 21 actuators and 25 joints. Actions are represented by vectors representing the actuator command values of these 21 joints.

The task is successful if the agent can grasp an object in a random position and orientation on a plane and move the object's center to the goal. Each episode consists of 100 steps, and each step executes a joint command value output from the DDPG and updates the state accordingly. The state is represented in 64 dimensions consisting of joint angles (25 Degrees of Freedoms: DoFs), joint angular velocities (25 DoFs), object position and posture (3 DoFs + 4 DoFs), object velocity and angular velocity (3 DoFs + 3 DoFs) and grasped object ID (1 DoF, seven objects).

As shown in Fig.1, postures that succeed in grasping objects are stored in the dataset. In this experiment, we use the joint angles (25 DoFs) as the type of posture to be preserved. Besides, five principal components will be used for the synergy space.

3.2. *Evaluation*

This experiment evaluates the task success rate during learning and the contribution of each principal component of the synergy. The contribution rate r_i is the ratio of the ith principal component to the total variance and is expressed as follows:

$$r_i = \frac{l_i}{\sum_{i=0}^{N} l_i}$$

l_i means the eigenvalue of the ith principal component. N is the number of the principal components of the synergy.

4. Results

Fig. 3 and Fig. 4 show the transition of the contribution rate (vertical bars) of each principal component of the learning synergy and the task success rate (black line) during the RL with $\lambda = 0, 1$, respectively. The contribution rate is calculated after the number of postures in the dataset exceeds 100. Until then, the penalty term λe in the reward function is zero, and learnings proceed similarly regardless of λ. In both settings, the contribution rate decreases significantly after adding the penalty term and then increases. In Fig. 3, the contribution rate decreases as the episodes pass, whereas in Fig. 4, the contribution rate increases. From these results, the feedback of the approximation error to the learned synergy space makes the synergy able to express grasping postures during the tasks.

An interesting fact is that considering synergy during RL improves both the contribution of synergy and the task success rate. Fig. 3 does not consider synergy, so the task success rate fluctuates around 70%. On the other hand, Fig.4 shows that the task success rate almost converges to 100%. This is because the grasping posture generated by the network follows the synergy, which includes a dataset of successful grasping postures. Thus, the task can be successful with a high probability.

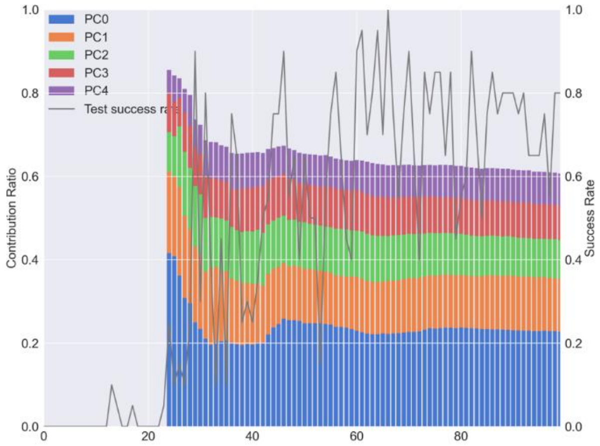

Fig. 3 Transition of contribution rate and task success rate during RL ($\lambda = 0$)

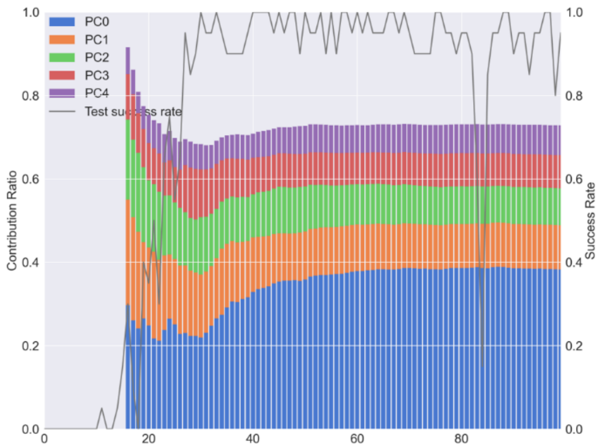

Fig. 4 Transition of contribution rate and task success rate during RL ($\lambda = 1$)

5. Conclusion

In this paper, we propose a platform for learning synergy among various robot models. We conduct experiments in which synergy is acquired in an object-grasping task using a humanoid hand. The results show that this RL platform realizes to make a synergy that can express various grasping postures and improves task success rate during RL.

This proposed system can learn not only grasping tasks but also more complex tasks such as tool manipulation that are appropriate for multi-fingered hands. In the future, we will compare the synergies obtained in this study with those obtained from humans, study the reward function for better synergy acquisition, train and assign synergy to various multi-fingered hand models, and verify the efficiency of RL when the acquired synergies are used as an action space.

References

1. Santello, Marco, Martha Flanders, and John F. Soechting. "Postural hand synergies for tool use." Journal of neuroscience 18.23 (1998): 10105-10115.
2. Ficuciello, Fanny. "Synergy-based control of underactuated anthropomorphic hands." IEEE Transactions on Industrial Informatics 15.2 (2018): 1144-1152.
3. Rodriguez, Diego, et al. "Learning Postural Synergies for Categorical Grasping through Shape Space Registration." 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids). IEEE, 2018.
4. Ficuciello, Fanny, et al. "A model-based strategy for mapping human grasps to robotic hands using synergies." 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. IEEE, 2013.
5. Silver, David, et al. "Deterministic policy gradient algorithms." International conference on machine learning. PMLR, 2014.
6. Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015).
7. Andrychowicz, Marcin, et al. "Hindsight experience replay." Advances in neural information processing systems 30 (2017).
8. Todorov, Emanuel, Tom Erez, and Yuval Tassa. "Mujoco: A physics engine for model-based control." 2012 IEEE/RSJ international conference on intelligent robots and systems. IEEE, 2012.
9. The Shadow Robot Company. https://github.com/shadow-robot

Authors Introduction

Mr. Kazuki Higashi

He received the M.E. degree from the Graduate School of Engineering Science, Osaka University, Japan in 2020. Currently, he is a Ph.D. candidate in Graduate School of Engineering Science, Osaka University, Japan. His research interests include low-dimensional control (Synergy) for multi-fingered hands and development of robotic hand utilizing synergy.

Mr. Tomohiro Motoda

He received the M.E. degree from the Graduate School of Engineering Science, Osaka University, Japan in 2020. Currently, he is a Ph.D. candidate in Graduate School of Engineering Science, Osaka University, Japan. His research interests deep learning in grasping and manipulation, motion planning, and manufacturing automation.

Mr. Akiyoshi Hara

He received the M.E. degree from the Graduate School of Information Science and Technology, Osaka University, Japan in 2020. Currently, he is a Ph.D. candidate in Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include transcranial current stimulation and reflex response, and tele-existence.

Dr. Kensuke Harada

He received the Ph.D. degree in engineering from the Graduate School of Mechanical Engineering, Kyoto University, Kyoto, Japan, in 1997. He is currently a Professor with the Graduate School of Engineering Science, Osaka University, Osaka, Japan. His current research interests include the mechanics and control of humanoid robots and robotic hands.