IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 September 2023, accepted 22 October 2023, date of publication 26 October 2023, date of current version 1 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327826

RESEARCH ARTICLE

Viewpoint Selection for the Efficient Teleoperation of a Robot Arm Using Reinforcement Learning

HAOXIANG LIU[1], REN KOMATSU[1], (Member, IEEE), SHINSUKE NAKASHIMA[1], (Member, IEEE), HIROYUKI HAMADA[1], NOBUTO MATSUHIRA[1], (Senior Member, IEEE), HAJIME ASAMA[1], (Fellow, IEEE), AND ATSUSHI YAMASHITA[2], (Senior Member, IEEE)

[1]Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan
[2]Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8563, Japan

Corresponding author: Ren Komatsu (komatsu@robot.t.u-tokyo.ac.jp)

This work was supported by the Japan Atomic Energy Agency (JAEA) Nuclear Energy Science and Technology (S&T) and Human Resource Development Project under Grant JPJA19H19210047.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Graduate School of Engineering, The University of Tokyo under Approval No. KE22-45.

ABSTRACT In this study, we developed a novel method to determine the optimal viewpoint from which an operator could realize faster and more accurate robot teleoperation using reinforcement learning. The reinforcement learning model was trained using images obtained from several candidate viewpoints from scratch, and the viewpoint at which the model achieved the highest rewards was considered the optimal viewpoint. The target robot, task, and environment were modeled using computer simulations and the candidate viewpoint images were obtained using those simulations. We employed the world model as our reinforcement learning model to maximize rewards in the reaching task of a robot arm. The reward function was designed to encourage the robot arm to reach the target position both quickly and accurately. The experimental results validated the choice of the world model as the reinforcement learning model. Moreover, subject experiments wherein subjects operated a robot arm remotely to reach the target position were conducted. The experiments produced results that strongly aligned with the performance obtained through computer simulations, indicating that the proposed method is capable of selecting the optimal viewpoint without handcrafted design and subject experiments.

INDEX TERMS Deep reinforcement learning, human interface, robot manipulation, teleoperation, viewpoint selection.

I. INTRODUCTION

Teleoperation, which plays a vital role in extreme environments, such as disaster-affected, nuclear, and underwater environments, is a widely used technology that allows operators to manipulate construction machinery and robots from a remote location. In particular, teleoperation is adopted in disaster response tasks, wherein in-vehicle operations are subject to danger from secondary disasters and extreme environmental conditions. For instance, when the Great East Japan Earthquake, representing one of the biggest disasters, occurred on March 11, 2011, several teleoperation robots were utilized because the affected places were dangerous to humans owing to radiation exposure. Nagatani et al. used a teleoperated robot called Quince to investigate the interiors of reactor buildings [1]. Additionally, numerous teleoperation robots were used in the Fukushima Daiichi Nuclear Power Station [2].

However, robot teleoperations are hindered by their comparatively low work efficiencies. Chayama et al. demonstrated that the efficiency of teleoperation is approximately 50% lower than that of in-vehicle operation during

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Liu.

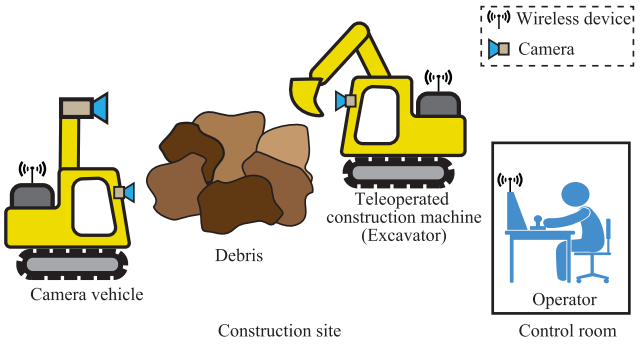


FIGURE 1. Example of an unmanned construction system. An operator remotely controls a teleoperated construction machine (excavator) from a safe operation room situated at a distance from the construction site. Alongside the excavator, a camera vehicle is present in the field to provide the operator with third-person views of the excavator.

excavation [3]. One reason for this is the difficulty in understanding the environment surrounding the robot. When an operator remotely controls a robot from a control room, the operator observes the monitor displaying images captured by the camera onsite, understands the situation based on these images, and then decides on commands to be sent to the robot. Therefore, the operator must understand the environment to efficiently control the robot.

One approach to obtain information on the surrounding environment of a robot involves viewing images captured by cameras installed on the robot, which provide first-person-viewpoint images. However, these images do not provide considerable depth information, unlike in-vehicle operations, where humans can perceive depth information owing to binocular vision. This limitation sometimes leads to an insufficient understanding of the surrounding environment by operators and can result in reduced work efficiencies.

To enhance efficiency, numerous studies have focused on determining the types of visual information beneficial for operators at remote sites. In unmanned construction systems, in addition to the target machines that are remotely controlled by operators, camera vehicles are employed to capture images from outside the construction field [3]. Fig. 1 presents an example of an unmanned construction system in which an operator controls an excavator in a safe operating room. As shown in Fig. 1, the camera vehicle offers the operator third-person views of the excavator, which helps the operator understand the situation of the excavator and perceive the depth.

Komatsu et al. [4] proposed a third-person-view generation system for the teleoperation of indoor robots using multiple fisheye cameras and two-dimensional distance sensors. In addition, Sugasawa et al. [5] improved the system for multiple robots, a dump truck and a hydraulic excavator, using fisheye cameras and a three-dimensional distance sensor.

In addition to providing third-person-view images, viewpoint selection is also crucial. Gualtieri et al. [6] demonstrated that choosing the correct viewpoint can significantly improve the average accuracy of a robot arm grasping task. Environmental cameras are often installed during unmanned construction activities. To reduce the blind spots caused by fixed viewpoints, Kamezaki et al. [7] proposed an autonomous camera-control system using six displays. However, multi-display systems increase the cognitive load on the operator and require skill and experience to determine the optimal viewpoint [8]. In addition, Yanco et al. suggested that displaying a large amount of sensor information to operators is not recommended as it tends to confuse the operators [9]. Therefore, selecting a single fixed optimal viewpoint is extremely crucial for efficient operation.

Chikushi et al. [10] proposed a method for automatically controlling the viewpoint based on the required specifications of construction machine operators involved in the construction of a dam. Their method was evaluated through subjective experiments. Sato et al. [11] proposed an allowable range of single viewpoints and single optimal viewpoints for the digging and releasing tasks of a hydraulic excavator by conducting subject experiments.

In this paper, we present a method for selecting a fixed single viewpoint that improves work efficiency, specifically enabling faster and more accurate teleoperation of a robot arm. In previous studies, the methods adopted for selecting viewpoints were manually crafted, meaning that they were designed by humans based on prior experience and task considerations; these methods were evaluated through subjective experiments. However, two major challenges still remain. First, in complex tasks, designing a suitable handcrafted viewpoint-selection method can be challenging. Second, previous methods require subject experiments to evaluate the handcrafted selection method. The optimal viewpoint may vary depending on factors such as the robot configuration, operational tasks, and environmental conditions. Therefore, the optimal viewpoint must be designed and evaluated through subjective experiments for every task.

The proposed method can automatically design an optimal viewpoint for the operator to realize faster and more accurate teleoperation, thereby overcoming the two challenges mentioned above by employing reinforcement learning. Here, we utilized the world model [12] as a reinforcement learning method that processes input images from a viewpoint using a variational autoencoder (VAE) [13], [14]. The world model was trained using images obtained from several candidate viewpoints from scratch, and the viewpoint that achieved the highest rewards was considered the optimal viewpoint. In this study, we obtained images from the candidate viewpoints using computer simulations, where the target robot, its tasks, and the environment were carefully modeled. Specifically, we focused on the reaching tasks of a robot arm as our target. The reward function was designed to encourage the robot arm to reach the target position both quickly and accurately. Moreover, we conducted subject experiments to illustrate the strong alignment between the results obtained from the world model and those obtained from the subjects.

This paper presents an extended version of our previous study [15], originally presented at a conference, which introduced the concept of utilizing reinforcement learning for viewpoint selection. In this extended work, we build upon our previous research by incorporating a comprehensive analysis of the validity of the chosen world model and introducing the results of a new human subject experiment.

The remainder of this paper is organized as follows. Section II revisits the fundamental concepts of reinforcement learning in viewpoint selection [15] and elaborates on the relationship between the human brain and world model. Section III explains the validity of the world model compared to other reinforcement learning methods. Section IV demonstrates that the results obtained by the world model in computer simulations agree well with the results obtained from the subject experiments. Finally, we discuss the results and present our conclusion in Sections V and VI, respectively.

II. METHODS

A. CONCEPT

The motivation for the development of the proposed method was to substitute the human response with a reinforcement learning model, as illustrated in Fig. 2. As shown in Fig. 2, during subject experiments, the human brain decides the next action based on visual signals and produces results. In the proposed method, we utilized a reinforcement learning model with the same responsibilities as the human brain, which receives a visual signal and produces the next action. In the proposed method, the reward function must be defined as explained later. If the response of the reinforcement learning model aligns well with the human response, the viewpoint obtained from reinforcement learning can also be considered optimal for the operator to efficiently control robots without additional evaluations through subject experiments.

Replacing the subject experiments with the reinforcement learning model also offers the advantage of training the model in computer simulations. This allows for the evaluation of numerous viewpoints without requiring the time-consuming process of gathering human subjects for experiments, provided adequate computer resources are available.

An overview of the proposed method is shown in Fig. 3. First, the target robot, task, and environment are modeled using computer simulations. Second, the candidates for the viewpoints are determined by considering the constraints on the poses of the viewpoints. Notably, viewpoint candidates can be placed randomly in cases with no constraints. Third, the reinforcement learning model is trained from scratch using images from the viewpoint of each candidate. Finally, the viewpoint that yields the highest reward is selected as the optimal viewpoint for teleoperation in the given situation.

B. REINFORCEMENT LEARNING MODEL ALIGNED WITH THE HUMAN BRAIN

To replace subject experiments with a reinforcement learning model, the reinforcement learning model must be chosen

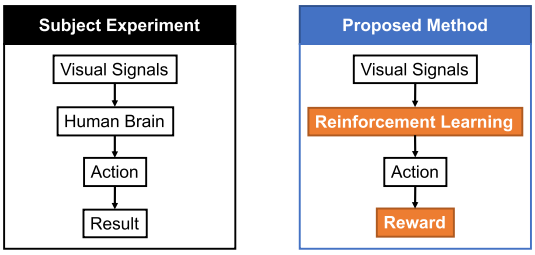


FIGURE 2. Comparison between the subject experiment and our proposed method.

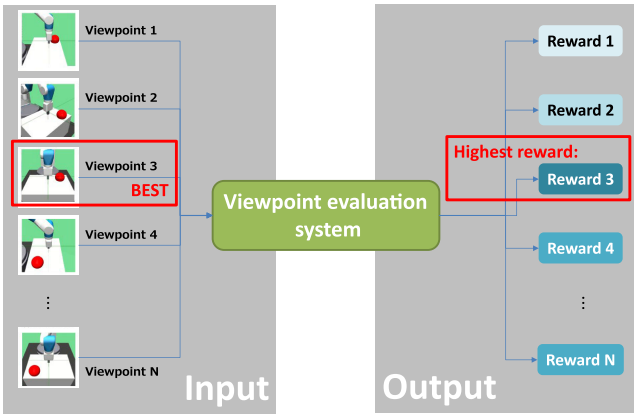


FIGURE 3. Overview of the proposed method. Numerous viewpoints are trained and evaluated using the reinforcement learning model, and the viewpoint that achieves the highest reward is selected as the best viewpoint for teleoperation in the situation.

carefully so that the response of the model aligns well with that of the human brain. In this study, we focus on the manifold hypothesis and motion extrapolation.

To process large amounts of information, the human brain learns to abstract spatial and temporal representations [16]. Humans can observe and describe scenes abstractly. For example, when making decisions while driving, the brain does not analyze every pixel of image information in the field of view. Instead, it transforms visual information into low-dimensional representations, which is called the manifold hypothesis.

In computational neuroscience, the manifold hypothesis posits that the distribution of real-world data can be perceived as low dimensional. Just as the human brain constantly converts high-dimensional information into abstract low-dimensional information, a VAE performs the same function.

Motion extrapolation refers to the dynamic object recognition mechanism among humans. Studies have shown that humans can use the previous trajectory of a moving object to predict its position [17]. To imitate the behavior of the motion extrapolation of humans, recurrent neural networks (RNNs) were utilized in this study. An RNN is a deep learning network structure that uses past information to improve the network performance for current and future inputs. An RNN is characterized by the inclusion of hidden states and loops in the network. By using a loop structure in the network,

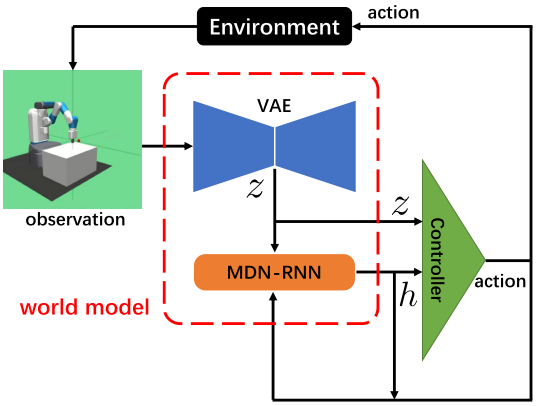


FIGURE 4. Schematic of the world model. Visual signals are processed through the VAE, and the processed low-dimensional latent space vectors are used in MDN-RNN to predict the state by incorporating past information. Finally, the controller decides the next action based on the latent vectors.

past information can be stored in hidden states and processed sequentially.

In this study, the world model proposed by Ha et al. [12] was selected as the reinforcement learning model because it utilizes a VAE to process the visual signals and an RNN to incorporate past information to improve state prediction, which is suitable for the proposed method.

C. WORLD MODEL

Here, the world model [12] is explained in detail. Fig. 4 shows a schematic of the world model. In the world model, the VAE [13], [14] is used to transform high-dimensional input observations into an abstract low-dimensional latent space. In this study, we take a three-channel 64 pixel $\times$ 64 pixel image as the input and convert it into a 32-dimensional latent vector.

Then, to predict the future based on the current information, the MDN-RNN [18], [19] is used. Using the RNN and mixture distribution networks (MDNs), the next latent vector is predicted from a mixture of Gaussian distributions.

Finally, the controller returns an action with the latent vector transformed by the VAE, representing the current state and the hidden state output by the RNN as the input. Owing to the absence of teaching data for the appropriate action in each scene, we need to train the controller using reinforcement learning. In reinforcement learning, the agent learns the actions of the robot arm through trial and error in an environment guided by the magnitude of the reward it receives. Specifically, we use the covariance matrix adaptation evolution strategy (CMA-ES) [20] to train the controller to maximize the reward.

D. DEFINITION OF REWARD

In this study, we focused on a reaching task in which the tip of the robot arm reached the target position as soon as possible. We defined x_{robot} as the tip of the robot arm and x_{target} as the target position. The reward function can be represented as

$$r_i = -\|x_{target} - x_{robot}\|, \tag{1}$$

where the negative Euclidean distance between the tip of the robot arm and the target in each frame was given as a reward.

Throughout all episodes, the tip of the robot arm x_{robot} started from the same position. Because an episode (one simulation) was set to n frames, the sum of the rewards in each frame r_{total} was defined as follows:

$$r_{total} = \sum_{i=1}^{n} r_i, \tag{2}$$

where n is set to 50 in this study. We designed the reward function for the following reasons:

1) The closer the robot arm to the target object, the greater the reward.
2) The faster the robot arm approaches the target, the greater the reward.

Therefore, a rapid and accurate approach is expected when learning yields the largest reward.

III. EXPERIMENTS ON THE VALIDITY OF THE WORLD MODEL

A. EXPERIMENTAL SETTING

In this section, we demonstrate the validity of the world model. In this study, the world model was selected as the reinforcement learning model. However, other reinforcement learning models could also be effective. Therefore, we trained the world model and another reinforcement learning model called deep Q-learning (DQN) [21] using images from different viewpoints to observe any differences in the rewards across the viewpoints. The DQN model is one of the most famous deep reinforcement learning algorithms, and unlike the world model, the DQN model is devoid of any VAE and RNN modules.

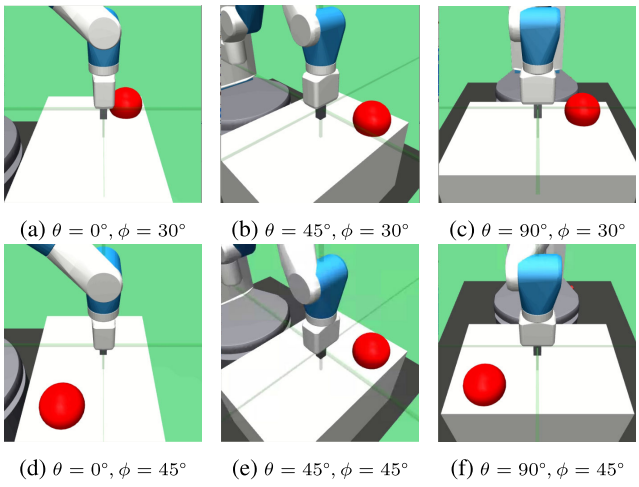
The simulation was implemented using OpenAI Gym [22], which is a commonly used reinforcement learning environment. Specifically, we customized `FetchReach` [23] environments to obtain images from arbitrary viewpoints as observations and to change the target objects. The algorithm was implemented using the PyTorch framework [24].

Six different viewpoints were selected as candidate viewpoints, as shown in Figs. 5(a)–(f). The polar coordinates of the body system are shown in Fig. 6. The viewpoint angles used in the simulations are presented in Table 1.

In the task, the tip of the robot arm was expected to reach the target position as soon as possible. The target position is indicated by the red spheres in Figs. 5(a)–(f). The model was trained four times for the world model and three times for the DQN model in the same environment. For the DQN model, the reward function is represented as

$$r_i^{DQN} = -2 \times \|x_{target} - x_{robot}\| + 1, \tag{3}$$

where the value of r_i is clipped between -1 and 1.



(a) $\theta = 0°, \phi = 30°$ (b) $\theta = 45°, \phi = 30°$ (c) $\theta = 90°, \phi = 30°$

(d) $\theta = 0°, \phi = 45°$ (e) $\theta = 45°, \phi = 45°$ (f) $\theta = 90°, \phi = 45°$

FIGURE 5. Viewpoints used for learning in Table 1.

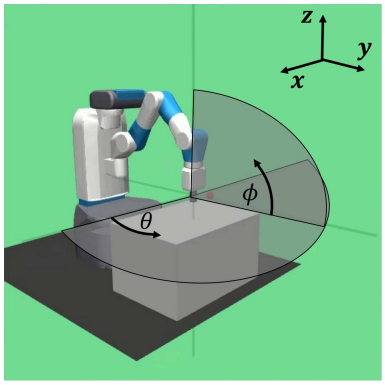


FIGURE 6. Polar coordinates of the robot arm. θ is the pan angle, and ϕ is the tilt angle.

TABLE 1. Candidate viewpoints.

	a	b	c	d	e	f
θ [°]	0	45	90	0	45	90
ϕ [°]	30	30	30	45	45	45

B. RESULTS

The results of the DQN and world models are presented in Figs. 7(a)–(c) and Figs. 8(a)–(c), respectively. The horizontal axis represents the epoch, and the vertical axis represents the total reward in the corresponding episode. The solid and dashed lines in Figs. 7(a)–(c) and 8(a)–(c) represent the mean values, and the semi-transparent region represents the 95% confidence interval.

As shown in Figs. 7(a)–(c) and 8(a)–(c), the rewards in DQN did not change across different viewpoints, whereas those in the world model increased for $\phi = 30°$ compared to those for $\phi = 45°$. Note that the reward was not significantly different at different θ values for either the DQN or world models. In other words, the DQN model is not suitable for selecting the best viewpoint, as viewpoint changes have a limited effect on the reward, whereas the world model is suitable, as the rewards differ depending on the viewpoint.

Furthermore, it is worth noting that the observation of higher rewards with a lower tilt angle (ϕ) aligns with the findings of a previous study conducted by Sato et al. [11]. Sato et al. conducted a subject experiment to investigate the optimal viewpoint of skilled operators and mentioned that a lower tilt angle ϕ in the release movement of a hydraulic excavator leads to higher operational efficiency because the operator can better gauge the vertical distance to the ground.

The difference between the DQN and world models may arise from the VAE module, where high-dimensional visual signals are transformed into low-dimensional representations. In the world model, the VAE is utilized so that small changes in visual information might disappear when the visual information is converted into a low-dimensional representation, which results in worse rewards when the world model uses a viewpoint where the robot movements do not cause significant visual changes. However, the world model performed better when the tilt angle was lower, and significant visual changes were observed when the robot arm moved vertically.

IV. EXPERIMENTS ON MODEL ALIGNMENT WITH SUBJECTS

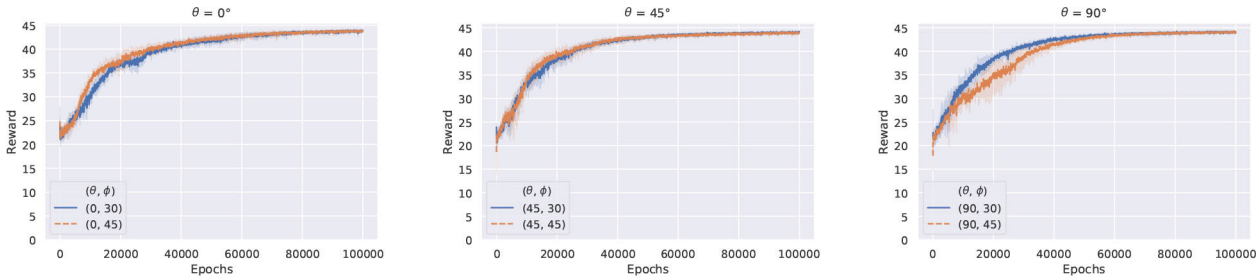
A. EXPERIMENTAL SETTING

In Section III, we demonstrated the efficiency of the proposed method and its alignment with the findings of a previous study [11]. However, the experimental and task conditions were not the same as those in the earlier study. In this section, we attempt to ensure that the experimental conditions and simulations are as similar as possible. Consequently, we aim to provide more evidence that our proposed method aligns well with the results obtained from experiments with participants.

We used a panda arm manufactured by Franka Emika for the experiments. The panda arm is a robot arm with seven degrees of freedom (DOF) that can be controlled easily using the Robot Operating System (ROS). As shown in Fig. 9, we implemented a simulation environment based on `PandaReach` in Panda-gym [25], which is built on top of the OpenAI Gym [22] environment with the panda arm. The simulation environment was customized to match an actual scene.

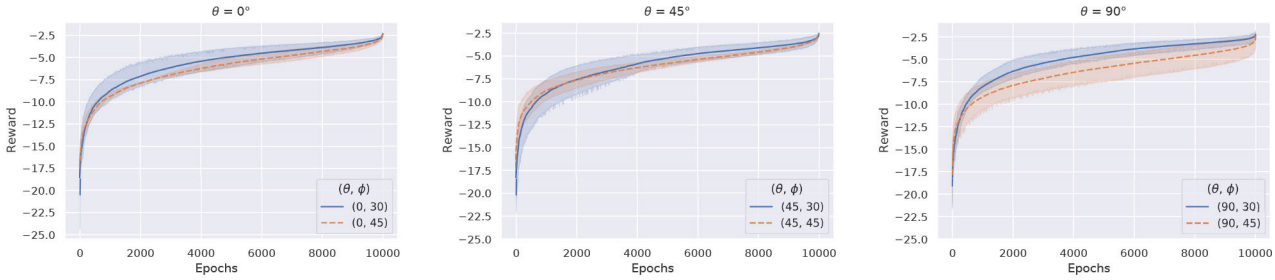
The experiments were conducted under three conditions.

- **1-DOF reaching task** without an obstacle. The tip of the robot arm only has 1-DOF motion available along the z-axis to reach the target. The candidates of the viewpoint are $(\theta, \phi) = (90°, 45°)$ and $(\theta, \phi) = (90°, 30°)$.
- **2-DOF reaching task** without an obstacle. The tip of the robot arm only has 2-DOF motion available along the y and z-axes to reach the target. The candidates of the viewpoint are $(\theta, \phi) = (0°, 0°)$ and $(\theta, \phi) = (90°, 0°)$.
- **2-DOF reaching task with an obstacle**. The task is the same as the **2-DOF reaching task** but with obstacles.



(a) $\theta = 0°$ (b) $\theta = 45°$ (c) $\theta = 90°$

FIGURE 7. Reward plots of DQN. The plots are separated based on the θ value. The horizontal axis represents the epoch, and the vertical axis represents the total reward in the corresponding episode.



(a) $\theta = 0°$ (b) $\theta = 45°$ (c) $\theta = 90°$

FIGURE 8. Reward plots of the world model. The plots are separated based on the θ value. The horizontal axis represents the epoch, and the vertical axis represents the total reward in the corresponding episode.

The coordinates of (θ, ϕ) are shown in Fig. 9. It should be noted that we set up simplified tasks with limited DOFs because complicated robot manipulation tasks can be decomposed into 1-DOF and 2-DOF robot arm movements. Furthermore, the efficiency of a complex task may vary depending on the chosen strategy, such as approaching the target from the right or left side. Selecting tasks with limited DOFs allowed us to evaluate the efficiency as influenced by different viewpoints.

The initial position of the tip of the robot arm was determined based on the target position, which allowed the robot to reach the target with limited DOFs. For instance, in the task with 1-DOF motion along the z-axis, the tip of the robot arm was placed immediately above the target so that only motion along the z-axis was required to reach the target.

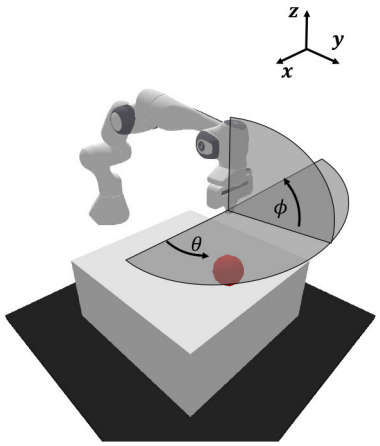
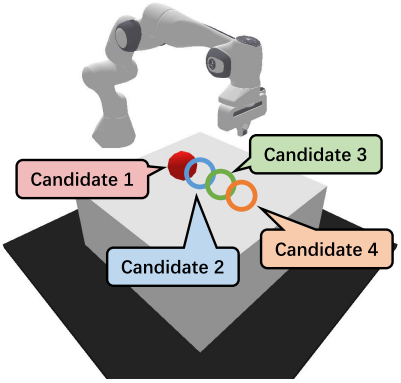


FIGURE 9. Computer simulation environment and coordinates of the robot arm. θ is the pan angle, and ϕ is the tilt angle. z is the vertical axis.

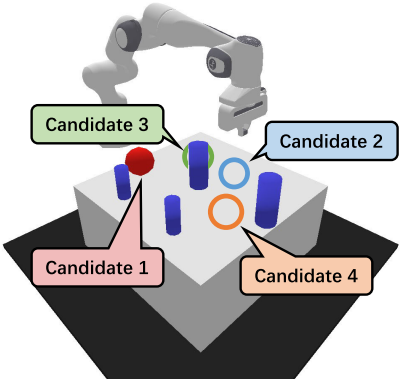
B. SIMULATION EXPERIMENTS

The target was randomly placed at four candidate positions, as shown in Fig. 10. The training of the VAE and MDN-RNN started with a learning rate of 0.001, and if the loss did not improve for five consecutive epochs, the learning rate was halved, and the training was continued. If the loss did not improve for more than 30 epochs, the model was considered to have stopped improving.

The results of the **1-DOF reaching task**, **2-DOF reaching task**, and **2-DOF reaching task with obstacles** are shown in Figs. 11(a)–(c), respectively. In each figure, the horizontal axis represents the epoch, and the vertical axis represents the total reward in the corresponding episode. The solid and dashed lines represent the mean values, and the semi-transparent region represents the 95% confidence interval. As shown in Fig. 11(a), for 1-DOF reaching task, the rewards are higher and converge faster for $\phi = 30°$ compared with those for $\phi = 45°$ and training, which aligns with the results in Section III. This suggests that even though the robot and environment are different, the proposed method extracts the viewpoint with a higher efficiency, and it can provide a better gauge for the vertical distance to the ground for the reaching



(a) 1-DOF and 2-DOF tasks without obstacles.



(b) 2-DOF task with obstacles.

FIGURE 10. **The candidate positions of target in simulation environment.**

task, which is further verified by subject experiments later in this section.

As shown in Figs. 11(b) and (c), regardless of the presence of obstacles, the rewards were higher for $\theta = 0^\circ$ compared with those for $\theta = 90^\circ$. This will be further evaluated using the results from the subject experiments later in this section.

C. SUBJECT EXPERIMENTS

Experiments were conducted with 11 subjects to validate the results of the simulation experiments. The experiments were conducted with the approval of the Research Ethics Committee of the Graduate School of Engineering, the University of Tokyo (approval no. KE22-45). Using statistical analysis, the effect of selecting viewpoints in teleoperation was demonstrated, and the effectiveness of the proposed method was proven.

To reproduce the teleoperation conditions as accurately as possible, the experimental room was separated into two parts using a whiteboard, as shown in Fig. 12. The left part in Fig. 12 shows the task area for placing the robot arm, camera, desk, and target, as shown in Fig. 13. The right part in Fig. 12 shows the teleoperation area in which the subjects manipulated the robot to complete the assigned task, as shown in Fig. 14. The participants were not allowed to view the task area directly before the experiments.

Logicool HD Pro Webcam c920r was used as the camera for live streaming the task area to the monitor in the teleoperation area. The participants controlled the robot arm using a Logicool F310r Gamepad, as shown in Fig. 15. The left joystick controls the movement of the tip of the robot arm in the horizontal plane, whereas the right joystick controls the movement of the tip of the robot arm in the vertical direction. The speed at which the robot arm moves can be adjusted by varying the pressure applied to the joystick. When the "start" button is pressed, the robot arm will return to the initial position automatically. When the "A" button is pressed, the gripper positioned at the tip of the robot arm can open or close.

The tasks for the subjects were the same as those in the simulation experiments described in Section IV-B, that is, the **1-DOF reaching task**, **2-DOF reaching task**, and **2-DOF reaching task with obstacles** were conducted. Moreover, the candidates for the viewpoints and target positions were the same as those in the simulation experiments.

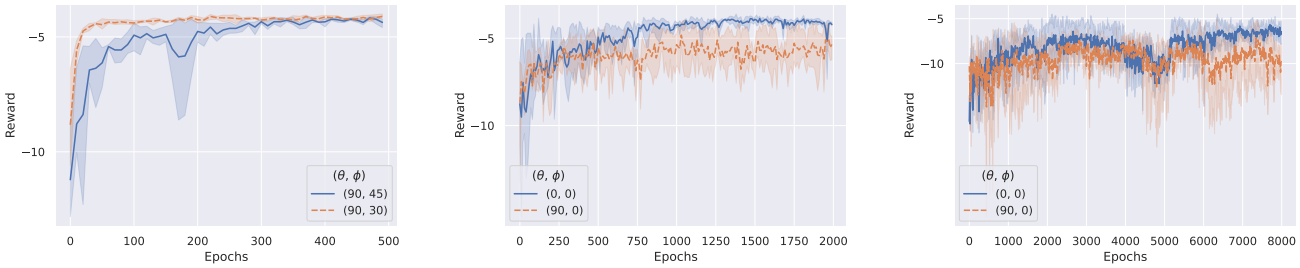
The operation time and distance errors were recorded during the experiments and two-sample t-tests were performed to compare the two viewpoints as follows:

- $(\theta, \phi) = (90^\circ, 30^\circ)$ and $(\theta, \phi) = (90^\circ, 45^\circ)$ in **1-DOF reaching task**.
- $(\theta, \phi) = (0^\circ, 0^\circ)$ and $(\theta, \phi) = (90^\circ, 0^\circ)$ in **2-DOF reaching task**.
- $(\theta, \phi) = (0^\circ, 0^\circ)$ and $(\theta, \phi) = (90^\circ, 0^\circ)$ in **2-DOF reaching task with obstacles**.

To accurately measure the operation time and distance error, the subjects were instructed to reach and grasp the red target with as much precision as possible, as shown in Fig. 16. Notably, when the "A" button on the controller is pressed to grasp the target, an operation is considered finished regardless of whether the grasp succeeds or not. The operation time and Euclidean distance between the tip of the robot arm and the target are then recorded. For each task, the target is placed at different positions four times, and the average operation time and distance error are calculated for the corresponding task.

Examples of the viewpoints from the **2-DOF reaching task** and **2-DOF reaching task with obstacles** are shown in Figs. 17 and 18, respectively. As shown in Fig. 18, even with obstacles, we carefully selected the target position such that the target was not completely occluded by the obstacles.

The participants received instructions before the experiments. First, the subjects were asked to sit in front of a desk, and the research objectives were explained. Second, the experimental description was presented, and the participants were asked to sign a consent form that explicitly informed them that they could withdraw from participating in the experiment. Afterward, the subjects sat in front of a monitor that displayed live streaming of the task area, including the robot arm, desk, target, and obstacles. The participants were provided with a controller manual and instructions on how to operate the robot arm. The subjects were given a controller for 5–15 min to familiarize themselves with the task and the operation of the robot arm while watching the



(a) 1-DOF reaching task. (b) 2-DOF reaching task. (c) 2-DOF reaching task with obstacles.

FIGURE 11. The results of reaching tasks in simulations. The horizontal axis represents the epoch, and the vertical axis represents the total reward in the corresponding episode.

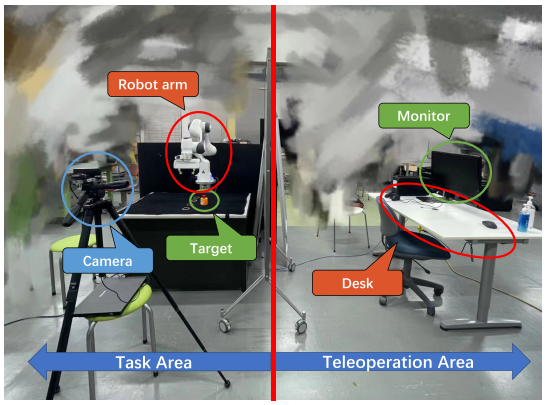


FIGURE 12. Overview of the experimental environment. The left and right sides are the task area where the robot arm is placed and the teleoperation area where the operator teleoperates the robot, respectively.

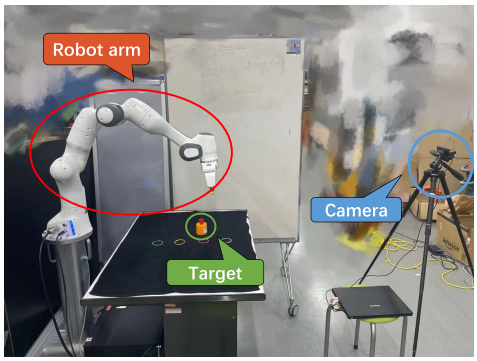


FIGURE 13. Task area from a side view. The teleoperation area is behind the whiteboard.

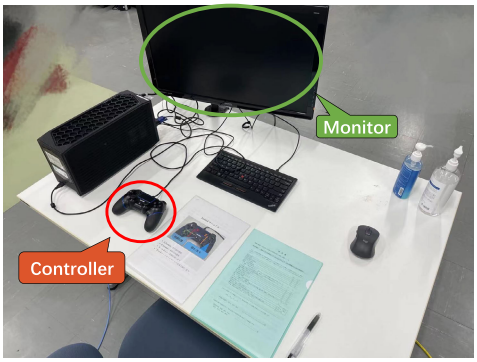


FIGURE 14. Teleoperation area.

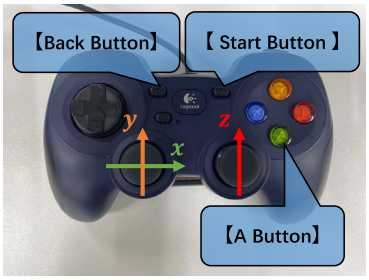


FIGURE 15. Controller.

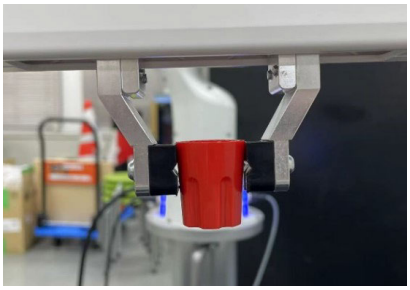


FIGURE 16. Goal of the task.

monitor. The experiment began after the exercise. The order of the viewpoints assigned to the subjects was randomized to minimize the impact of proficiency improvement during the experiments.

D. SUBJECT EXPERIMENT RESULTS

The statistical analysis of the subject experiments using two-sample *t*-tests is shown in Table 2, and the plots are depicted in Figs. 19, 20, and 21. We set a significance level of $\alpha = 5\%$ for all the statistical tests. As shown in Table 2 and Figs. 19(a) and (b), in the **1-DOF reaching task**, an analysis of the distance error between the viewpoints $(\theta, \phi) = (90°, 30°)$ and $(\theta, \phi) = (90°, 45°)$

TABLE 2. Averaged results and *p*-value of the two-sample *t*-test.

		1-DOF		2-DOF		2-DOF with obstacles	
Viewpoint	θ [°]	90	90	0	90	0	90
	ϕ [°]	30	45	0	0	0	0
Task time (s)		**10.66**	12.11	**17.29**	25.94	**17.57**	29.17
Distance error (cm)		**0.94**	1.56	**1.74**	6.23	**1.22**	5.94
p-value of task time		0.149		0.044		0.021	
p-value of distance error		0.034		0.00036		0.00078	

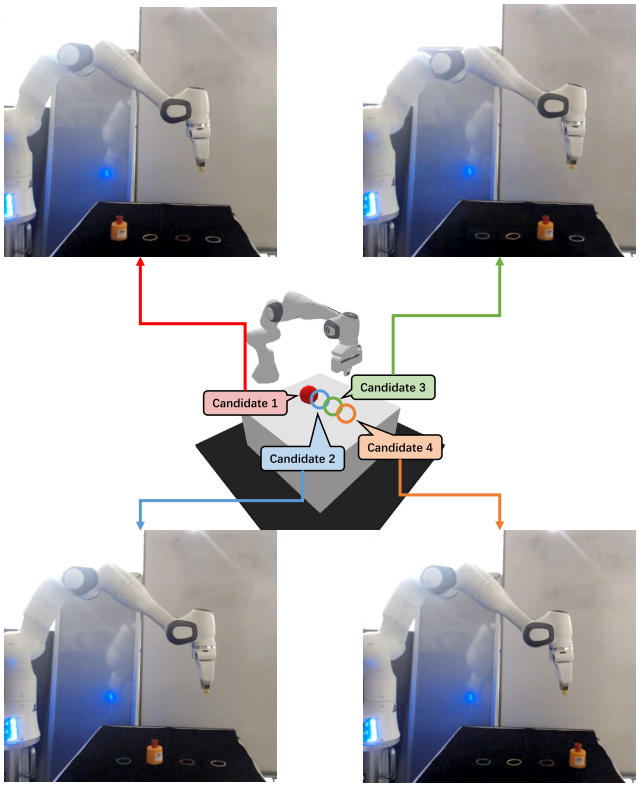


FIGURE 17. Example of images captured from viewpoint $(\theta, \phi) = (0°, 0°)$ for the **2-DOF reaching task**. Four trials consist of four different target positions.

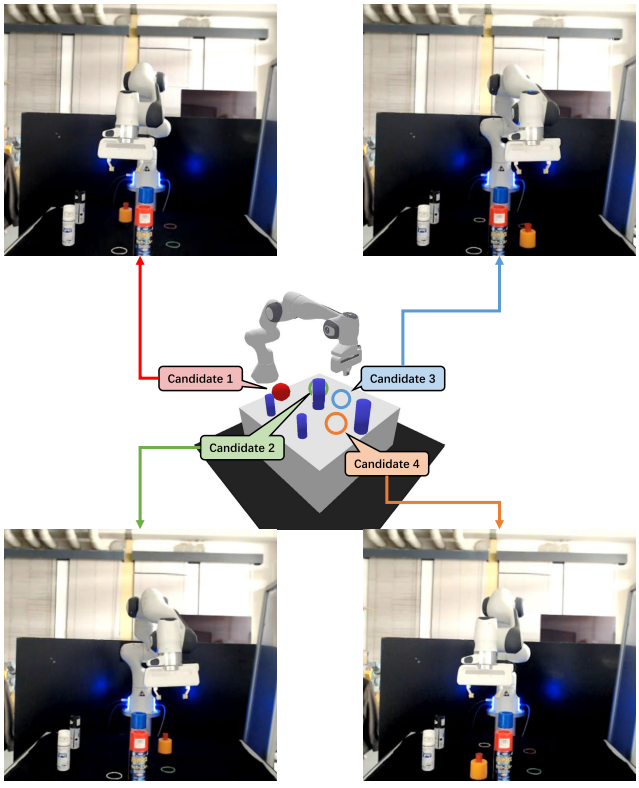


FIGURE 18. Example of images captured from viewpoint $(\theta, \phi) = (90°, 0°)$ for the **2-DOF reaching task with obstacles**. Four trials consist of four different target positions.

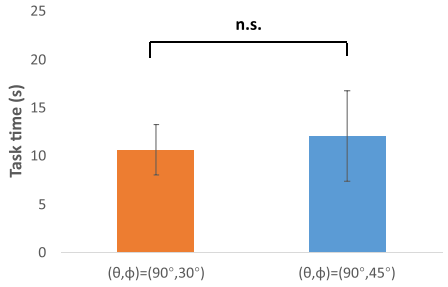
revealed a significant difference between the two viewpoints ($p = 0.034$). However, the task completion time did not exhibit a statistically significant difference ($p = 0.149$) under the same experimental conditions. In terms of the distance error, the viewpoint of $(\theta, \phi) = (90°, 30°)$ is better than the viewpoint of $(\theta, \phi) = (90°, 45°)$, which is in good agreement with the results obtained from the simulation experiment described in Section IV-B.

As shown in Table 2 and Figs. 20 and 21, the task completion time and distance error are significantly different in both the **2-DOF reaching task** (p=0.044 and p=0.00036, respectively) and **2-DOF reaching task with obstacles** (p=0.021 and p=0.00078, respectively). The results show that a viewpoint with $(\theta, \phi) = (0°, 0°)$ produces a shorter task completion time and fewer task errors than a viewpoint with $(\theta, \phi) = (90°, 45°)$. This result suggests that the viewpoint with $(\theta, \phi) = (0°, 0°)$ is better than the others, regardless of the presence of obstacles, which is well aligned with the results obtained from the simulation experiment detailed in Section IV-B.

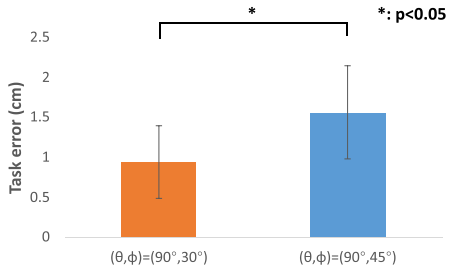
In summary, the rewards obtained by the world model in the simulation aligned well with the speed and accuracy of the subject experiments. This suggests that the viewpoints that provide the world model with higher rewards are likely viewpoints where a human operator can achieve faster and more accurate teleoperation. However, the task completion time for the **1-DOF reaching task** did not show a significant difference, as discussed in Section V.

V. DISCUSSION

The experimental results suggested a strong alignment between the rewards obtained by the world model in the computer simulation and the results of human teleoperation using images captured from various viewpoints. However, there was no significant difference in the task completion time for the **1-DOF reaching task**.

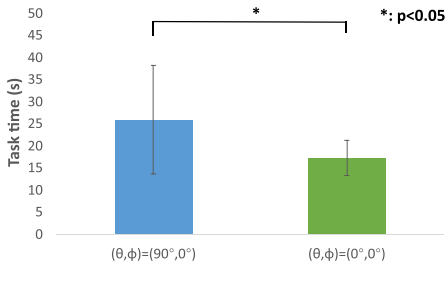


(a) Time.

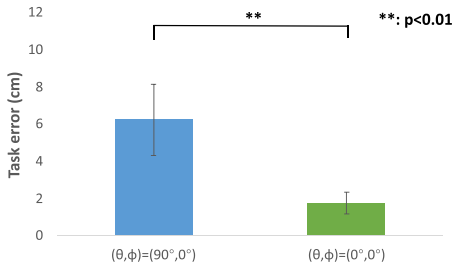


(b) Error.

FIGURE 19. Statistical result on the 1-DOF reaching task with viewpoints $(\theta, \phi) = (90°, 30°)$ and $(\theta, \phi) = (90°, 45°)$.



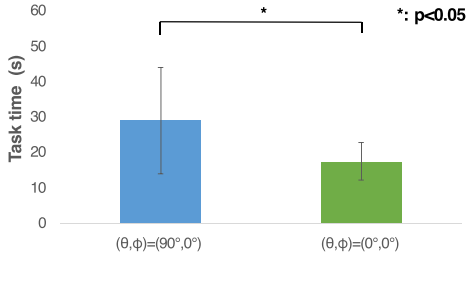
(a) Time.



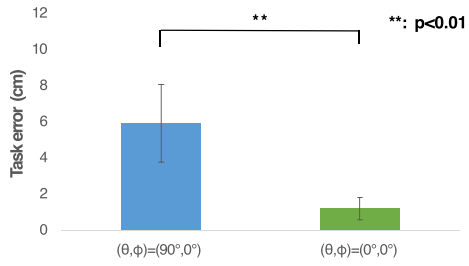
(b) Error.

FIGURE 20. Statistical result on the 2-DOF reaching task with viewpoints $(\theta, \phi) = (0°, 0°)$ and $(\theta, \phi) = (90°, 0°)$.

One hypothesis is that the **1-DOF reaching task** is overly simple. In this task, the subjects were required to control only the 1-DOF movement of the robot arm. This means that once the subjects apply pressure to the joystick in a specific direction, the subjects are only required to release the pressure on the joystick when the tip of the robot arm reaches the target and press the "A" button on the controller to close the gripper.



(a) Time.



(b) Error.

FIGURE 21. Statistical result on the 2-DOF reaching task with obstacles with viewpoints $(\theta, \phi) = (0°, 0°)$ and $(\theta, \phi) = (90°, 0°)$.

When the gripper is closed, the task is considered complete regardless of whether the target is grabbed and the time is recorded. Hence, there is not much room for variation in the task completion time across different viewpoints.

Although the task completion time may not exhibit significant differences across various viewpoints, the distance error metric quantifies how closely the robot arm reaches the target, which is crucial for assessing the performance of a teleoperation system. Therefore, the absence of significant differences in the task completion time for the **1-DOF reaching task** would not affect the validity of the proposed method.

VI. CONCLUSION

This paper proposes a novel method to find the optimal viewpoint from which an operator can realize faster and more accurate robot teleoperation using a reinforcement learning model, namely the world model. The world model was trained from scratch using images obtained from several candidate viewpoints, and the viewpoint where the model achieved the highest rewards was considered the optimal viewpoint. The target robot, task, and environment were modeled using computer simulations and the candidate viewpoint images were obtained using those simulations. The experimental results validated the selection of the world model as the reinforcement learning model and demonstrated well-aligned results between the rewards obtained by the world model in computer simulations and the task performance, namely the accuracy and speed, obtained from the subject experiments. This suggests that the proposed method can determine the optimal viewpoint without requiring manual design or

subjective experiments, as long as a computer simulation is provided.

In this study, simplified tasks with limited DOFs were selected to evaluate the proposed method because complicated tasks can be decomposed into only limited DOF movements. Because the proposed method can generate a viewpoint to improve the work efficiency for limited DOF movements, it is likely that the viewpoint provided by the proposed method can improve the work performance, even for complicated tasks. However, it is also possible that the effectiveness of our method is confined to specific scenarios, such as reaching tasks. The generalizability of the proposed method can only be demonstrated by applying it to a wide range of teleoperation tasks. In future studies, the proposed method will be applied to more complex tasks to validate the effectiveness of the proposed method.

REFERENCES

[1] K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, and Y. Hada, "Redesign of rescue mobile robot Quince," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, Nov. 2011, pp. 13–18.

[2] S. Kawatsuma, M. Fukushima, and T. Okada, "Emergency response by robots to Fukushima-Daiichi accident: Summary and lessons learned," *Ind. Robot: Int. J.*, vol. 39, no. 5, pp. 428–435, Aug. 2012.

[3] K. Chayama, A. Fujioka, K. Kawashima, H. Yamamoto, Y. Nitta, C. Ueki, A. Yamashita, and H. Asama, "Technology of unmanned construction system in Japan," *J. Robot. Mechtron.*, vol. 26, no. 4, pp. 403–417, 2014.

[4] R. Komatsu, H. Fujii, Y. Tamura, A. Yamashita, and H. Asama, "Free viewpoint image generation system using fisheye cameras and a laser rangefinder for indoor robot teleoperation," *ROBOMECH J.*, vol. 7, no. 1, pp. 1–10, Dec. 2020.

[5] Y. Sugasawa, S. Chikushi, R. Komatsu, J. Younes L. Kasahara, S. Pathak, R. Yajima, S. Hamasaki, K. Nagatani, T. Chiba, and K. Chayama, "Visualization of dump truck and excavator in bird's-eye view by fisheye cameras and 3D range sensor," in *Proc. Int. Conf. Intell. Auton. Syst.* Cham, Switzerland: Springer, 2021, pp. 629–640.

[6] M. Gualtieri and R. Platt, "Viewpoint selection for grasp detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 258–264.

[7] M. Kamezaki, J. Yang, H. Iwata, and S. Sugano, "An autonomous multi-camera control system using situation-based role assignment for tele-operated work machines," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 5971–5976.

[8] D. E. Crundall and G. Underwood, "Effects of experience and processing demands on visual information acquisition in drivers," *Ergonomics*, vol. 41, no. 4, pp. 448–458, Apr. 1998.

[9] H. Yanco, J. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Hum.-Comput. Interact.*, vol. 19, no. 1, pp. 117–149, Jun. 2004.

[10] S. Chikushi, Y. Moriyama, H. Fujii, Y. Tamura, H. Yamakawa, K. Nagatani, Y. Sakai, T. Chiba, S. Yamamoto, K. Chayama, A. Yamashita, and H. Asama, "Automated image presentation for backhoe embankment construction in unmanned construction site," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2020, pp. 22–27.

[11] R. Sato, M. Kamezaki, M. Yamada, T. Hashimoto, S. Sugano, and H. Iwata, "Environmental camera placements for skilled operators in unmanned construction," *Autom. Construct.*, vol. 119, Nov. 2020, Art. no. 103294.

[12] D. Ha and J. Schmidhuber, "Recurrent world models facilitate policy evolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1–13.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[14] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1278–1286.

[15] H. Liu, R. Komatsu, H. Woo, Y. Tamura, A. Yamashita, and H. Asama, "Viewpoint selection without subject experiments for teleoperation of robot arm in reaching task using reinforcement learning," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2022, pp. 1015–1020.

[16] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, no. 6, pp. 1013–1028, Jun. 2017.

[17] H. Hogendoorn, "Motion extrapolation in visual processing: Lessons from 25 years of flash-lag debate," *J. Neurosci.*, vol. 40, no. 30, pp. 5698–5705, Jul. 2020.

[18] C. M. Bishop, "Mixture density networks," Tech. Rep., 1994. [Online]. Available: https://publications.aston.ac.uk/id/eprint/373/

[19] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*.

[20] N. Hansen, "The CMA evolution strategy: A tutorial," 2016, *arXiv:1604.00772*.

[21] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.

[22] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," 2016, *arXiv:1606.01540*.

[23] M. Plappert, M. Andrychowicz, A. Ray, B. McGrew, B. Baker, G. Powell, J. Schneider, J. Tobin, M. Chociej, P. Welinder, V. Kumar, and W. Zaremba, "Multi-goal reinforcement learning: Challenging robotics environments and request for research," 2018, *arXiv:1802.09464*.

[24] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–12.

[25] Q. Gallouédec, N. Cazin, E. Dellandréa, and L. Chen, "Panda-gym: Open-source goal-conditioned environments for robotic learning," 2021, *arXiv:2106.13687*.

HAOXIANG LIU received the B.E. degree from the Division of Mechanical, Materials and Manufacturing Science, Osaka University, Japan, in 2021, and the M.E. degree from the Department of Precision Engineering, The University of Tokyo, Japan, in 2023. His research interests include computer vision, deep learning, and robot teleoperation.

REN KOMATSU (Member, IEEE) received the B.S. degree in engineering from Yokohama National University, Japan, in 2014, and the M.S. and Ph.D. degrees in engineering from The University of Tokyo, Japan, in 2016 and 2020, respectively. From 2017 to 2018, he was a Visiting Scholar with the Robotics Institute, Carnegie Mellon University, USA. From 2020 to 2022, he was a Project Assistant Professor with The University of Tokyo, where he has been an Assistant Professor, since 2022. His research interests include computer vision, visual SLAM, deep learning, and robot teleoperation.

SHINSUKE NAKASHIMA (Member, IEEE) received the B.S. degree in engineering from Yokohama National University, Japan, in 2013, and the M.S. and Ph.D. degrees in information science and technology from The University of Tokyo, Japan, in 2015 and 2021, respectively. From 2021 to 2022, he was a Project Researcher with The University of Tokyo. His research interests include bio-inspired robots, self-healing robots, and robot mechanism.

HIROYUKI HAMADA received the M.M.Sc. and Ph.D. degrees from Chiba University, in 2012 and 2016, respectively. From 2008 to 2015, he was a Physical Therapist with the Ichikawa-City Rehabilitation Hospital. From 2017 to 2018, he was a Trainee with the Study Center for Cognitive Multisensory Rehabilitation. From 2018 to 2020, he was an Assistant with Bunkyo Gakuin University. Since 2020, he has been a Project Assistant Professor with the Department of Precision Engineering, The University of Tokyo. His research interests include rehabilitation and neuroscience. He is a member of JPTA, WCPT, SfN, and JNS.

NOBUTO MATSUHIRA (Senior Member, IEEE) received the B.E., M.E., and D.E. degrees from the Tokyo Institute of Technology, in 1978, 1982, and 1997, respectively. From 1982 to 2011, he was with the Toshiba Corporate Research and Development Center. From 2011 to 2023, he was a Professor with the Shibaura Institute of Technology. Since 2022, he has been a Project Researcher with The University of Tokyo and the Shibaura Institute of Technology. His research interests include teleoperation, robot networks, human–robot interaction, and community service robot. He is a fellow of JSME, RSJ, and SICE.

HAJIME ASAMA (Fellow, IEEE) received the M.S. and Dr.Eng. degrees from The University of Tokyo (UTokyo), in 1984 and 1989, respectively. He was with RIKEN, Japan, from 1986 to 2002; a Professor with the Research Into Artifacts, Center for Engineering (RACE), UTokyo, in 2002, and with the School of Engineering, since 2009; and the Director of RACE, since 2019. His main research interests include service robotics, distributed autonomous robotic systems, embodied brain science systems, and cognitive ergonomics. He is a fellow of JSME and RSJ. He received the SICE System Integration Division System Integration Award for Academic Achievement, in 2010, and the JSME Award (Technical Achievement), in 2018. He was the Vice President of RSJ, from 2011 to 2012, and an Ad Com Member of the IEEE Robotics and Automation Society, from 2007 to 2009. He has been the President of IFAC, since 2020. He has been a Council Member of the Science Council of Japan, since 2017.

ATSUSHI YAMASHITA (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Faculty of Engineering, The University of Tokyo, in 1996, 1998, and 2001, respectively. From 2001 to 2008, he was an Assistant Professor with Shizuoka University. From 2006 to 2007, he was a Visiting Associate with the California Institute of Technology. From 2008 to 2011, he was an Associate Professor with Shizuoka University. From 2011 to 2022, he was an Associate Professor with The University of Tokyo, where he has been a Professor, since 2022. His research interests include robot vision, image processing, and intelligent robot systems. He is a member of ACM, JSPE, RSJ, IEICE, JSAE, JSCE, JSME, IEEJ, IPSJ, ITE, SICE, and the Society for Serviceology.