

# Development of interface for teleoperation of humanoid robot using task model method

Masaya Ogawa<sup>1</sup>, Katsuya Honda<sup>2</sup>, Yoshihiro Sato<sup>3</sup>, Takeshi Oishi<sup>3</sup> and Katsushi Ikeuchi<sup>4</sup>

**Abstract**—Teleoperation systems for humanoid robots that replace human tasks in emergency situations have been widely studied. Humanoid robots have the capability of utilizing the equipment designed for human manipulation. When the humanoid robot works on behalf of human beings, rapid deployment can be achieved when a disaster occurs. However, since the humanoid robot and the human have different physical abilities, it is difficult to control the robot by mapping the human motion directly to the robot.

Recently, the interfaces that map the human motion, taken by the vision sensors, to the humanoid robot have become popular. Though the vision-based teleoperation system can be rapidly constructed anywhere, the interfaces require some sort of feedback to smoothly control the robot. The tactile and force feedback system has limitations on human motion because of mechanical structures. The teleoperation system also has the potential issue of time delay.

In this paper, we propose an interface for teleoperation of the humanoid robot with vision sensors. The inner states of the robot and the recognized human motions are visualized for the feedback to the operator. We also introduce the task model to avoid the time delay. In the visual feedback, the task states are visualized to assist the smooth operations.

## I. INTRODUCTION

In recent years, the research of a humanoid robot that replaces human tasks in emergency situations has been widely studied [1][2]. At a time when a disaster occurs, the rapid deployment of robots is required. Developing new robots, designed for these tasks and adjusted to the environments, wastes much time. On the other hand, the humanoid robot has the capability to utilize the same equipment that is designed for human operation. It would be extremely valuable if the humanoid robot could work in such emergency situations on behalf of the human (Fig. 1).

We need an intuitive interface to effectively control the humanoid robot from a distant place. Telexistence systems for immersive control provide a highly realistic operating environment by using Virtual Reality (VR) techniques [3][4]. But, the system is generally huge and the human motion is limited by mechanical structures. The interfaces that capture the human motion by using vision sensors have been recently studied [5][6]. The depth-sensing camera has become widely available and can be used for capturing the human motion

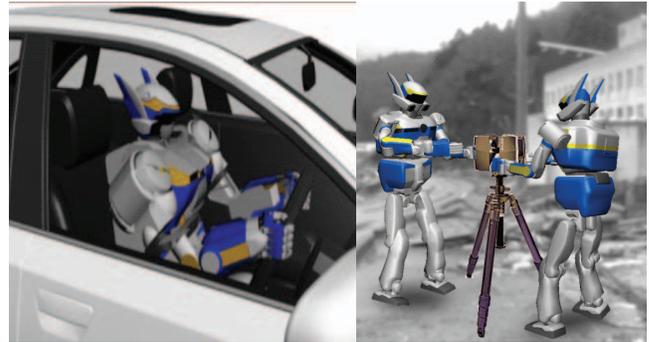


Fig. 1. Teleoperation of humanoid robot in emergency situations

without mechanical sensing. A real-time human imitation system for biped robots has been developed. However, since the humanoid robot and the human have a different joint structure, physical ability, and weight balance, it is difficult to map the human motion directly to the robot.

Teleoperation has the potential issue of time delay between the operator and the robot. The operator cannot recognize the state of the robot and the environment at the same moment. Human motion is not fluently reflected in the robot motion. That is, the intuitive operation of the robot is difficult by such interfaces. To avoid the influence of time delay, it is desirable that the operator provides global directions and the robot run semi-autonomously. In addition to time delay, lack of operability occurs because there is not enough information to check the state of the work object and the robot itself. It is difficult to do the work by the visual information obtained from the robot, and it puts a large burden on the operator.

In this paper, we present an interface for a teleoperation system to facilitate teleoperation by visual feedback of operating information. The recognition and transmission of human motions are performed by the task model [7]. The task model can simultaneously solve the issues of the structural differences between the human beings and the humanoid robot, and time delays. The task is recognized with the vision systems by observing the human operation. In addition, information of task recognition and motion generation is presented along with the visual information as the visualization of task model. The operator can recognize their own behavior and the corresponding behavior of the robot by this operation auxiliary interface.

Considering operations in disaster areas, humanoid robots are not good at moving over uneven terrain with many obstacles.

<sup>1</sup>Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Japan [ogawa@cvl.iis.u-tokyo.ac.jp](mailto:ogawa@cvl.iis.u-tokyo.ac.jp)

<sup>2</sup>Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, Japan [honda.k@cvl.iis.u-tokyo.ac.jp](mailto:honda.k@cvl.iis.u-tokyo.ac.jp)

<sup>3</sup>Institute of Industrial Science, The University of Tokyo, Japan [{yoshi,oishi}@cvl.iis.u-tokyo.ac.jp](mailto:{yoshi,oishi}@cvl.iis.u-tokyo.ac.jp)

<sup>4</sup>Interfaculty Initiative in Information Studies, The University of Tokyo, Japan [ki@cvl.iis.u-tokyo.ac.jp](mailto:ki@cvl.iis.u-tokyo.ac.jp)

Here, we target the operation to driving a four-wheel buggy for moving over rough terrain. The buggy is controlled by a steering wheel, acceleration and brake levers.

## II. RELATED WORK

Recently, interfaces that map the human motion to the humanoid robot have become popular. The interfaces require some sort of feedback to smoothly control the robot because of the lack of environmental information.

To compensate for the lack of environmental information in the teleoperation, an interface that presents force-tactile information is proposed. Tachi et al. are using the system to reproduce gripping force by tension of wire, and given operability [3]. Song et al. are getting tactile feedback by a mobile teleoperation system which includes a tactile glove [8]. An interface which uses tactile feedback gives the operational feeling of direct manipulation, but the system is increased in size by tactile presentation devices and has limitations on the human motion because of mechanical structures.

On the other hand, interfaces which improve operability by visual feedback are also proposed. Otsuki et al. introduce physical model of operation for the object behavior. Otsuki improves operability by the presentation of physical state visually and given auditory feedback depending on the operation [9]. Song et al. have proposed the operation by the rod-like operation metaphor [10]. By operation through the rod-like operation metaphor, Song et al. reduce the difficulty of the 3D virtual object grasping, by solving the accuracy problem of recognition hand rotation.

The operational interface to enhance the understanding of the behavior of the robot by the visualization of AR-MR techniques have even been suggested [11][12]. Kobayashi et al. presented in overlays a process and conditions for generating the autonomous behavior of humanoid robots using the mixed reality to the real space [11]. The humanoid robot system is so complex, and there is a problem that the processes of motion generation and recognition are difficult to check in the operation verification by real humanoid robots in real space. Kobayashi developed a system which is useful to understand intuitively how each module in the humanoid affects the total system and how one is affected by the humanoid's surrounding environment.

The teleoperation system also has the potential issue of time delay. A teleoperation system considering time delay using VR environment has been proposed [13][14][15]. The method separates the entire teleoperation feedback loop into operator-VR feedback loop and VR-robot feedback loop. The system is stably controlled by satisfying the loops independently. Thus, the system does not lose control because of time delay.

We also introduce the task model to avoid the issues of the physical differences and time delay. The task model is known as a technique for generating robot motions from human motions while eliminating the physical differences between the human and robot [7][16][17]. The task model describes a human motion by tasks and skill parameters: "what to do"

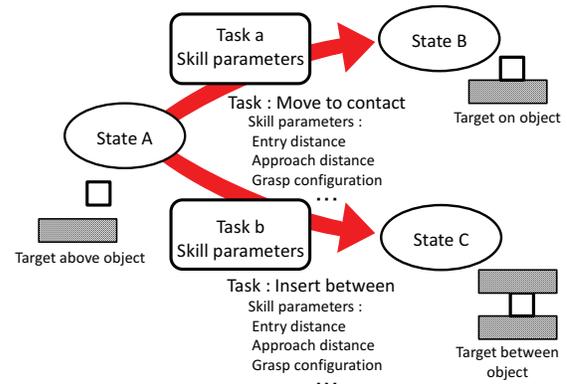


Fig. 2. Definition of task model [7]

and "How to do". Fig. 2 shows the overview of the task model. To recognize the human motions by the task model, the states of the target and its transitions must be extracted. The operation of a state transition is called task. To execute the task, the parameters for moving the robot are required. Example parameters include the inner state of the robot and the state of the target object. These parameters are called skill parameters.

APO (Assembly-Plan-from-Observation) paradigm is designed to produce robot motion for assembly works from the observation of human operation [7]. The states of the target objects are classified and described by the state of face contact. A motion template for the robot arm is prepared for each state transition. The proposed system successfully assembled the polyhedral objects by using the task model.

Nagata et al. proposed shared autonomy system by task instruction in the work space [16]. The Object Template Model and Task Space Model are defined for describing the work space model. In this system, the user recognizes the object type and its state in the environment. The user instructs a robot to execute the task of picking an object by selecting a task model according to the object type and situation.

We propose a vision-based interface for teleoperation of the humanoid robot by using task model. The inner states of the robot and the recognized human motions are visualized for the feedback. The task model is used for recognizing the human motion and the generation of the robot motion while avoiding the issues of the physical differences and time delay. In addition, internal information of task recognition and motion generation is presented along with the visual information as the visualization of task model. We execute teleoperation of humanoid robot smoothly by the visualization of task model.

## III. TELEOPERATION BY TASK MODEL

The operational interface for teleoperation system by the task model consists of the following four procedures (Fig. 3).

- 1) The environmental views taken by the vision systems on the robot are transmitted to the operator through a

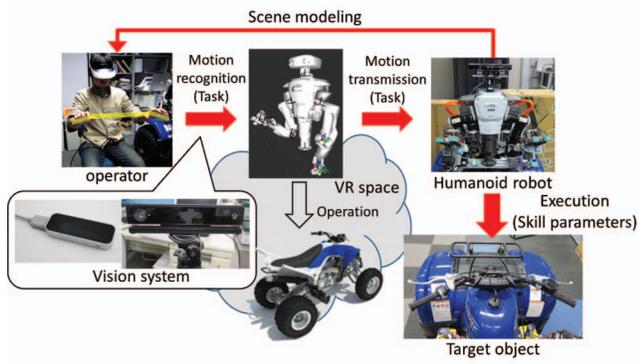


Fig. 3. Overview of teleoperation system

VR space.

- 2) The operator performs motions in the VR space. The tasks are recognized by observing the human motions.
- 3) The task information is transmitted to the robot. The robot executes the tasks by generating the motions based on the task information and skill parameters.
- 4) The states of the operator and the robot are visualized in the VR space for visual feedback.

#### A. Task recognition

The human motion is observed by vision-based sensors and is analyzed to find state transitions. Motion capture systems based on low cost depth-cameras such as Kinect and LeapMotion are widely available. The posture of the human body and fingers can be obtained with reasonable stability in real time. We discriminate status of operator corresponding to the state of the target using bone data from depth-cameras. By observing the state transition of the operator, the task that causes the transition can be recognized.

#### B. Motion generation

The trajectories of hands and fingertips are generated by skill parameters. As described above, we assume that rough skill parameters can be previously given. The trajectory is directly generated from the rough parameters. However, it does not work well because the state of the target object dynamically changes because of the vibration of the body and the change of contact states. Thus, the trajectory is adjusted according to the relative states of the target object and the hand. For estimating the relative positions and contact states, the tactile information is used. The position of the fingertip is sequentially controlled by calculating the destination point at each control period[18]. The trajectories of arms are generated by simple PTP control. Pose of each state(key pose) is given previously. Robot moves by PTP control between key points.

#### C. Visual feedback

We also apply the information of task model to the feedback for the operator. The proposed method recognizes tasks from gestures and generates motion. If it is possible for the operator to recognize the state of the robot, the operator

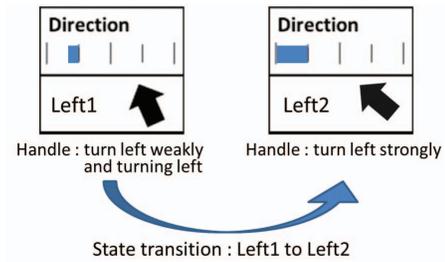


Fig. 4. Visualization of state transitions

can facilitate the selection of a task since it is possible to understand whether the operator can transition to any state. Additionally, the state of the operator is presented in order to determine the state transition. By this operation auxiliary interface, the operator can understand their own status and which corresponding task the robot tries to perform. (Fig. 4).

In our system, information of operator motion and visualized operation parameters are overlaid to camera image from the robot. In addition, virtual robot model is also displayed. The operator skeleton data captured from a depth-sensing camera is presented as information of operator motion. It is visualization of input data for task recognition. The parameter of operation indication is displayed on an operation auxiliary panel like Fig. 4. It represents visualization of internal data for task recognition. The virtual robot model is also overlaid. The virtual robot is given the same joint angle data as given to the real robot. Thus it means the visualization of an output data for task execution. The operator recognizes this output data without motion delay of robot by VR model indication. By the visualization of this information of the task recognition, our system improves the operability using only visual feedback.

## IV. DESIGN OF TASK MODEL

Driving operation is a continuous and complex operation. We disassemble driving operation as the combination of simple tasks. Target states are extracted from the structure of the target. State and transition are designed considering teleoperation.

#### A. Driving operations of buggy

The operations of driving the buggy-type vehicle consist of mainly handle, accelerator, and brake operations. The handle operation controls the traveling direction. The accelerator and brake operations control traveling acceleration. By this structure, each operation is limited to one trajectory and two directions. Therefore, the following six types of operations can be listed for the driving operation.

- Turn the handle to the right
- Turn the handle to the left
- Gripping accelerator lever
- Release accelerator lever
- Gripping brake lever
- Release brake lever

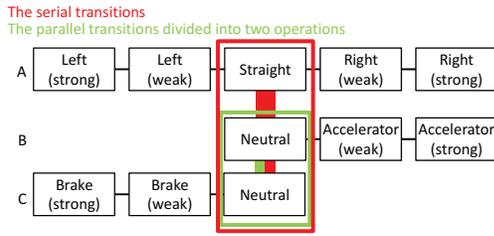


Fig. 5. State transition of driving operation

### B. State transition of driving operation

The driving operations are represented by several discrete states. For driving buggy-type vehicles, we need to operate handle, accelerator and brake as described above. Human beings can operate them continuously. But, it is difficult to recognize the continuous human motions accurately because of recognition errors and the temporal and spatial resolutions. Especially in the teleoperation system, the continuous recognition and transmission of motion doesn't work well because of the time-delay and the limitation of the network bandwidth. Therefore, we represent the driving operations by several states.

The handle operation is represented by five steps: straight, right (weak), right (strong), left (weak) and left (strong) see Fig. 5 A. The accelerator and brake operations are each represented by three steps: neutral, grip (weak), grip (strong) (Fig. 5 B, Fig. 5 C).

### C. Task model design

The tasks are defined as the transitions between the states. The three types of state transitions can be considered as the representations of the driving operation.

- 1) The operations are concurrently performed: the parallel transitions.
- 2) The operations are exclusively performed: the serial transitions.
- 3) The acceleration and the brake operations are exclusively performed.

(1) is based on the assumption that the human operates the target objects in parallel. This is the most flexible and is used by professional drivers. (2) assumes that the human operates the target objects one by one. This is a beginner's operation. The driving operation is divided into two operations: controlling the speed and changing the traveling direction in (3). In the preliminary experiments, it is found that (3) is the most natural and easy way to operate the buggy. The task model (3) is used in the following experiments.

## V. IMPLEMENTATION

### A. Teleoperation interface

Kinect RGB-D camera is put in front of the operator for recognizing the operator's body motion. LeapMotion with infrared depth cameras is put under the hands for recognizing the finger motions [19]. The views sent from the remote robot and the visual feedback superimposed into the scene

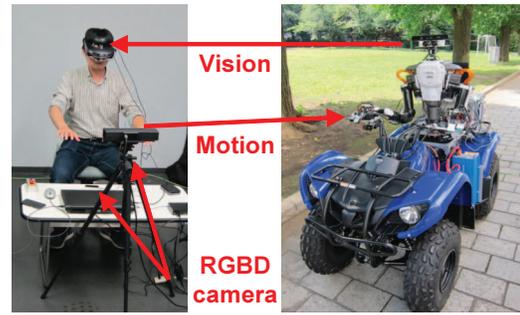


Fig. 6. Overview of the proposed system

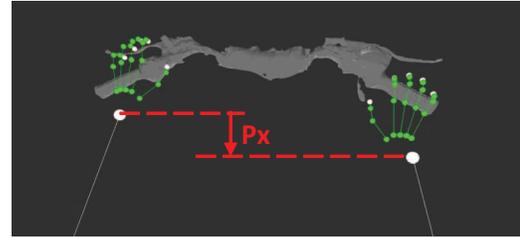


Fig. 7. Recognition of handle recognition

is displayed on the HMD. That is, the operator controls the robot in a Mixed Reality (MR) environment. The recognized motions are transmitted to the robot as tasks and applied to the robot.

### B. Task recognition

1) *Handle operation*: The handle parameter  $p_h$  is computed by  $p_x$  that is the distance between the left and right hands in x-coordinate as shown in Fig. 7.  $p_h$  is from 1 to  $-1$ . The state transition of the handle operation is recognized by simply thresholding  $p_h$  (Fig. 8. a).

2) *Accelerator operation*: The accelerator task is recognized by the angle of the thumb relative to the palm. Figure 9 shows the estimation of the acceleration parameter. The accelerator parameter  $p_a$  is represented by the angle  $\theta_a$  between vector  $\vec{v}_1$  and  $\vec{v}_2$ . The state transition is recognized by thresholding  $p_a$  (Fig. 8. b).

3) *Brake operation*: The brake task is recognized by a gesture of the hand. Opening or closing the hand is recognized by the relative positions of the fingers and translates to brake parameter  $p_b$  (Fig. 8. b). The task of gripping the brake or releasing the brake is executed by the state transition of opening or closing the human hand.

### C. Motion generation

We assume that the 3D structure of the buggy has been previously acquired. The rough positions and movable directions of the handle, accelerator, and brake levers are known. The motions of the operations are generated by considering the contact states of the robot hand and fingers with the target objects.

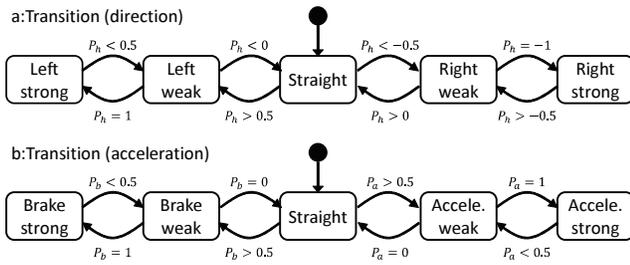


Fig. 8. State machine diagram

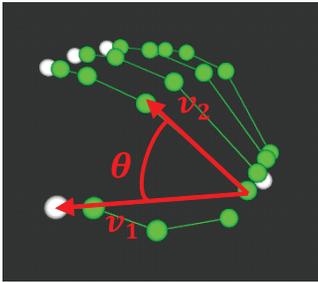


Fig. 9. Recognition of acceleration parameter

## VI. EXPERIMENT AND USER STUDY

In this experiment, we focus on visual feedback to the operator. Robot motion is assumed to be guaranteed to complete at a constant time from the instruction.

### A. Experimental setup

Upper body humanoid robot HIRO (Kawada industries, inc) is used for teleoperation robot. The robot does not have a lower body, and is fixed on the seat of the four-wheel buggy. The head of the robot moves, synchronizing with the human head motion taken by a 6DOF motion sensor. The robot hand consists of three fingers. Each finger has three degrees of freedom. A tactile sensor, ShokacCube(Touchence Inc.), is attached on each fingertip to measure the contact point. All tests are performed without running the buggy; the engine had not been started.

In this experiment, we evaluated the operability of the proposed interface with the visual feedback by user studies. The subjects were asked to operate the pre-defined tasks. The time from when the task is given to when the state transition is achieved is measured in two cases: with only visualization of recognition states (test A), and with visualization of both recognition states and tasks (test B). Test A shows skeleton model and VR model overlaid to robot camera image, and Test B shows task visualization added to Test A. The number of subjects was five, ages ranged from 22 to 27, 3 males and 2 females.

All subjects were asked to attempt a pre-defined sequence of tasks 15 times by following instructions displayed on the HMD. Task instruction was indicated on the upper side of the operator's view(Fig. 10). When the given target state is maintained for 1 second, the task is assumed to be performed.

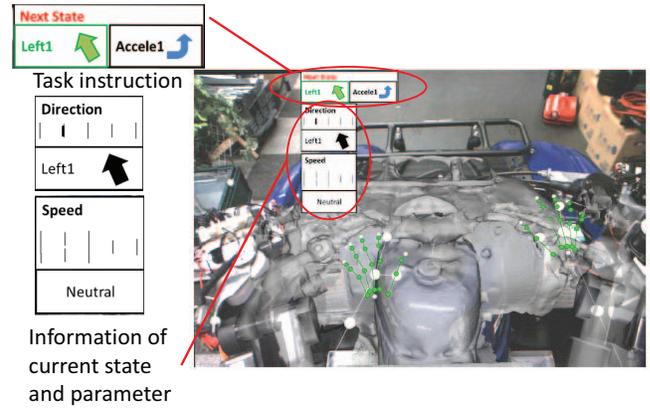


Fig. 10. Visualization of task recognition

TABLE I

	Test A [sec] (without task visualization)	Test B [sec] (with task visualization)
Subject a	373.818	193.627
Subject b	315.279	237.977
Subject c	545.545	174.835
Subject d	997.426	524.922
Subject e	461.443	388.838

The initial states of the task sequence were Straight (Handle) and Neutral (Acceleration, Brake).

### B. Experimental results

The average time required for each task is compared with Test A and Test B (Table I). The results show that the required time for Test B is shorter than that of Test A for every subject. Figure 11 shows an example of the operation time for each task (subjects a and e). In figure 11, the horizontal axis indicates the number of task from start, and the vertical axis indicates the required time for execution of each task. Results of experiment without visualization are colored red and with visualization are colored blue. The graph shows that when the tasks are visualized by the subject, the operation time becomes stable. It is clear that the visualization of the inner state of the robot and the motion observed by the sensors are important. Moreover, the visualization of state transitions is helpful to stably control the robot because the operator can comprehend all the states of the robot and the operator in real time.

### C. Discussion

The reason why subject e could achieve better results with Test A than with Test B for more than half of the transitions is because subject e tried Test B first and Test A later. Though the order of the experiments was randomly given, it is clear that the operation interface is easy to learn. However, in some tasks, subject e spent much more time in the later test. Even if the operator can easily get the hang of the system, the visual feedback is still necessary for checking the various states.

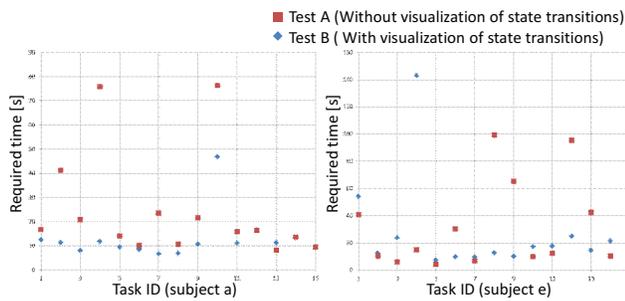


Fig. 11. Operation time of each task (subject a)

It is also expected that the difference of operation times between Test A and B becomes smaller as the operator continues the training. After the training, the operator may be able to recognize the states only by the views sent from the robot. The task visualization is required for reconfirming the concrete state. It can be said that the visual feedback plays the same role as the speedometer in a vehicle. The visual feedback is useful for operating the vehicle with a similar feeling of the actual operation.

## VII. CONCLUSION

In this paper, we proposed an interface to remotely operate the humanoid robot by the task model and the visual feedback system. The task model can simultaneously solve the issues of the structural differences between the human and the humanoid robot, and time delays. The proposed interface improves the operability by showing the operator motion as skeleton model, the state transitions parameters, and VR model that are overlapped into the real scenes.

In this work, we treat the driving of a buggy as a subject of study for a humanoid robot which operates equipment that is designed for a human. We implemented an interface of teleoperation system for driving a four-wheel buggy, and performed user studies for evaluating the operability of the interface. The results of the user study showed that the task visualization system is effective for the teleoperation.

In future work, experiments in an actual environment must be attempted. Moreover, the evaluation of the proposed method with large time-delay is required. It is required to evaluate how the operator changes his behavior when there are large time delays.

## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 24240034.

## REFERENCES

- [1] "Darpa robotics challenge," <http://www.theroboticschallenge.org/>.
- [2] R. O'Flaherty, P. Vieira, M. Grey, P. Oh, A. Bobick, M. Egerstedt, and M. Stilman, "Humanoid robot teleoperation for tasks with power tools," in *Technologies for Practical Robot Applications (TePRA)*, 2013 *IEEE International Conference on*, April 2013, pp. 1–6.
- [3] S. Tachi, K. Komoriya, K. Sawada, T. Nishiyama, T. Itoko, M. Kobayashi, and K. Inoue, "Telexistence cockpit for humanoid robot control," *Advanced Robotics*, vol. 17, no. 3, pp. 199–217, 2003.

- [4] K. Negishi, N. Ito, A. Namiki, T. Tanaka, A. Ueda, Y. Murakami, S. Ikeda, T. Wada, T. Moromoto, H. Fujimori, and S. Tachi, "i-marionet : Remote operated humanoid robot system controlled by telexistence fst(in japanese)," in *Proceedings of The Robotics and Mechatronics Conference2013(In Japanese)*, 2013, pp. 1A2–E06.
- [5] F. Wang, C. Tang, Y. Ou, and Y. Xu, "A real-time human imitation system," in *Intelligent Control and Automation (WCICA)*, 2012 *10th World Congress on*, July 2012, pp. 3692–3697.
- [6] V. V. Nguyen and J.-H. Lee, "Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot," in *System Integration (SII)*, 2012 *IEEE/SICE International Symposium on*, Dec 2012, pp. 93–98.
- [7] K. Ikeuchi and T. Suchiro, "Towards an assembly plan from observation. i. assembly task recognition using face-contact relations (polyhedral objects)," in *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, may 1992, pp. 2171 –2177 vol.3.
- [8] Y. E. Song, M. Niitsuma, T. Kubota, H. Hashimoto, and H. I. Son, "Mobile multimodal human-robot interface for virtual collaboration," in *Cognitive Infocommunications (CogInfoCom)*, 2012 *IEEE 3rd International Conference on*, Dec 2012, pp. 627–631.
- [9] M. Otsuki, T. Oshita, A. Kimura, F. Shibata, and H. Tamura, "Touch & detach: Physics-based unbinding and observation of complex virtual objects in 3d space," in *CHI Conference on Human Factors in Computing Systems Extended Abstract (CHI 2012)*, May 2012, pp. 1661–1666.
- [10] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, "A handle bar metaphor for virtual object manipulation with mid-air interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 1297–1306. [Online]. Available: <http://doi.acm.org/10.1145/2207676.2208585>
- [11] K. Kobayashi, K. Nishiwaki, S. Uchiyama, H. Yamamoto, and S. Kagami, "Viewing and reviewing how humanoids sensed, planned and behaved with mixed reality technology," in *Humanoid Robots, 2007 7th IEEE-RAS International Conference on*, Nov 2007, pp. 130–135.
- [12] T. Collett and B. MacDonald, "An Augmented Reality Debugging System for Mobile Robot Software Engineers," *Journal of Software Engineering for Robotics*, vol. 1, no. 1, pp. 18–32, 2010.
- [13] H. Kawasaki, K. Nakayama, and G. Parker, "Robot teaching based on intention of human motion in virtual reality environment(in japanese)," *Transactions of the virtual reality society of japan(In Japanese)*, vol. 5, no. 2, pp. 899–906, jun 2000.
- [14] Y. Wang, F. Sun, L. Hu, Z. Li, and H. Liu, "Space teleoperation with large time delay based on vision feedback and virtual reality," in *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*, July 2009, pp. 1200–1205.
- [15] L. Yunhong, W. Yubo, W. Mingxin, and Z. Pei, "Vision based teleoperation robot virtual reality circumstance modeling in real-time," in *Optoelectronics and Image Processing (ICOIP)*, 2010 *International Conference on*, vol. 2, Nov 2010, pp. 726–729.
- [16] K. Nagata, Y. Wakita, and E. Ono, "Task instruction by putting task information in work space," in *Robotics and Automation, 2007 IEEE International Conference on*, April 2007, pp. 305–310.
- [17] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, H. Hirukawa, and K. Ikeuchi, "Learning from observation paradigm: Leg task models for enabling a biped humanoid robot to imitate human dances," *Int. J. Rob. Res.*, vol. 26, no. 8, pp. 829–844, aug 2007.
- [18] M. Ogawa, K. Honda, Y. Sato, S. Kudoh, T. Oishi, and K. Ikeuchi, "Motion generation of the humanoid robot for teleoperation by task model," in *Robot and Human Interactive Communication (RO-MAN)*, 2015 *24th IEEE International Symposium on*, Aug 2015, pp. 71–76.
- [19] "Leapmotion sdk," <https://developer.leapmotion.com/>.