

Previewed Reality: Near-future perception system

Yuta Horikawa¹, Asuka Egashira¹, Kazuto Nakashima¹, Akihiro Kawamura², and Ryo Kurazume²

Abstract—This paper presents a near-future perception system named “Previewed Reality”. The system consists of an informationally structured environment (ISE), an immersive VR display, a stereo camera, an optical tracking system, and a dynamic simulator. In an ISE, a number of sensors are embedded, and information such as the position of furniture, objects, humans, and robots, is sensed and stored in a database. The position and orientation of the immersive VR display are also tracked by an optical tracking system. Therefore, we can forecast the next possible events using a dynamic simulator and synthesize virtual images of what users will see in the near future from their own viewpoint. The synthesized images, overlaid on a real scene by using augmented reality technology, are presented to the user. The proposed system can allow a human and a robot to coexist more safely by showing possible hazardous situations to the human intuitively in advance.

I. INTRODUCTION

An informationally structured environment (ISE) is a promising solution for realizing a service robot working with humans in their daily lives. In an ISE, a number of sensors are embedded, and information such as the position of furniture, objects, humans, and robots, is sensed and stored in a database. Therefore, the information required for a service robot to perform a task can be obtained on-demand and in real time if the robot accesses the database. Some examples of ISEs include “Robotic Room” [1] and “Intelligent Space” [2] at Tokyo University, “Smart Room” at MIT MediaLab [3], “Intelligent Room” at AILab [4], “Aware Home” at Georgia Tech. [5], and “Wabot House” at Waseda University [6]. Many of these ISEs are still being studied actively in laboratories [7] [8] [9] [10] [11].

We have been developing an ISE software platform named ROS-TMS 4.0 [12], which provides an information processing system for an ISE based on the Robot Operating System (ROS), as shown in Fig. 1. Fig. 2 shows the architecture of ROS-TMS 4.0. This system consists of several modules such as TMS.UR (user interface), TMS.TS (task planner and scheduler), and TMS.RC (robot controller). Each module uses multiple nodes, which is a minimal component in ROS. The total number of nodes in ROS-TMS 4.0 is over 100 [13].

In addition, an ISE hardware platform named “Big Sensor Box” (B-Sen) has been developed [14]. In B-Sen, a number of sensors, including eighteen optical trackers (Vicon Bonita), nine RGB-D cameras (Microsoft Kinect for

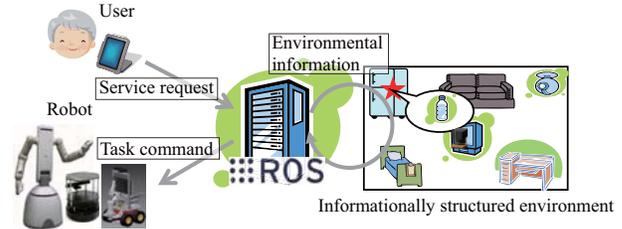


Fig. 1. ROS-TMS

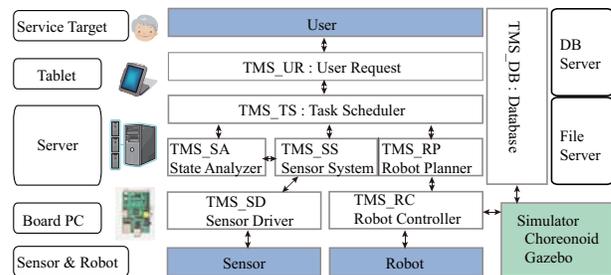


Fig. 2. ROS-TMS 4.0 architecture [12]

Xbox One), several laser range finders (URG-04LX-UG01, Hokuyo), and RFID tag readers are installed in a house with a bedroom, a dining room, and a kitchen (Figs. 3 and 4). Service robots are controlled by ROS-TMS 4.0 based on measured sensory data in B-Sen.



(a) Robots in B-Sen

(b) Sensors in B-Sen

Fig. 3. Big Sensor Box (B-sen)

Various immersive virtual reality (VR) displays have been put on the market in recent years. For example, Oculus Rift DK2 (Oculus VR) is a goggle-type immersive VR display device, which measures face direction using a gyro sensor, and VR images for the measured face direction are displayed on left- and right-eye screens. In [15], to increase the immersive feeling in the VR, human motion is measured by an optical tracker and stored in the VR world in real time. We have also been developing an immersive interface for ISE, which consists of an immersive VR display and an optical tracking system [16]. This system tracks the position of the immersive VR display in real time and displays a

¹Yuta Horikawa, Asuka Egashira, and Kazuto Nakashima are with Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan {horikawa, egashira, k.nakashima}@irvs.ait.kyushu-u.ac.jp

²Akihiro Kawamura and Ryo Kurazume are with Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan {kawamura, kurazume}@ait.kyushu-u.ac.jp

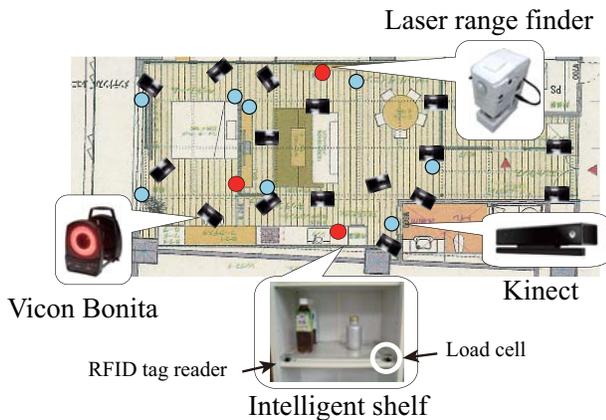


Fig. 4. Sensors in the Big Sensor Box

virtual image from the sensed eye position on the VR display. Unlike conventional immersive VR displays, not only the orientation but also the position of the display is measured. Therefore, a user wearing this device can walk in a room while wearing a goggle device and watching the VR image only.

This can be applied to situations in which humans and robots coexist with each other. If the user knows the next movement of the robot in advance, a hazardous situation can be avoided. However, if the user does not know the movement of the robot and the robot starts to move suddenly, a collision with the robot could occur. In ISE, all robot motions are planned in advance based on sensory information. It is possible to show the planned motion to the user before the robot starts moving. Although a warning light and beeping sound are simple ways to alert the user about the robot motion, it is more natural and intuitive if the user can directly see the next robot motion.

In this paper, we extend our immersive interface for ISE [16] and propose a new near-future perception system named “Previewed Reality”. With this system, users can intuitively perceive future events with their own eyes.

II. RELATED WORK

Immersive VR interfaces have been reported mainly in research on tele-robotics. Several applications using an immersive VR interface have been presented. The following are some examples: a surgical robot such as the da Vinci Surgical System (Intuitive Surgical, Inc.) [17], a surgical simulator [18], a space robot with motion control [19] and a rescue robot [20].

Immersive VR interfaces can be categorized into two groups. One group includes surround screen projection systems such as CAVE [21]. These utilize multiple screens and projectors to display whole directional images of walls, floors, and roofs independently of the user’s line of sight. The second group comprises wearable devices such as Google Glass or Epson Moverio (Fig. 5). These devices directly overlay a VR image onto a real image by using transparent glasses. Such devices have been attracting much attention because they are highly suitable for augmented reality (AR),

mixed reality (MR), and first-person vision. In recent years, new wearable devices such as Oculus Rift, HTC Vive, and Sony’s PlayStation VR have been produced as wearable immersive VR interfaces. These devices, which consist of a small display and an inertial measurement unit (IMU), synthesize a virtual image corresponding to the head direction. Wearable devices are simple and low-cost in comparison with systems with large screens and projectors, and thus have become very popular in recent years.



(a) Google Glass (b) Epson Moverio

Fig. 5. Smart glasses (Google Glass and Epson Moverio)

The time shift effect in VR, AR, or MR technologies is a popular technique, particularly in digital museums for showing lost cultural heritage objects. In the time shift effect, users can view ancient buildings overlaid on current scenery through a head mount display, a tablet, or a smart phone. For example, the Archeoguide project [22] reconstructed 3D models of the remains of ancient Greece. Kakuta et al. [23] developed a MR reproduction of the ancient capital of Asuka-Kyo in Japan, which was displayed as a video in a see-through head mounted display. In these systems, the MR content is created and stored beforehand. When a user views scenery through a display device, the MR content is overlaid on the screen. All of this content is static. On the other hand, in our system, near-future scenes are estimated and created in real time based on current sensory information. Thus, all of the content is dynamic. In addition, the time shift (a few seconds) is much shorter for systems such as Previewed Reality than that for displaying ancient heritage scenes (many years).

III. IMMERSIVE VR INTERFACE FOR INFORMATIONALLY STRUCTURED ENVIRONMENT [16]

This section introduces an immersive VR interface [16] that can connect a VR world and a real world in ISE. This system consists of an immersive VR display (Oculus Rift DK2, Oculus VR), a stereo camera (Ovrvision, Wizapply), shown in Fig. 6(a), an optical tracking system (Bonita, Vicon) shown Fig. 6(b), and a simulator (Choreonoid or Gazebo).

Oculus Rift DK2 is a wearable display that produces binocular stereo vision by displaying left- and right-eye images on each screen. The device has a three-axis gyro sensor that measures the head orientation and sends it to the PC. The position of Oculus Rift DK2 in the room is tracked by attached optical markers (Fig. 6(a)) and the Vicon Bonita optical tracking system (Fig. 6(b)) with an accuracy of less than one millimeter. Based on the measured position and orientation of Oculus Rift DK2, VR images of the room from the same viewpoint are synthesized and displayed on Oculus Rift DK2. Fig. 7 shows a dataflow for Oculus Rift DK2, Vicon Bonita, and a simulator (Choreonoid, Gazebo).

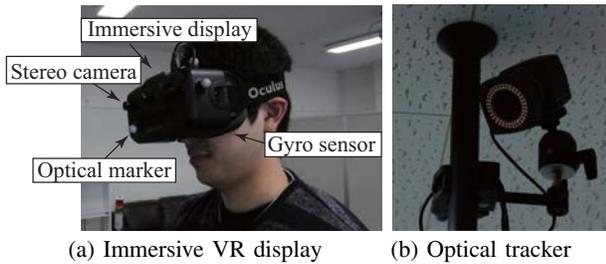


Fig. 6. Immersive VR display device consisting of (a) Oculus Rift DK2 (immersive stereo display) and Ovrvision (stereo camera), and (b) Vicon Bonita (optical position tracker)

The scene is first created using environmental information such as the position of objects and robots stored in the database in ROS-TMS 4.0. Then, the OpenGL-based graphics engine synthesizes distorted stereo images by calling a distortion function in libOVR (LibOVR 0.5.0, Oculus VR) in Choreonoid or by plug-in software in Gazebo. Finally, stereo images are individually displayed on the left- and right-eye screens in Oculus Rift DK2. The stereo camera (Ovrvision)

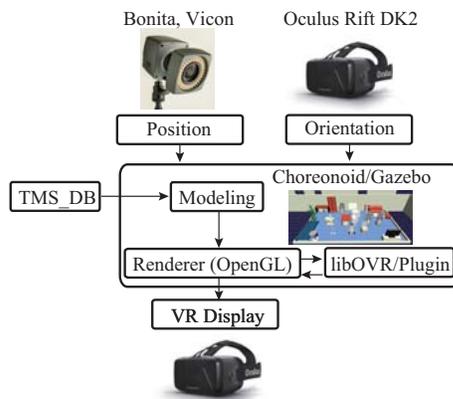


Fig. 7. Dataflow of the proposed immersive VR interface. The position of Oculus Rift DK2 is tracked by Vicon Bonita and the pose is sensed by the three-axis gyro sensor in Oculus Rift DK2. VR stereo images of the real world from the current viewpoint of the user are synthesized and displayed on Oculus Rift DK2 in real time.

is mounted on the front chassis of Oculus Rift DK2 and captures images from the viewpoint of the person wearing the Oculus Rift DK2 device. The images are displayed on the left- and right-eye screens of Oculus Rift DK2, and thus the surrounding real images can be seen without detaching Oculus Rift DK2.

In an experiment, we set up a walk-through path, shown in Fig. 8. Fig. 9 shows some synthesized stereo VR images for Oculus Rift DK2 and real images captured by the stereo camera when a person wearing the system walks through the room (Fig. 8). From this experiment, we can confirm that the VR images are quite similar to the real images. The VR images are created and displayed based on head position and orientation, even if the person wearing this system walks anywhere in the room. For example, a user can sit on a chair in the VR world or pick up a virtual object from a shelf.

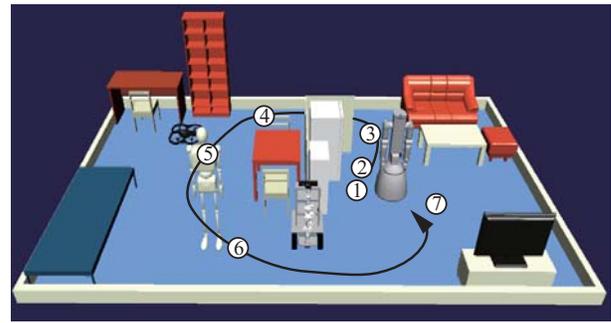


Fig. 8. Walk-through path in a room

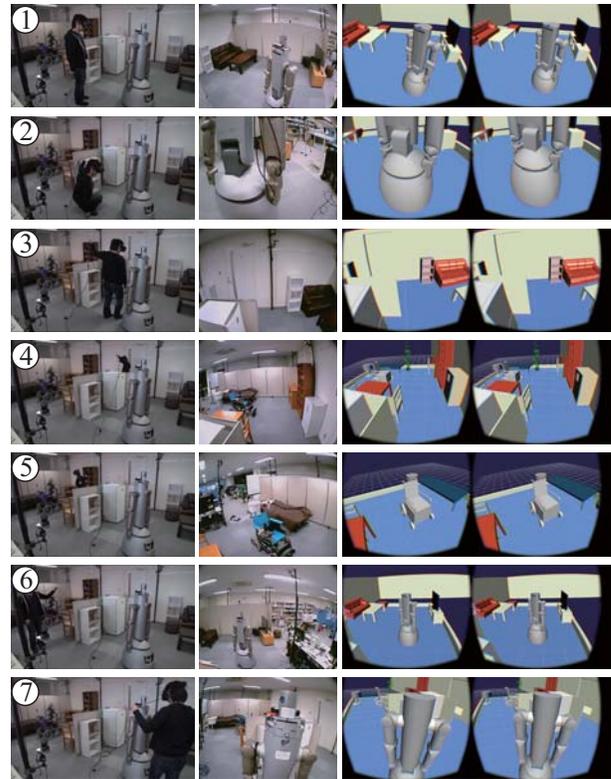


Fig. 9. Walk-through images of a room displayed on the stereo screens of Oculus Rift DK2. The two columns on the left show real images, and the two columns on the right show synthesized VR images from the current viewpoint. The VR images are quite similar to the real images.

IV. NEAR-FUTURE PERCEPTION SYSTEM NAMED PREVIEWED REALITY

In this section, we propose Previewed Reality, which lets a user perceive near-future events (i.e., before the events occur) by using the immersive VR interface presented in Section III.

The key technique of Previewed Reality is the time shift effect. In an informationally structured environment such as B-Sen, information is obtained using a sensor network, and robot motion can be planned beforehand. The near-future events can be forecast by the robot planner or simulators such as Choreonoid or Gazebo. A simulation is performed based on the current real status of objects, furniture, humans, and robots sensed and stored by ROS-TMS. Therefore, we can show highly realistic VR images of near-future events

to a human via an immersive VR display or smart glasses.

A. System overview

In B-sen, information such as the position of objects on a shelf or table or in a refrigerator, and the pose of a human and a robot, are sensed in real time and stored in the database using TMS_SD in ROS-TMS 4.0 (Fig. 2), as shown in Fig. 10. Fig. 11 shows a robot, a can, and a table registered

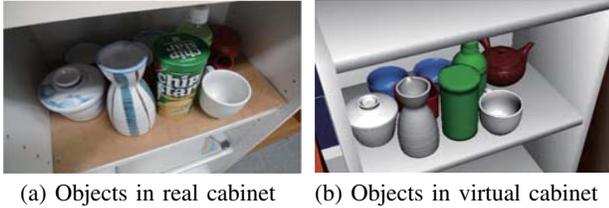


Fig. 10. Objects in (a) real and (b) virtual worlds. The location and type of objects on a virtual shelf are automatically registered in the database by RFID tags and load cells.

automatically in the database. These images are displayed in real time on the immersive VR interface.

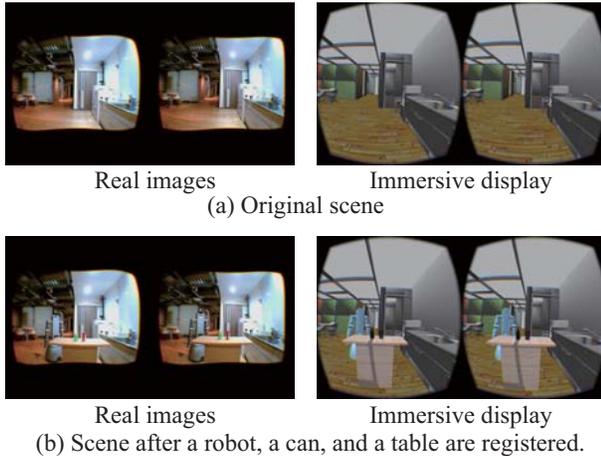


Fig. 11. Robots, objects, and furniture automatically registered in the simulator and displayed as the images on the immersive VR interface.

Simulations reflecting the current situation of objects, robots, and humans in B-sen can be performed in ROS-TMS 4.0 using a kinematic simulator such as Choreonoid or a dynamic simulator such as Gazebo. For example, we can estimate in advance whether a robot will grasp an object using its current finger configuration, or if it fails, in which direction the object will bounce, as shown in Fig. 12. Thus, if we can estimate the motion of objects or robots in B-sen, we can predict near-future events.

We developed the near-future perception system, Previewed Reality, by combining the immersive VR interface and the estimation of near-future events in B-sen. For example, we consider the case in which a robot raises its hand to give an object to a user. If the user can see the robot motion a few seconds before it occurs and senses danger, the user can leave his/her current position or stop the robot motion in advance. Therefore, Previewed Reality allows us to co-exist with robots more safely.

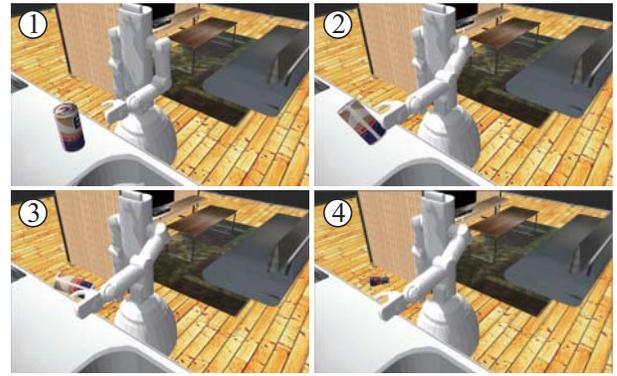


Fig. 12. Dynamic simulation of object grasping by Gazebo. In this case, the robot fails to grasp the object, which then bounces away.

B. System configuration

To realize Previewed Reality, the system has to show an actual scene to a user through the immersive VR interface. Though we can use see-through smart glasses such as Moverio (Epson), we adopted a stereo camera (Ovrvision) that can be attached on Oculus Rift DK2, as shown in Fig. 6. We captured RGB images at 60 Hz. As a dynamic simulator, we used Gazebo 6, which supports Oculus Rift DK2. Virtual images synthesized by Gazebo are overlaid on real images taken by the stereo camera and provided to the left and right eyes individually. In our system, the eye position in the real world must coincide very precisely with the eye position in the simulator. An error in these positions induces an uncomfortable feeling in the user, because the real image and the virtual image do not overlap precisely. Therefore, we carefully calibrated the camera parameters of the stereo camera and Oculus Rift DK2, and the display parameters in Gazebo. We also adjusted dynamic parameters such as the mass and shape of the objects in B-Sen so that we could precisely reflect the real-world environmental information in the dynamic simulation.

The motions of robots in B-sen are planned by ROS-TMS, which plans the performance of each robot, the tasks, the positions, and the objects [12]. Therefore, the planned motions can be examined in advance by the Gazebo simulator. By showing the planned motions on the screens of Oculus Rift DK2, the user can see the motions before the robot actually performs them. For example, hand motions or joint trajectories of the arm for grasping tasks (e.g., Fig. 12) are designed by the motion planner MoveIt! in the TMS_RP module of ROS-TMS (Fig. 2). Therefore, before sending planned commands to the actual robot, the virtual robot in Gazebo can perform the desired motions by sending the same commands to Gazebo. In addition, because Gazebo is a dynamic simulator, it is possible to predict interactions such as collision or slippage involving the robot and objects. The trajectory of the robot to a desired location is planned in ROS-TMS as Voronoi edges [12]. As a result, we can display the robot movement along the desired trajectory on the screens of Oculus Rift DK2 before the robot actually moves on the floor.

V. EXPERIMENT OF PREVIEWED REALITY

We conducted two kind of experiments by the developed system to confirm the validity and the performance of Previewed Reality.

First, we conducted experiments to confirm that the user can see the virtual image from the current viewpoint in B-Sen. Fig. 13 shows the human positions and corresponding images displayed on the screens of Oculus Rift DK2. Because the user can see that the virtual images almost coincide with the actual images from the viewpoints measured by Vicon Bonita, the user can safely walk in B-Sen.

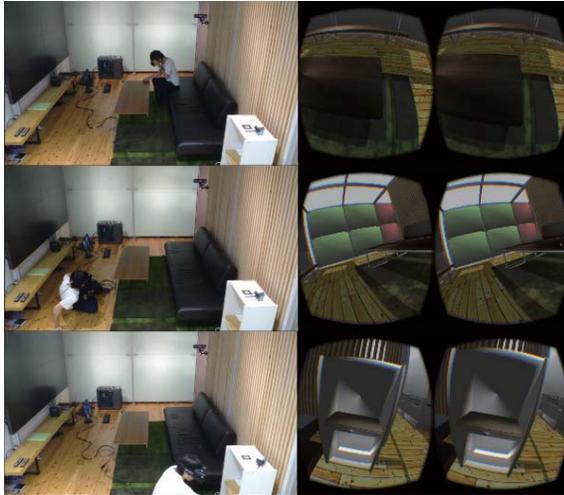
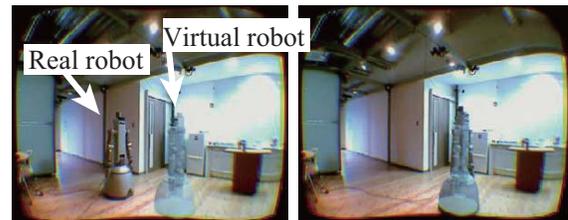


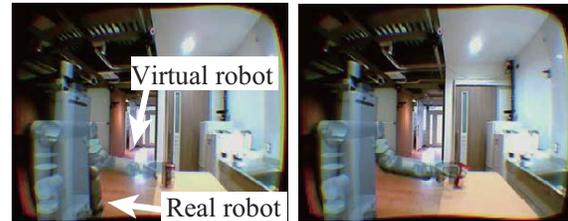
Fig. 13. Real images (left) and virtual images displayed on the immersive VR display (right). The current viewpoint of the user is measured by the optical tracking system in B-Sen.

Next, we conducted basic experiments to confirm Previewed Reality using the developed system. We used a humanoid service robot, SmartPal V (Yaskawa Electric) in the experiment. Fig. 14 shows the overlaid images of real scenes taken by the stereo camera and the virtual robot, whose motions are planned by ROS-TMS in advance. Fig. 15 shows another example of Previewed Reality. In this case, if the user approaches the robot and the robot has been instructed to start moving within a short time, the virtual robot colored red, which means the user will collide with the robot, appears on the screens of Oculus Rift DK2, and the user can see the motion that will soon be performed by the real robot. More precisely, the same commands determined by the motion planner MoveIt! are sent to the virtual and actual robots. However, by using AR techniques to show the planned motion by the virtual robot to the user in advance, the user can recognize the robot motion directly and avoid hazardous situations that could occur a few seconds later.

One of the current problems in the experimental system is the latency between the measurement of the optical tracker and image production. Because a large latency causes motion sickness, it is necessary to make the latency as small as possible. At present, the typical latency is about 0.2 s in our system. Although we employed a gyro sensor in Oculus Rift DK2 to suppress the latency, currently the sensor is not



(a) Virtual robot moves and real robot traces it.



(b) Virtual robot grasps and real robot traces it.

Fig. 14. Previewed Reality displayed on the screens of Oculus Rift DK2

satisfactory. Another problem is the calibration error in the immersive VR display and the stereo camera. Because both optical systems are subject to a large amount of distortion, as shown in Fig. 16, we removed the distortion as much as possible. However, precise calibration is quite difficult and remains an open problem.

VI. CONCLUSIONS

The proposed Previewed Reality system, which displays near-future scenes to a user directly and intuitively, is beneficial in situations where humans and robots are close together, such as in a house or a production line in a factory. Most of studies in robotics have been focused to plan and execute collision free trajectories and tasks by robots. On the other hand, in the present study, the user avoids the collision with the robot at once by watching future robot behaviors, and we believe this would be another solution for human and robot to coexist safely. Currently, the developed system uses a goggle-type immersive VR display and a stereo camera. However, we can use smart glasses with screens, such as Epson Moverio (Fig. 5), which is lighter and easy to use. Thus, we believe the proposed Previewed Reality can be used in our daily lives in the near future. In addition, because the timing for displaying near-future events is quite interesting from the standpoint of psychology, we intend to investigate the proper timing for a human to feel comfortable.

ACKNOWLEDGMENT

This research was supported by The Japan Science and Technology Agency (JST) through its “Center of Innovation Science and Technology based Radical Innovation and Entrepreneurship Program (COI Program)” and by a Grant-in-Aid for Challenging Exploratory Research (16K14199).

REFERENCES

- [1] T. Sato, Y. Nishida, and H. Mizoguchi, “Robotic room: Symbiosis with human through behavior media,” *Robotics and Autonomous Systems*, vol. 18, no. 1-2, pp. 185–194, 1996.

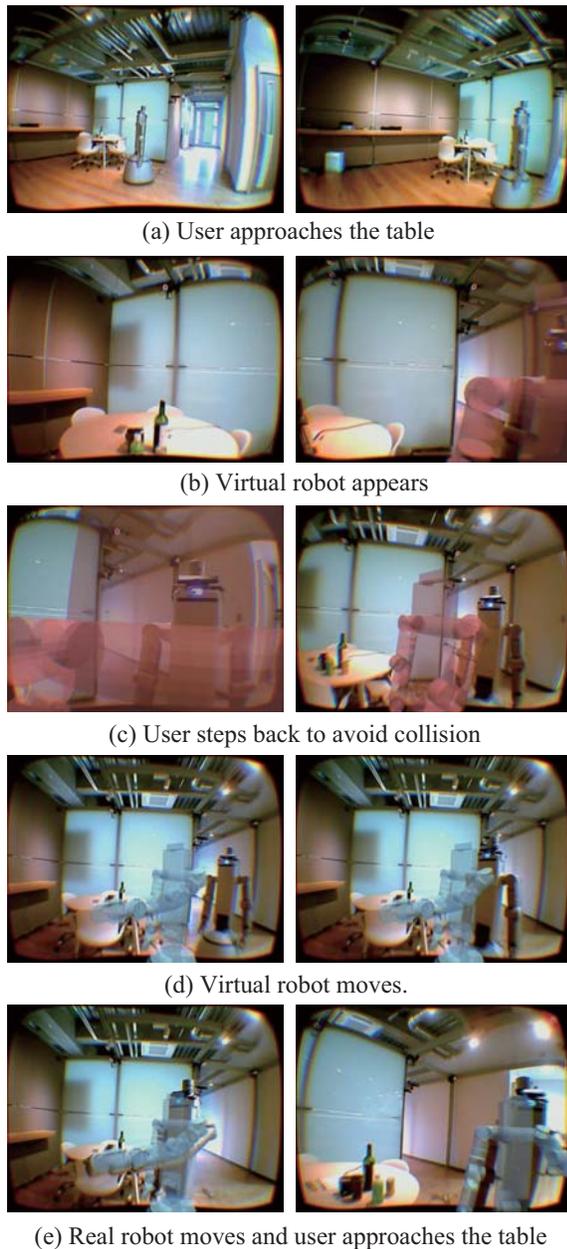


Fig. 15. If the user approaches the robot, which is instructed to start moving, the virtual robot is displayed to notify the user about this movement in advance.

- [2] J.-H. Lee, N. Ando, and H. Hashimoto, "Design policy of intelligent space," in *Proceedings on 1999 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, pp. 1077–1082 vol.3, 1999.
- [3] A. P. Pentland, "Smart rooms," *Scientific American*, vol. 274, no. 4, pp. 54–62, 1996.
- [4] R. A. Brooks, "The intelligent room project," in *2nd International Conference on Cognitive Technology (CT '97)*, CT '97, (Washington, DC, USA), pp. 271–278, IEEE Computer Society, 1997.
- [5] J. A. Kientz, S. N. Patel, B. Jones, E. Price, E. D. Mynatt, and G. D. Abowd, "The georgia tech aware home," in *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, (New York, NY, USA), pp. 3675–3680, ACM, 2008.
- [6] S. Sugano and Y. Shirai, "Robot design and environment design - waseda robot-house project," in *International Joint Conference SICE-ICASE, 2006*, pp. 1–31–I–34, Oct 2006.
- [7] H. Noguchi, T. Mori, and T. Sato, "Automatic generation and connection of program components based on rdf sensor description in network

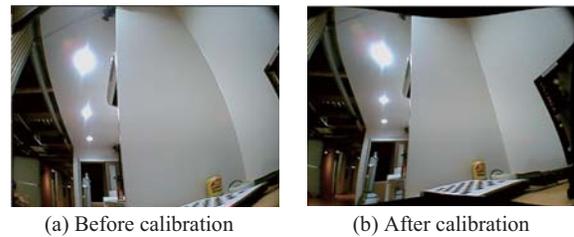


Fig. 16. Distorted stereo image and calibrated image.

- middleware," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2008–2014, 2006.
- [8] K.-H. Park, Z. Bien, J.-J. Lee, B. K. Kim, J.-T. Lim, J.-O. Kim, H. Lee, D. H. Stefanov, D.-J. Kim, J.-W. Jung, *et al.*, "Robotic smart house to assist people with movement disabilities," *Autonomous Robots*, vol. 22, no. 2, pp. 183–198, 2007.
- [9] Y. Kato, T. Izui, Y. Tsuchiya, M. Narita, M. Ueki, Y. Murakawa, and K. Okabayashi, "Rsi-cloud for integrating robot services with internet services," in *IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society*, pp. 2158–2163, 2011.
- [10] H. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, C. Martin, T. Langner, and M. Merten, "I'll keep an eye on you: Home robot companion for elderly people with cognitive impairment," in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2481–2488, 2011.
- [11] M. Tenorth, A. Perzylo, R. Lafrenz, and M. Beetz, "The robearth language: Representing and exchanging knowledge about actions, objects, and environments," in *IEEE International Conference on Robotics and Automation*, pp. 1284–1289, 2012.
- [12] Y. Pyo, K. Nakashima, S. Kuwahata, R. Kurazume, T. Tsuji, K. Morooka, and T. Hasegawa, "Service robot system with an informationally structured environment," *Robotics and Autonomous Systems*, vol. 74, no. Part A, pp. 148–165, 2015.
- [13] "Ros-tms." http://irvs.github.io/ros_tms/.
- [14] R. Kurazume, Y. Pyo, K. Nakashima, T. Tsuji, and A. Kawamura, "Feasibility study of iort platform "big sensor box"," in *IEEE International Conference on Robotics and Automation*, 2017.
- [15] S. Marks, J. E. Estevez, and A. M. Connor, "Towards the holodeck: Fully immersive virtual reality visualisation of scientific and engineering data," in *29th International Conference on Image and Vision Computing New Zealand (IVCNZ '14)*, pp. 42–47, 2014.
- [16] Y. Pyo, T. Tsuji, Y. Hashiguchi, and R. Kurazume, "Immersive vr interface for informationally structured environment," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM2015)*, pp. 1766–1771, 2015.
- [17] J. F. Boggess, "Robotic surgery in gynecologic oncology: evolution of a new surgical paradigm," *Journal of Robotic Surgery*, vol. 1, pp. 31–37, 3 2007.
- [18] M. Bro-Nielsen, J. L. Tasto, R. Cunningham, and G. L. Merril, "Preop endoscopic simulator: a pc-based immersive training system for bronchoscopy," *Stud Health Technol Inform*, vol. 62, pp. 76–82, 1999.
- [19] <http://www.engadget.com/2013/12/23/nasa-jpl-control-robotic-arm-kinect-2/>.
- [20] H. Martins, I. Oakley, and R. Ventura, "Design and evaluation of a head-mounted display for immersive 3d teleoperation of field robots," *Robotica*, vol. FirstView, pp. 1–20, 1 2015.
- [21] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the cave," in *20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, (New York, NY, USA), pp. 135–142, ACM, 1993.
- [22] V. Vlahakis, M. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, D. Stricker, T. Gleue, P. Daehne, and L. Almeida, "Archeoguide: an augmented reality guide for archaeological sites," *IEEE Computer Graphics and Applications*, vol. 22, no. 5, pp. 52–60, 2002.
- [23] T. Kakuta, T. Oishi, and K. Ikeuchi, "Development and evaluation of asuka-kyo mr contents with fast shading and shadowing," *Journal of Image Information and Television Engineers*, vol. 62, no. 9, pp. 1466–1473, 2008.