Proceedings of the 2014 IEEE/SICE International Symposium on System Integration, Chuo University, Tokyo, Japan, December 13-15, 2014

Cloud VR System with Immersive Interfaces to Collect Human Gaze-Controls and Interpersonal-Behaviors

Yoshinobu Hagiwara, Yoshiaki Mizuchi, Yongwoon Choi, and Tetsunari Inamura

***Abstract*— In this study, we present a cloud VR system with immersive interfaces for collecting human gaze-controls and interpersonal-behaviors. Human beings can log in to a VR space and naturally communicate with each other and a robot by using immersive interfaces. An HMD (Head Mounted Display) device and a Kinect sensor provide immersive visualization and motion control, respectively. If users have these devices, they can attend experiments with human-robot interactions in the VR space from around the world. Human gaze-controls and interpersonal-behaviors are collected to database while human beings and a robot interact with each other. The proposed VR system enables to collect huge number of human gaze-controls and interpersonal-behaviors by using immersive interfaces. Application experiments to learn object attributes from human beings and to observe greetings by two persons demonstrate the effectiveness of the proposed VR system to collect human gaze-controls and interpersonal-behaviors.**

I. INTRODUCTION

Recently, the development of service robots to support human daily activities have been attracted attention. One of important issues to realize these robots is to provide services based on natural interactions between human beings and a robot.

Yamaoka et al. have proposed a method that a robot obtains a communication-protocol to explain about items by using the database of human behaviors [1]. Yamakata et al. have proposed the method for the disambiguation of object reference by learning object attributes from human beings [2]. Sumi et al. have proposed IMADE room to collect human behaviors in daily activities by using wearable and fixed sensors [3]. Several other researches have been proposed to realize natural interactions between human beings and a robot [4], [5], [6].

However, in these studies, the number of experiments to collect human behaviors remains within about ten times, because costs for preparing experimental environments and the maintenance of a real robot become higher. As one of solutions for this problem, the use of a simulator might reduces the cost of interaction experiments with human beings and a robot.

Kamide et al. have performed human-robot interactions with realistic visualization using a cave VR system [7]. Weiss et al. have also performed human-robot cooperations in a VR system in oder to observe human emotions [8]. However, if these VR systems apply to cloud computing, the physical constraint of these systems becomes a critical problem. To solve this problem, cloud VR system can reduce the cost of interaction experiments in oder to collect human behaviors.

As a research to collect human behaviors by using a cloud VR system, Bosc et al. have proposed a data mining method using human-behaviors collected from an on-line game [9]. However, collected human-behaviors were limited to control-motions such as handling. Human gaze, entire body-motion, and voice should be collected for high-level human-robot interactions.

In this study, we propose a cloud VR system with immersive interfaces in oder to collect human gaze-controls and interpersonal-behaviors. Multiple users can log in to the proposed VR system and naturally interact with each other and a robot by using immersive interfaces. Human gaze-controls and interpersonal-behaviors are collected to database while human beings and a robot interact with each other. The proposed VR system enables to collect huge number of human behaviors through the Internet. Application experiments to learn object attributes and to observe greetings demonstrate the effectiveness of the proposed VR system to collect human gaze-controls and interpersonal motions.

II. PROPOSED VR SYSTEM

A. Cloud VR System

The proposed VR system was developed based on a VR system called SIGVerse [10]. SIGVerse makes it possible for multi users to login to a VR space as an avatar and communicate with a robot through the Internet as shown in Fig. 1. Therefore, human-robot interactions can be effectively performed in a VR space by using SIGVerse.

Fig. 2 shows the system construction of SIGVerse. SIGServer provides physics, dynamics, and communication simulations through a inter process communication manager. An agent controller which is a program for a robot and an avatar sends control information for them to SIGServer and receives a simulation result. A service provider which is a program for recognition and perception receives sensing information from SIGServer and sends a calculation result. Users can access to SIGServer using a user client called SIGViewer. SIGViewer sends user motions captured by user interfaces to SIGServer and receives a simulation result. For example, an HMD(Head Mounted Display) device and a Kinect sensor enable immersive visualization and motion-control respectively.

This work was not supported by any organization

Y. Hagiwara is with the National Institute of Informatics, Tokyo, Japan yhagiwara@nii.ac.jp

Y. Mizuchi and Y. Choi are with Soka University, Tokyo, Japan choi@soka.ac.jp

T. Inamura is with the National Institute of Informatics/the Graduate University for Advanced Studies, Tokyo, Japan inamura@nii.ac.jp

978-1-4799-6944-9/14/$31.00 ©2014 IEEE

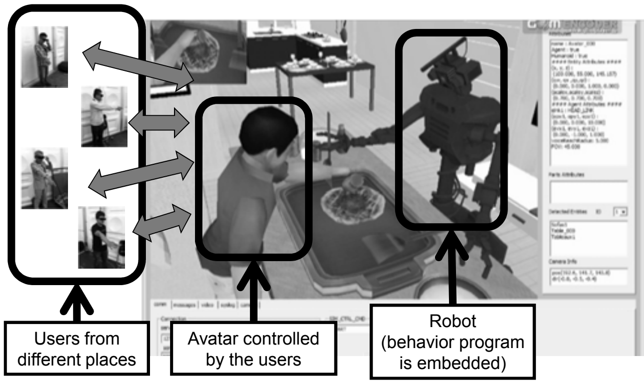


Fig. 1. Multi-users can login to the VR space as an avatar and communicate with a robot

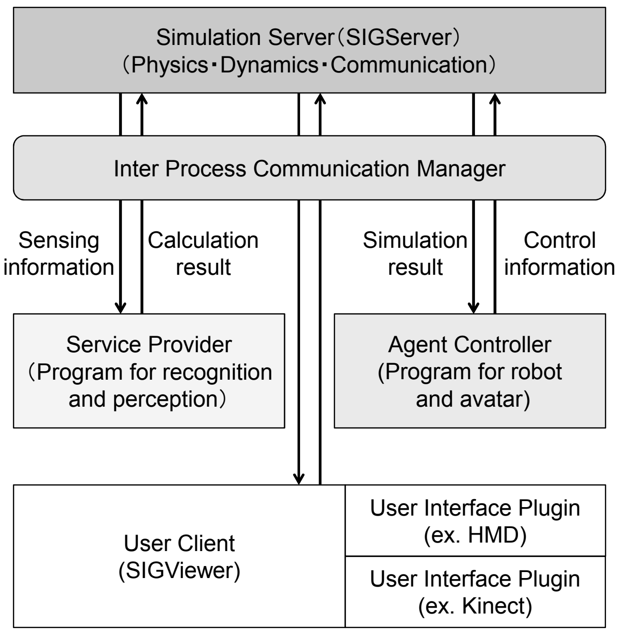


Fig. 2. System construction of SIGVerse

B. Immersive interfaces

Fig. 3 shows the available user-interface plug-ins in SIGVerse. (a) and (b) show an HMD and a 3D VR Headset to provide immersive visualization and to capture gaze motions. (c), (d), and (e) show a 3D motion sensor, a hand motion controller, and a finger motion controller to provide immersive motion-controls for a body, a hand, and fingers, respectively. (f) shows a 3D haptic device to provide haptic feedback from a VR space. Users can chose suitable user-interfaces for the purpose of their studies and download its plug-ins from the Internet. In addition, users can make the original plug-ins of user interfaces using SIGService library and SDK for their devices. In following sections, the effectiveness to collect human behaviors by the proposed VR system with immersive interfaces is evaluated from experiments to learn human knowledge and to observe human behaviors.

Fig. 3. Available immersive interfaces in SIGVerse

III. Learning Experiment for Human Knowledge

A. Experimental Set-up

We performed an experiment to evaluate the effectiveness of the proposed VR system in the learning of human knowledge. Fig. 4 shows an experimental set-up to learn human knowledge. A human being in the real space and a virtual robot in a VR space can communicate by their voice input through a microphone. An avatar in a VR space moves based on the motion of a real human being by using a 3D motion sensor. A first-person view of an avatar in a VR space is provided to a human being in the real space through a 3D VR headset. Also a human beings can naturally look around a VR space through a 3D VR headset. A human being can observe a VR space by using these immersive interfaces. Arbitrary human beings can control an avatar in a VR space from remote places. Therefore, it is expected to efficiently learn human knowledge without time and place restrictions.

B. Experimental Condition

As an example to learn human knowledge, a robot learn object attributes from human beings by using the proposed VR system. Fig. 5(a) shows the environment of learning experiments in a real space. A 3D headset called Oculus Rift was used to provide immersive visualization and to capture gaze-controls. A 3D motion sensor called Xtion was also used to capture entire-motion of the human subject in a real space. A subject can teach object attributes in the VR space through an operator instead of a microphone. In a VR space,

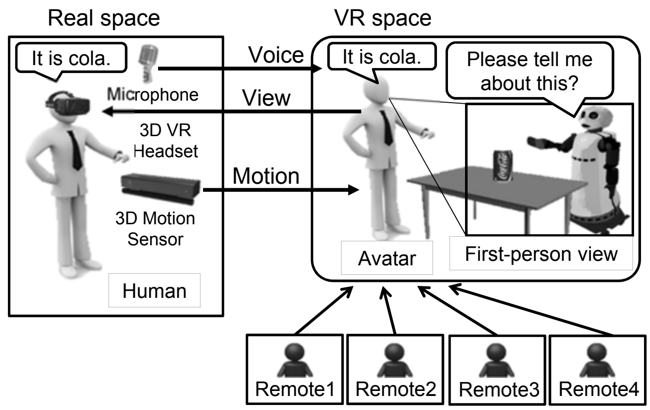
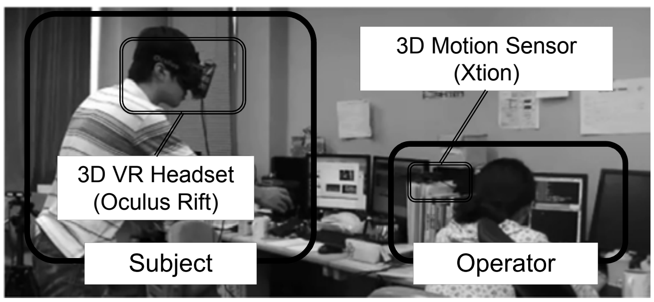
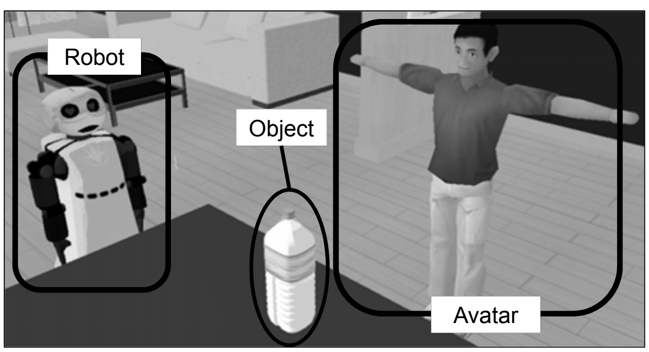


Fig. 4. Experimental set-up to learn human knowledge



(a) Environment in real space



(b) Environment in VR space

Fig. 5. Experimental environments to learn object attributes

a robot and an avatar are in front of an object placed on a table as shown in (b).

In the experiment, the first-person view of an avatar is provided to a subject by using Oculus Rift, as shown in Fig. 6. Since different views are projected to left and right screen of Oculus Rift, a subject can perspectively observe an object. A subject can grasp a target object in the VR space as shown in Fig. 6 in order to observe the attributes of the target object in detail while considering the scale of the target object depending on a relative distance.

In this experiment, we collect the attributes of three virtual-objects from two persons. The robot in the VR space questioned the human subject about the attributes of the three objects: name, color, shape, state, and how to use. Fig. 7 shows the three objects used in the experiment. We chose

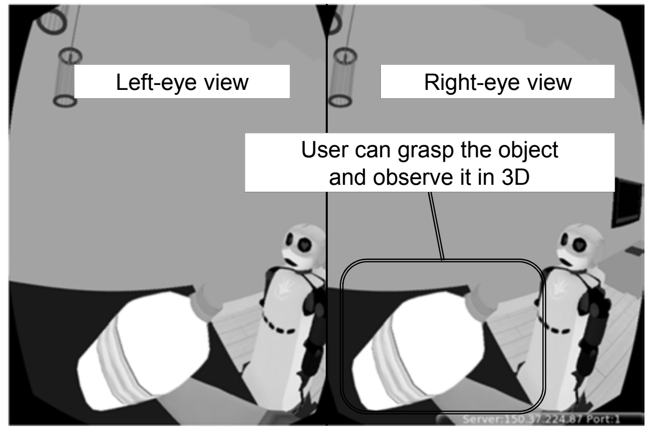


Fig. 6. Immersive visualization by Oculus Rift

(a) PET bottle1 (b) PET bottle2 (c) PET bottle3

Fig. 7. Objects for learning experiments

three types of PET bottles: 2L size of full PET bottle, 2L size of empty PET bottle, and 500mL size of empty PET bottle.

C. Experimental Result

Fig. 8 shows object attributes collected from two persons in the experiment. For example, we assume that a human being orders "Please bring me a green PET bottle". A robot searches for an object that has object attribute of “green” and “PET bottle” from database of object attributes to identify the ordered object. In this case, the robot can identify “PET bottle 1” based on the database of object attributes. As the other example, we assume that a human being orders "Please throw away an empty PET bottle". The robot can find “PET bottle 2” and “PET bottle 3” as an object which has attributes of “empty” and “PET bottle” based on the database of object attributes. In such case, the robot can ask additional object attributes to identify the ordered object, such as "What is size of target objects?". If the human being answers that "It is small one", the robot can identify the “PET bottle 3”. In addition, the proposed VR system can collect suitable motion for each object by using a 3D motion sensor. Human gaze-controls to observe objects is also collected by a 3D headset. The experimental result demonstrates the feasibility of the proposed VR system with immersive interfaces in oder to learn human knowledge.

Object	Subject	Name	Color	Shape	State	How to use
PETbottle 1	Subject A	PET bottle	Transparent color	Elongate shape	Full	Drink
		Green tea	Green	Square	Unopened state	Store
		A drink	Yellowish green			Pour
	Subject B	Green tea	Green	2L	Full	Pour
		PET bottle	White	Large	New	Drink
			Yellowish green		Unopened state	
PETbottle 2	Subject A	PET bottle	Sky blue	Elongate shape	Empty	Pour water
		Water	Transparent color	Sharp at the end		
				Uneven surface		
	Subject B	PET bottle	Sky blue	2L	Empty	Drink
		Water	Transparent color	Large		Throw
			Blue	Uneven surface		
PETbottle 3	Subject A	Coca cola	Red	Elongate shape	Empty	Pour
		PET bottle	Transparent color	Round shape		Twisting
			White	Sharp at the end		
	Subject B	PET bottle	Red	Small	Empty	Drink
		Coca cola	Transparent color	500mL	Popular	Throw
				Elongate shape		Brew

Fig. 8. Collected object-attributes in learning experiments

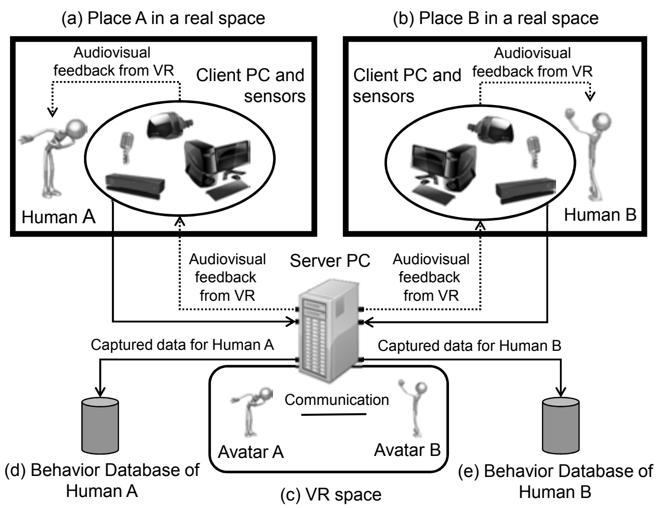


Fig. 9. Experimental set-up to observe human behaviors

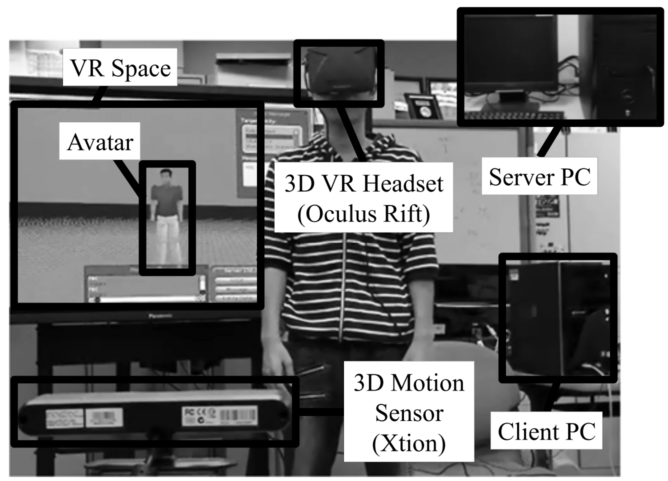


Fig. 10. An experimental environment to observe greetings

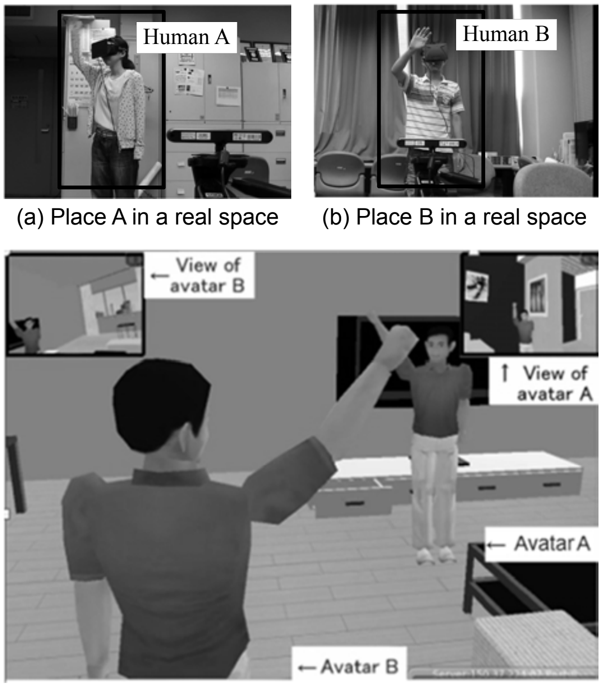
IV. Observing Experiment for Human Behaviors

A. Experimental Set-up

We performed an experiment to evaluate the effectiveness of the proposed VR system to observe human behaviors. Fig. 9 shows an experimental set-up to observe human behaviors. As shown in (a) and (b), a microphone, a 3D VR headset, and a 3D motion sensor were used to capture voice, motions and gazes, respectively. Then captured data by these sensors are sent to the server PC through the client PC placed at each place. The server PC calculates a VR space to perform communications between two persons and collects human behaviors to a database for each person as shown in (c). Two persons at the places A and B in a real space can communicate with each other through avatars in the VR space. Communications in the VR space are realized by applying captured data in each client PC to another avatar. In addition, as the feedback of communications in the VR space, the server PC sends a view on the eye sight of each avatar to each person in a real space through a 3D VR headset. By using this system, a person at the place A can communicate with another person at the place B in a real space through the VR space. In order to analyze communications between two persons, behaviors by two avatars in the VR space are observed and collected to the behavior database of human A and B shown in (d) and (e).

B. Experimental Condition

As an example to observe human behaviors, a greeting between two persons are observed by using the proposed VR system with immersive interfaces. Fig. 10 shows the experimental environment to observe greeting between two persons. Three kinds of sensors were used to capture human behaviors from a real space. Human gaze-controls were captured by a 3D VR headset called Oculus Rift which has a gyro sensor to detect head directions. Human motions were captured by a 3D motion sensor called Xtion. Human voice was captured by a microphone with a voice recognition



(a) Place A in a real space (b) Place B in a real space

(c) Greeting between two persons in a VR space

Fig. 11. Scenes of an observing experiment

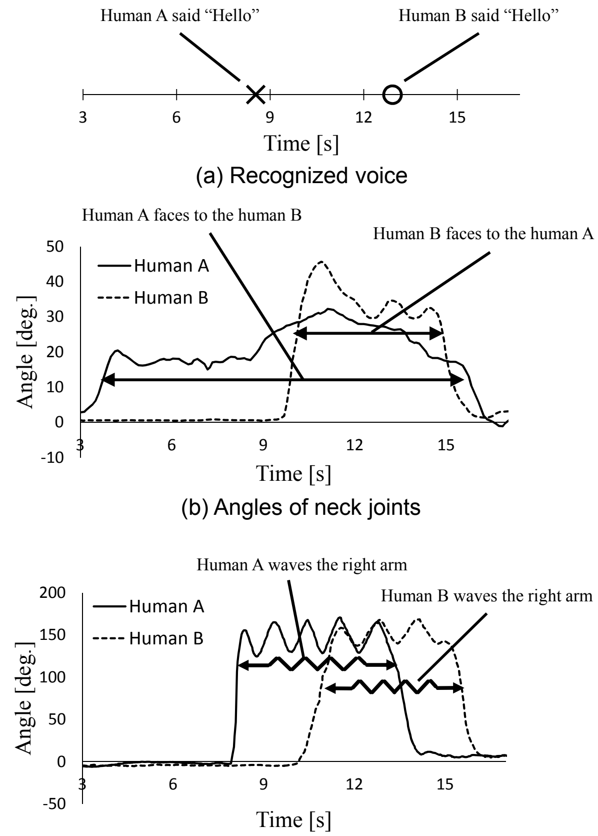
service. These captured data are sent to the server PC on the network through the client PC in order to compute between two avatars in a VR space.

Fig. 11 shows the scene of a greeting between two persons in a VR space. The human A and B shown in (a) and (b) log in to the VR space shown in (c) via the network, then two avatars greet with each other. This greeting is computed in the server PC by using measured data sent from two client PCs in place A and B. As experimental conditions, two avatars face each other. The avatar B is placed to the right side of the front of the avatar A.

C. Experimental Result

Fig. 12(a) and (b) show behavior-databases of the human A and B obtained from the experiment of a greeting. Each table has collected data for the voice, the neck joint angle, and quaternions for the right shoulder joint, the right elbow joint, and the body of each human. These data were selected from the behavior database to analyze correspondence relations between the human A and B in the greeting.

Fig. 13 shows the relationship of behaviors between the human A and B in the greeting. Graphs were plotted from the behavior databases shown in Fig. 12(a) and (b). The horizontal axis of each graph shows time. The graph of (a) shows the recognized voices for two persons A and B in a greeting. As shown in the graph, the human A said "Hello" at the time of about 9 [s]. The graph of (b) shows the angles of neck joints for two persons. The vertical axis of the graph in (b) shows the neck-joint angle on yaw axis. The angle of the neck joint is set as 0 degree from the front of the avatar and right rotation is positive direction. In the graph



(a) Recognized voice

(b) Angles of neck joints

(c) Angles of right-shoulder joints

Fig. 13. Graphs of human behaviors in an observing experiment

of (b), the solid line shows that the human A faced to the human B during about 3 [s] to 16 [s]. The dashed line also shows that the human B faced to the human A during about 9 [s] to 15 [s]. From the graphs of (a) and (b), we can see that the human B opportunely started facing to the human A after the human A said "Hello". In the graph of (c), the vertical axis shows the angle of right shoulder on roll axis. Angles of right shoulders on horizontal arm are set as 0 degree and lower rotation is positive direction. In the graph of (c), the solid line shows that the human A waved the right arm during about 8 [s] to 14 [s]. The dashed line of shows that the human B waved the right arm during about 10 [s] to 16 [s]. From the graphs of (b) and (c), we can see that the human B opportunely started waving the right arm after the human A waved the right arm.

V. CONCLUSIONS AND FUTURE WORKS

We proposed a cloud VR system with immersive interfaces to collect human gaze-controls and interpersonal-behaviors. In the proposed VR system, multiple users can log in to a VR space and naturally interact with each other and a robot by using immersive interfaces. Human gaze-controls and interpersonal-behaviors were collected to database while human beings and a robot interact with each other. Two application experiments were performed to evaluate the effectiveness of the proposed VR system with immersive interfaces. In the learning experiment for object attributes,

Time [s]	Voice	Neck Joint Angle[rad.]			Right Shoulder Joint Quaternion			
		yaw	pitch	roll	w	x	y	z
8.5	Hello	-0.314	0.039	0.027	0.861	0.000	0.109	-0.497
9.0		-0.393	0.010	0.027	0.914	0.000	0.123	-0.387
9.5		-0.456	0.005	0.042	0.809	0.000	0.110	-0.578
10.0		-0.489	0.031	0.031	0.927	0.000	0.152	-0.343
10.5		-0.521	0.032	0.041	0.780	0.000	0.096	-0.618
11.0		-0.539	0.042	0.050	0.934	0.000	0.152	-0.323
11.5		-0.546	0.032	0.030	0.764	0.000	0.100	-0.637
12.0		-0.507	0.036	0.043	0.907	0.000	0.135	-0.399
12.5		-0.486	0.014	0.016	0.873	0.000	0.106	-0.476
13.0		-0.476	0.016	-0.016	0.855	0.000	0.092	-0.510

(a) Human A

Time [s]	Voice	Neck Joint Angle[rad.]			Right Shoulder Joint Quaternion			
		yaw	pitch	roll	w	x	y	z
8.5		-0.011	-0.018	0.014	0.682	0.000	-0.004	0.731
9.0		-0.009	-0.019	0.008	0.680	0.000	0.004	0.733
9.5		-0.014	-0.024	0.007	0.677	0.000	0.004	0.736
10.0		-0.309	0.029	0.007	0.687	0.000	0.000	0.727
10.5		-0.700	0.070	-0.019	0.854	0.000	0.090	0.513
11.0		-0.794	0.083	-0.040	0.995	0.000	0.042	-0.091
11.5		-0.669	0.067	-0.034	0.819	0.000	0.281	-0.500
12.0		-0.608	0.044	-0.046	0.908	0.000	0.033	-0.418
12.5		-0.519	0.036	-0.039	0.845	0.000	0.131	-0.518
13.0	Hello	-0.558	0.024	-0.021	0.809	0.000	0.136	-0.572

(b) Human B

Fig. 12. Collected human-behaviors in an observing experiment

a robot in a VR space learned object attributes from human beings through an avatar. Human could observe three objects with immersive visualization and motion-controls by using Oculus Rift and Xtion. In the observing experiment for greetings, a robot in a VR space observed greetings by two persons who are in remote places. Human gaze-controls, entire-body motions, and voice were collected to the behavior database by using immersive interfaces. The results of these application experiments to learn object attributes and to observe greetings demonstrate the effectiveness of the proposed VR system with immersive interfaces.

As future works, we will try to collect the huge number of multimodal informations from experiments with human-robot interactions in several experimental settings. For example, experimental settings for different distances between two persons and room arrangements will be performed in VR spaces. Also we will try to perform application experiments to collect to high-level human-robot interactions by using several other immersive-interfaces: a hand motion controller, finger motion controller, and 3D haptic device.

VI. ACKNOWLEDGMENTS

A part of this work was supported by JSPS KAKENHI Grant Number 23300077.

REFERENCES

[1] F. Yamaoka, T. Kanda, H. Ishiguro and N. Hagita, "Developing a Model of Robot Behavior to Identify and Appropriately Respond to Implicit Attention-Shifting", *in Proc. of 4th ACM/IEEE International Conference on Human-Robot Interaction*, 2009, pp. 133-140.

[2] Y. Yamakata, T. Kawahara, H. G. Okuno, and M. Minoh, "Belief Network based Disambiguation of Object Reference in Spoken Dialogue System", *in Trans. of the Japanese Society for Artificial Intelligence*, Vol.19, 2004, No.1, pp.47-56.

[3] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels, S. Iwasawa, K. Mase, K. Kogure, and N. Hagita, "Collaborative capturing, interpreting, and sharing of experiences", *in Journal of Personal and Ubiquitous Computing*, Vol. 11, 2007, Issue 4, pp. 265-271.

[4] K. Sugiura, N. Iwahashi, H. Kashioka, and S. Nakamura, "Learning, Generation, and Recognition of Motions by Reference-Point-Dependent Probabilistic Models", *in Journal of Advanced Robotics*, Vol. 25, 2011, No. 6-7, pp. 825-848.

[5] T. Nakamura, K. Sugiura, T. Nagai, N. Iwahashi, H. Okada, and T. Omori"Learning Novel Objects for Extended Mobile Manipulation", *in Journal of Intelligent and Robotic Systems*, Vol. 66, 2012, Issue 1-2, pp. 187-204.

[6] K. Abe, A. Iwasaki, T. Nakamura, T. Nagai, A. Yokoyama, T. Shimotomai, H. Okada and T. Omori, "Playmate Robots That Can Act According to Child's Mental State", *in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4660-4667.

[7] H. Kamide, M. Yasumoto, Y. Mae, K. Ohara, T. Takubo, and T. Arai, "Comparative Evaluation of Virtual and Real Humanoid with Robot-Oriented Psychology Scale", *in Proc. of IEEE International Conference on Robotics and Automation*, 2011, pp. 599-604.

[8] A. Weiss, R. Buchner, H. Fischer and M. Tscheligi, "Exploring Human-Robot Cooperation Possibilities for Semiconductor Manufacturing", *in Proc. of Collaboration Technologies and Systems*, 2011, pp. 173-177.

[9] G. Bosc, M. Kaytoue, C. Raissi, J. Boulicaut, and P. Tan, "Mining Balanced Sequential Patterns in RTS Games", *in Proc. of the European Conference on Artificial Intelligence*, 2014, pp. 975-976.

[10] T. Inamura et al, "Simulator platform that enables social interaction simulation -SIGVerse: SocioIntelliGenesis simulator-", *in Proc. of IEEE/SICE International Symposium on System Integration*, 2010, pp. 212-217.