2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)
(Invited Paper)

Cloud Based VR System with Immersive Interfaces to Collect Multimodal Data in Human-Robot Interaction

Yoshinobu Hagiwara
Dept. of Human & Computer Intelligence
Ritsumeikan University
Shiga, JAPAN
yhagiwara@em.ci.ritsumei.ac.jp

***Abstract*—This paper presents a cloud based VR system with immersive interfaces to collect multimodal data in human-robot interaction and its applications. The proposed system enables a subject to log in to the VR space as an avatar and to naturally interact with a virtual robot by immersive interfaces. A head mounted display and a motion capture device provide immersive visualization and natural motion control in the VR system, respectively. The proposed VR system can simultaneously perform natural human-robot interaction in a VR space and collect visual, physical, and voice data during human-robot interaction by the immersive interfaces. Two application experiments to learn object's attributes and to learn communication protocol demonstrate the availability of the proposed system.**

Keywords—cloud robotics; multimodal data; human-robot interaction; virtual reality

I. Introduction

Recently, various types of human-support robots have been developed by companies and research institutes. These robots have to provide physical and mental supports to a person by natural human-robot interaction. In order to realize natural human-robot interaction, Yamaoka et al. have proposed a method that a robot obtains communication protocol by analyzing human behaviors when a person explains about exhibits to another person [1]. Yamakata et al. have also proposed a method for disambiguation of object reference by learning object's attributes from human beings [2]. However, in these studies, the number of experiments based on human-robot interaction was limited within about ten subjects, because of the high cost for the construction of experimental environment and the maintenance of actual robots as shown in Fig. 1.

As a solution of this problem, the use of a simulator might reduce the cost of experiments for human-robot interaction. Kamide et al. have performed experiments based on human-robot interactions using a cave VR system which provides realistic visualization in order to design safe robot's motions [3]. Weiss et al. have also performed human-robot co-operations in a VR space in order to observe human emotions [4]. However, it is difficult to directly apply these VR systems to cloud based VR system because these systems need to use special and large-size equipment.

Learning object concept   Learning communication protocol

Fig. 1. Interaction experiments between human and robot.

In order to solve the problem, authors have proposed a cloud based VR system with immersive interfaces such as Kinect sensor and Oculus Rift [5]. The proposed system can not only perform experiments based on human-robot interaction in a VR space but also collect multi-modal data such as voice, visual and physical data. In this study, we performed two application experiments to learn object's attributes and to learn communication protocol. The performance of the proposed system to collect multi-modal data in human-robot interaction was evaluated by the experiments.

II. Cloud based VR system

The proposed system was developed based on the simulator called SIGVerse [6]. Fig. 2 shows the overview of the proposed system. The system enables multiple subjects can log-in to a VR space as avatars from remote places and naturally interact with a virtual robot. By using the proposed system, human-robot interactions can be effectively performed in a VR space. The feature of the proposed system is the use of immersive interfaces that enable immersive visualization and natural body-motion control in a VR space. In addition, these immersive interfaces collect human gaze and body motion behaviors during human-robot interaction.

Oculus Rift, Xtion, PlayStation Move, and Leap Motion are able to use as 3D headset, body motion sensor, arm motion sensor, and finger motion sensor, respectively. Low price of these devices makes many users can attend experiments with human-robot interaction using the proposed system. Plug-in software to use these immersive interfaces in the proposed system can be downloaded from the following URL: http://www.sigverse.com/wiki/en/.

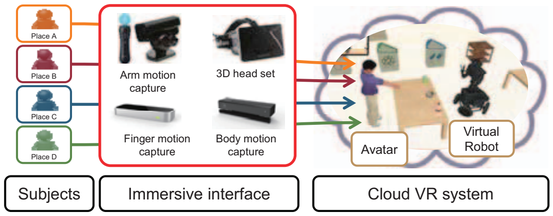


Fig. 2. Overview of the proposed system.

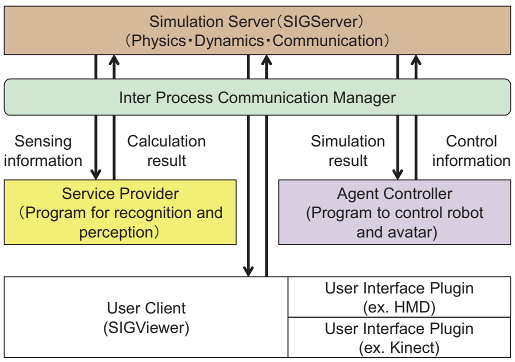


Fig. 3. System construction of SIGVerse.

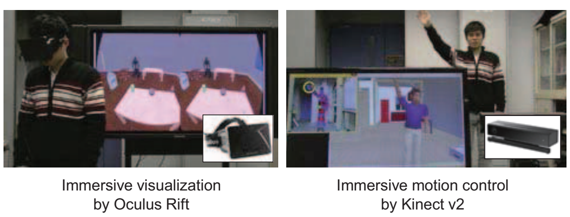


Fig. 4. Immersive interfaces in the proposed system.

Fig. 3 shows the system construction of SIGVerse. SIGVerse consists of a simulation server called SIGServer, agent controller, service provider and a user client called SIGViewer. SIGServer provides physics, dynamics and communication simulation. Agent controller is the program to control the motion of a robot and an avatar in the VR space. Service provider is program for perception and recognition for a robot. Software libraries such as Open CV and Julius can be used in service provider. Multiple users can access to SIGServer via SIGViewer with user interfaces such as a head mounted display and a Kinect senor. The system can simultaneously perform human-robot interaction and collect human behaviors during the human- robot interaction.

Fig. 4 shows immersive interfaces used in the proposed system. Oculus Rift provides immersive 3D visualization and collects human face direction. Kinect for windows v2 provides natural body motion controls in a VR space and collects human skeleton model and joint angles such as a neck and shoulders.

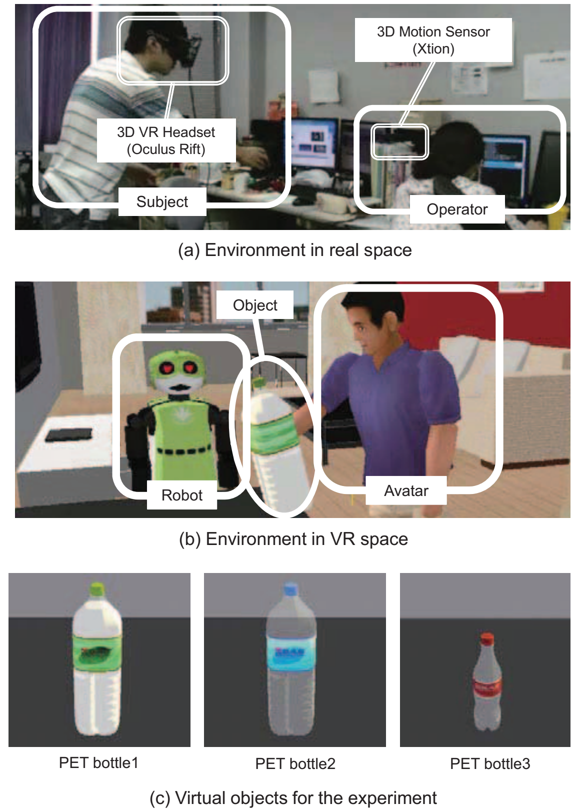


Fig. 5. Experimental environment to learn object's attributes.

III. Application Experiments

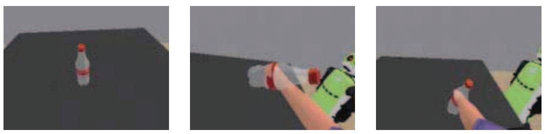
Two application experiments were performed in order to confirm what kind of data can be collected by the proposed system and whether the collected multi-modal data is effective for analyzing human behaviors. As the first experiment, a robot learns object's attributes from subjects by human-robot interaction. As the second experiment, a robot learns a greeting protocol by analyzing human behaviors from the greeting by two persons.

A. Experiment to learn object's attributes

As one of applications using the proposed system, we performed an application experiment that a robot learns object's attributes from subjects in a VR space. Fig. 5 (a) and (b) show experimental setups in a real space and a VR space, respectively. Fig. 5 (c) shows virtual objects used for the experiment. In the real space, the subject wore a 3D headset device called Oculus Rift which was used as immersive visualization and to collect human face direction. A motion capture device called Xtion was used as natural body motion controls and to capture the entire-body motions of the subject. The operator inputs a text message instead of voice recognition.

TABLE I. COLLECTED OBJECT'S ATTRIBUTES

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



Name: Coca Cola | Color: Red | Shape: Round Shape

(a) Memorized views with words

How to use: drink

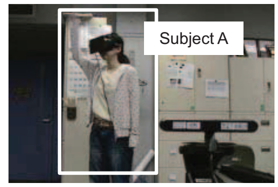
(b) Memorized motion with a word

Fig. 6. Collected human views and motion with words.

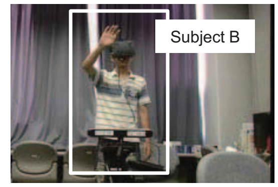
The both of a robot and an avatar were placed around the table in the VR space. The subject can observe an object from different angles by grasping and manipulating the object via the avatar. The subject also teaches how to use the object by ones gesture via the Xtion sensor. In the experimental setup, object's attributes of three virtual objects in Fig. 5 (c) were collected from two subjects A and B by the proposed system. In the experiments, the virtual robot asks four questions about object's attributes. The questions were "Name of the object", "Color of the object", "Shape of the object" and "How to use the object". Then the subjects answer the questions by their voice and motion.

Table 1 shows the collected object attributes by using the proposed system. The proposed system can collect not only words but also views and motions with these words.

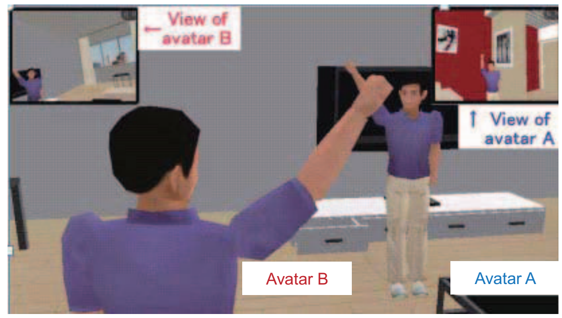
Fig. 6 shows the collected views and a motion at boxed words in Table 1. The three pictures in Fig. 6 (a) show the collected views when the subject A said "Coca cola", "Red", and "Round shape", respectively. In the picture of "Round shape", the robot can understand that the word was given to the view from the top side of the object. Fig. 6 (b) shows the collected motion-data when the subject said "Drink". The collected body motion data is useful information when the robot understands human body language.



(a) Place A in a real space



(b) Place B in a real space



(c) Greeting between two persons in a VR space

Fig. 7. Experimental environment to learn greeting protocol.

B. Experiment to learn communication protocol

As another application using the proposed system, we performed an application experiment to learn communication protocol from a greeting between two subjects in a VR space. Fig. 7 shows the experimental environment to learn communication protocol. In the experiment, subject A and B log in to a VR space from remote places as shown in Fig. 7 (a) and (b). Oculus Rift, Xtion and microphone were used to capture human behaviors in the real space. In the VR space, another avatar was place to the right side of their body front as shown in Fig. 7 (c). They performed a greeting via their avatar A and B in the VR space. In the greeting, they found another avatar and wave their hand and said "Hello".

In the experiment, real time processing within 100ms and natural avatars motion can be performed. Fig. 8 shows the collected voice and body motion data from two subjects by the proposed system. We focused on three data to analyze human behaviors. The graph (a) shows the recognized voice by the voice recognition software called Julius. The graph (b) and (c) show the angle of neck joint and the angle of right-shoulder joint angles which were captured by the Xtion sensor. The horizontal axis and vertical axis show time and the angle of each joint, respectively. The solid and dashed lines show the collected data of subject A and B. The timing to say "Hello" was shown in the graph (a). The motions to find another avatar and to wave their hand were shown in the graphs (b) and (c), respectively.

Fig. 9 shows the analysis result of the greeting protocol form the collected data shown in Fig. 8. The timing to start and finish facing another avatar and to start and finish waving their hand and to say "Hello" were extracted by thresholds for each joint.

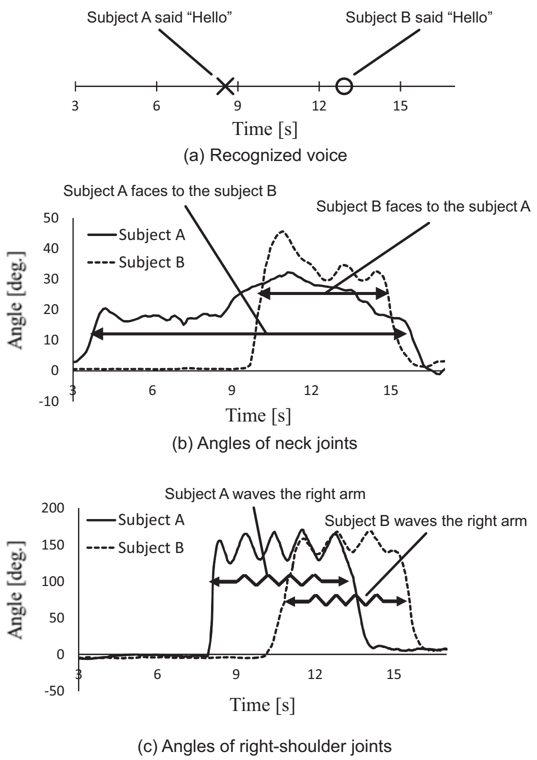


Fig. 8. Collected human behaviors in the greeting experiment.

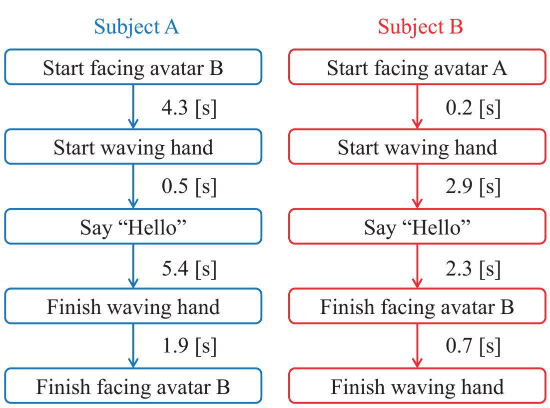


Fig. 9. Analysis result of greeting protocol.

These sequences will be useful to analyze the greeting protocol and to generate a robot's behavior model for greeting. Table 2 shows the detail parameters of the hand waving motion of subjects A and B. Motion time, period, and angular width were extracted by simple functions.

TABLE II. PARAMETERS TO WAVE HAND

Subject	Time [s]	Period [s]	Angular width [°]
A	6.0	1.2	36.9
B	5.5	1.8	24.6

It is expected that these parameters are effective in generating natural hand waving motion for robots. The proposed system can collect these parameters to generate communication protocols and motions from many subjects in remote places via a VR space.

IV. CONCLUDING REMARKS

In this study, we proposed a cloud based VR system with immersive interfaces to collect multi-modal data in human-robot interaction and its applications. Two application experiments to lean object attributes and to learn communication protocols demonstrated that the proposed system can collect multi-modal data and the collected data are effective for analyzing human behaviors to realize natural human-robot interaction. In the experiments, the learning by one or two persons was performed as the first step of this study. The proposed system enables a robot to learn knowledge to realize natural human-robot interaction from multiple subjects via a VR space. As future works, we will try to perform interaction experiments to collect the huge number of multi-modal data from many subjects and the behavior model generation by the analysis of the collected multi-modal data.

ACKNOWLEDGMENT

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

REFERENCES

[1] F. Yamaoka et al. “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 et al. “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] H. Kamide et al. “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.

[4] A. Weiss et al. “Exploring Human-Robot Cooperation Possibilities for Semiconductor Manufacturing”, in Proc. of Collaboration Technologies and Systems, 2011, pp. 173-177.

[5] Y. Hagiwara et al. “Cloud VR System with Immersive Interfaces to Collect Human Gaze-controls and Interpersonal-behaviors”, in Proc. of IEEE/SICE International Symposium of System Integration, 2014, pp.169-174.

[6] 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.