Initial Design on VR-based Data Acquisition Environments for Human-Robot Interaction: Scenarios and Functional Requirements

Bok Cha[1,2] and Joo-Haeng Lee[2,1],

***Abstract*— Data-driven approach is prevailing in various robotics research fields with advancement of deep learning. However, data related to human-robot interaction are still quite insufficient both in domain diversity and data volume itself. In this paper, we introduce our initial design ideas on a VR-based data acquisition environments for human-robot interaction, specially focusing on scenarios and relevant functional requirements. For example, a human performer who wears a VR HMD and trackers can have a role of an elderly person or a robot. Then, he can interact with an object or another character in a virtual environment, which is driven by a game engine such as Unreal 4. In this data acquisition system, all the behavioral data as well as visual and contextual information can be stored and transformed for target learning system. We define the related interaction types and data organization from the perspective of individual scenarios derived. We believe this design ideas will guide us for on-going research and current development.**

I. INTRODUCTION

Recently, with the technological development of computer graphics and VR devices, the quality of synthetic imagery and motions came to be very realistic, which accelerates their adoption in deep learning as a synthetic dataset. The virtual learning data has an advantage in that the system configuration such as data acquisition environment, pipeline, and data format can be flexibly defined according to data requirements in machine learning. In addition, such a system is relatively free from physical constraints and significantly lowers a cost barrier.

Especially, collecting deep learning data for a human-robot interaction domain imposes various difficulties for constructing a complex acquisition environment in reality. Costs for real robots and recruiting multiple experimenters should be seriously considered. In addition, even after acquiring a dataset following a precise plan in such a physical environment, data amendment and extra acquisition are almost mandatory in the reality of deep learning, which is hard to be re-considered after one physical acquisition session has finished.

We believe that such limitations of acquiring human-robot interaction data in a real-world can be solved using a virtual environments based on a game engines and VR technologies. VR allows a person to have a immersive experience in a simulated virtual environment. Hence, once a virtual environment is properly constructed as a digital twin of a real-world target domain, it would be possible to generate various types of human behavioral data towards robots, and vice versa, even when objects are engaged in the scenario.

[1]Bok Cha is with the University of Science and Technology, Daejeon, Republic of Korea. bokc@etri.re.kr

[2]Joo-Haeng Lee is with Human Robot Interaction Lab, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea. joohaeng@etri.re.kr. Corresponding author.

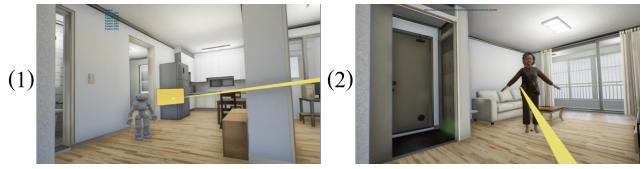

Fig. 1. Exchanging views using VR controller and Unreal Engine. The background environment is the living lab Asset by KIST [1], which was generated similarly with real apartment used by capturing dataset of human behavior. Figure (1) is the view of person. And (2) is the robot view exchanged by the hand controller inputs.

In this paper, we introduce our initial design ideas on a VR-based data acquisition environments for human-robot interaction, specially focusing on scenarios and relevant functional requirements. For example, a human performer who wears a VR HMD and trackers can have a role of an elderly person or a robot. Then, she/he can interact with another character and/or objects in a virtual environment, which can be efficiently driven by a modern game engine such as Unreal 4. In this data acquisition system, all the behavioral data as well as visual and contextual information can be simultaneously stored, and further transformed for a target learning system. We define the related interaction types and data organization from the perspective of individual scenarios derived.

II. RELATED WORK

Research on human-robot interaction is being attempted to show that it is possible to experience a dynamic experience in a virtual environment while wearing VR. For this reason, research has been conducted to train a real robot in a virtual environment while wearing VR for beginners who are first at manipulating industrial robots. [2] On the other hand, research and attempts to learn the behavior of a virtual robot by using VR and planning it in a virtual environment by teleoperating it with a real robot [2] and an experimenter who wears VR and interacts with a virtual robot to become a human-robot Even, It is being tried as a way of evaluating interactions. [3]

In addition, you can wear a VR and act as a virtual robot or person to experience the virtual environment in a vivid way and naturally direct the movement of the human body. Research [4],[5] in which an experimenter wears VR and becomes a robot in a virtual environment to control a real robot through teleporation, and a technology that can create

an avatar that follows the player's movement as a character in a virtual environment[6] was recently studied. Fig. 1 is the initial result for the interaction scenarios, which contains the exchange of view in virtual environment by using VR controller inputs.

A study on a platform [7] that can acquire high quality data in quantitative and qualitatively by deep learning data in a virtual environment, and research to enhance and improve performance of deep learning data using virtual data. [8]

In this paper, we propose an environment and scenario to generate data by creating natural behaviors and interactions while experiencing a virtual environment like reality.

III. System Overview

In this section, we describe the system overview while focusing on the target scenarios. Details of each scenario are explained at Table 2.

A. System Architecture

The VR-based data acquisition environment is implemented as a Blueprint using the Unreal Engine 4 game engine as described in Table 1 and utilizing the UE4 basic blueprint API and custom API for the virtual environment. It is a structure that allows the player to wear HTC Vive Head Mount Display(HMD) and Tracker, enter the virtual environment implemented, interact as in reality, and generate data by creating actions

Figure 2 has the flow of whole system from player inputs with VR tracker to deep learning dataset, through the blueprint developments on Unreal game engine.

TABLE I
System Environment

Items	Env.
OS	Windows 10
Game Engine	Unreal Engine 4.25
Hardware	HTC Vive HMD, 2 Hand Controllers, 3 Vive Trackers
Languages	UE4 Blueprint
API	UE4 Blueprint API, Mocap Custom Plugin
Third Party	UnrealCV
Data format	PNG, FBX, BVH

TABLE II
Scenario-Centric Summary

Scenario Class	Interaction Types	Functional Requirements	Data type
S1.	Human Behavior	S1.FR=[Possession, GUI, data saving]	S1.Data=[RGBD, mask, motion data]
S2.	Human-Object interaction	Interacting with object and S1.FR	Human-Object interaction pair data and S1.Data
S3.	Human-Robot interaction	Robot motion simulation, human reaction and S1.FR	Human-Robot interaction pair data and S1.Data
S4.	Human-Human interaction	Dual VR approach module and S1.FR	Human-Human interaction pair data and S1.Data

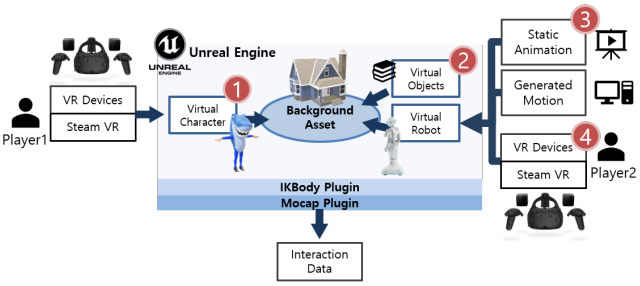

Fig. 2. System architecture and scenarios for proposed VR-based data acquisition environments: (1) the first scenario is the case of the behavior of one player who wear the VR devices, (2) the interaction scenario with virtual objects. (3) the scenario with the motion of virtual robot which are simulated by the sequences of the recorded animation or the generated motion from deep learning model, and (4) the last scenario about the duo-player interaction.

B. Interaction Scenarios

The following four scenarios are defined according to the type of interaction. Based on Table 2 below, scenarios, interactions, functional specifications, and data types are described.

- S1. Human Behavior Data
  Players wearing VR and trackers move freely in a virtual environment as they move in response to human characters in a virtual environment and generate daily behavior data.
- S2. Human-Object Interaction Data
  The player becomes a virtual character, interacts with objects in the virtual environment, touches, moves, etc. and stores these interaction data.
- S3. Human-Robot Interaction Data
  The player becomes a virtual character and interacts with a virtual robot that simulates predefined actions to generate data. The predefined movement of the virtual robot at this time is a result of simulating behavior data generated by static animation or learning. The player sees the behavior of the virtual robot, expresses the natural reaction, and stores it as data.
- S4. 2-VR Interaction Data
  Using two VR devices, two players connect to the same virtual environment, discuss with each other a human character and a robot character, create a situation or exchange responses to each other's actions, and store it as a pair of interactive data.

C. Functional Requirements

This section describes in detail the functions required for the scenario defined above. As the FR section of Table 2 describes the functional requirement for each scenario. the first human behavior scenario is the base of functional requirements for the other scenarios. And Figure 3 have the details about implementing pipeline of the system applying custom plugin and motion retargeting method.

- Human Behavior (S1. FR)

Functional requirement of S1. has three main implementation of the possession, VR-based Graphic User Interface and data achieving environment. The first possession is implementation of point of view switching function and Including motion retargeting function in the data acquisition environment for immersive experience to player. Second, VR-based GUI implementing has conducting with VR hand controller to manipulate the graphic interface, such as UMG menu. VR-based GUI could control and set by using hand controller input in virtual environment, and the most important thing is connecting assets and processing each modules on UE4 engine by graphic interface. Last, Data achieving environment cover the many parts of GUI Blueprint development. Moreover, the environment could save the data files as needed each learning. The details of dataset are explained below section.

- Human-Object Interaction(S2. FR)
  The functions of human behavior are based of human behavior's functional requirement(S1. FR). After than, Human-Object Interaction function has additional behavioral function of human with the virtual objects.
- Human-Robot Interaction (S3. FR) Describing the functions of human-robot interaction is based on human behavior's functional requirement(S1. FR) as well. The additional functions are about the reaction of player with virtual robot animation. the robot motion can be two types, the simulated static animation and the simulated robot motion which is the acquired results from learning.
- Human-Human Interaction (S4. FR) Last functional requirement is the usage of dual VR devices simultaneously. The base functions same with other FR details as S1. FR. Human-Human interaction need the function of allowing the dual VR HMD and hand controllers input as two virtual characters.

D. Organization of Dataset

Basically, the format of data is composed of various types of images such RGB, depth, and mask information through a virtual camera from a VR wearer's point of view, a virtual character or a robot's point of view, and an arbitrary position. In addition, it was designed to output motion data of the person wearing VR and tracker in BVH or FBX form by implementing the motion capture function and to process it into the data form of the input of the learning model.

The form of the interaction data is motion data of BVH or FBX files. The types are (1) the experimenter's interaction with a virtual object, and (2) the experimenter's behavior for a predefined character's behavior, (3) It can be expected that there are three actions that two experimenters interact with each other and create in a virtual environment.

The below S1.Data is the base of other scenario data.

- Image sequences (RGB, depth, mask) data
- Motion data (FBX/BVH)

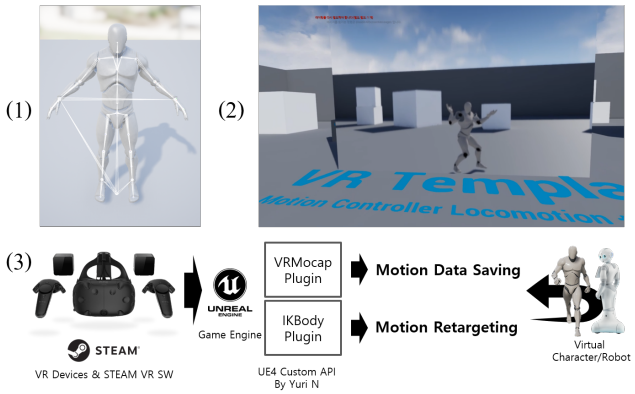

Fig. 3. Implementation idea: (1) Virtual character and robot should have our skeleton rig. The skeleton structure is assumed to the Unreal default. (2) A screen shot of motion retargetting using custom plugin IKBody. (3) The process of how motion data are generated from VR devices to data files.

E. Scenario-Centric Summary

The main idea in this paper is explained with Table 2, the summary focused on the scenario. Each scenario are expressed as class, S1, S2, S3 and S4. Scenario 1 is about human behavior and it's functional requirements(S1.FR) are the union, possession, GUI, data saving. This S1.FR reuse repeatedly for other scenario functional requirements. Data of scenario 1 are the image sequences and motion data. And this data(S1.Data) are reused for other scenario class. Human-Object interaction scenario has the function of interacting with object and S1.FR. About data, Human-object interaction pair data and S1.Data are contained. S3. Human-Robot interaction scenario are centered with human reacting to robot motion. The functional requirement and data have the contents of interaction pairs. S4. scenario additionally has the function using the dual VR devices and the data interacting pair data with two player.

IV. Conclusion and Future Work

In this paper, we outlined initial design ideas on VR-based data acquisition environments for human-robot interaction, specially focusing on the system overview, scenarios, and relevant functional requirements. The actual implementation of the acquisition environment for the proposed scenarios is in progress to boost the intelligence of an elderly-supporting robot with deep learning technology.

ACKNOWLEDGMENT

This work was supported by the Institute of Information communications Technology Planning Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2017-0-00162, Development of Human-care Robot Technology for Aging Society)

References

[1] Park, G.W., Cho, J. and Kim, I.J., 2019. A Synthetic Data Creation Platform for Daily Activity Recognition of Elderly People.

[2] Pérez, L., Diez, E., Usamentiaga, R. and García, D.F., 2019. Industrial robot control and operator training using virtual reality interfaces. Computers in Industry, 109, pp.114-120.

Savkin, P., Quinn, N. and Wilson, L., 2019. Spooky action at a distance: real-time VR interaction for non real-time remote robotics. In ACM SIGGRAPH 2019 Real-Time Live! (pp. 1-1).

[3] Duguleana, M., Barbuceanu, F.G. and Mogan, G., 2011, July. Evaluating human-robot interaction during a manipulation experiment conducted in immersive virtual reality. In International Conference on Virtual and Mixed Reality (pp. 164-173). Springer, Berlin, Heidelberg.

[4] Lipton, J.I., Fay, A.J. and Rus, D., 2017. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. IEEE Robotics and Automation Letters, 3(1), pp.179-186.

[5] Whitney, D., Rosen, E., Ullman, D., Phillips, E. and Tellex, S., 2018, October. Ros reality: A virtual reality framework using consumer-grade hardware for ros-enabled robots. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1-9). IEEE.

[6] Caserman, P., Achenbach, P. and Göbel, S., 2019, August. Analysis of Inverse Kinematics Solutions for Full-Body Reconstruction in Virtual Reality. In 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH) (pp. 1-8). IEEE.

[7] Garcia-Garcia, A., Martinez-Gonzalez, P., Oprea, S., Castro-Vargas, J.A., Orts-Escolano, S., Garcia-Rodriguez, J. and Jover-Alvarez, A., 2018, October. The robotrix: An extremely photorealistic and very-large-scale indoor dataset of sequences with robot trajectories and interactions. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 6790-6797). IEEE.

[8] Ros, G., Sellart, L., Materzynska, J., Vazquez, D. and Lopez, A.M., 2016. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3234-3243).