Agricultural Training System with Gazing-Point Detection Function using Head-Mounted Display: HTC Vive Pro Eye and Virtual Reality–based Unity System

Shinji Kawakura and Ryosuke Shibasaki

ABSTRACT

In this study, we develop a virtual reality (VR)–based agricultural (hereinafter, agri-) support technique to assist in training newcomers and trainees in agricultural work. The system consists of a head-mounted display (HMD)—a HTC Vive Pro Eye—gaming personal computer, and peripheral components. The HMD-based system consists of the following: (1) a precision eye-tracking system for tracking and interpreting eye movements to enable lifelike interactions, better manage GPU workload, and simplify input and navigation; (2) a room area tracker up to 100 m²; and (3) a VR-space experience with unmatched tracking accuracy with SteamVR™ Tracking. The agri-situation considered in this study is non-specific tomato harvesting performed by inexperienced agri-workers. The study aims to (1) measure and analyze persons’ cognition and behavior indicators in VR-based environments that simulate the work site, (2) provide suggestions regarding analyses of processes from the cognition of targets to specific behaviors using objective indexes; and (3) realize in advance the verification and comparison of improvements of measures related to specific agri-works without the need to utilize actual agri-work sites. Specifically, we utilize and apply the eye-tracking function incorporated in the HMD, in addition, we develop a Unity-based VR system with sound notification to indicate the validity of eye-tracking and the motion of manual agri-workers and managers. Subsequently, we conduct experimental trials in a non-specific room. In the trials, subjects wearing the HMD system sequentially gaze at figures of small red tomatoes in the VR spaces. In the process, the VR system plays alerts to notify subjects when they gaze at or miss the target tomatoes. The system provides quick and accurate operations, and the eye-tracking function of the system differs from existing agri-training-oriented techniques and products. The system has several advantages such as lower cost compared with existing similar mechanical systems found in the literature and similar commercial products. In addition, the Unity-based system is a minimal and flexibly scalable system that can be adjusted to suit future studies and expansion for different agri-situations. This study consists of five phases: (1) designing and confirming the validity of the system; (2) accumulating image data on outdoor farmland; (3) constructing the entire system and tuning various minor system settings like program parameters and other specifications; (4) executing experiments in an indoor room; and (5) assessing and discussing the results and gathering comments from the subjects and presenting the characteristics of these trials. We consider that, from the limited trials, the system can be judged to be valid to some extent in certain situations. However, we could not perform broader or more generalizable experiments using the system. We present experimental characteristics and numerical ranges related to the trials, particularly noting speed and likelihood of mistakes concerning the system’s practical operations. The novel achievements of this study lie in the fusion of the latest HMD and Unity-based agricultural training facilities. In future, agri-workers and managers can use the proposed system for training, particularly for eye movement. Furthermore, we believe that, by combining this system with other existing systems, agriculture can be greatly improved.

Keywords: agricultural training system, eye tracking, point of gaze, VR.

I. INTRODUCTION

Training new workers in the agricultural (hereinafter, agri-) field is an important aspect of practical agri-management sites. Agri-business entrepreneurs and agri-workers have developed a wide range of attractive and useful systems and techniques (hereinafter, tech) for inexperienced agri-workers and trainees [1], [2], which improve productivity and reduce costs along with other advantages. These promising and effective systems have continued to
evolve with the combination of VR game-fields and agri-engineering.

Furthermore, recent studies in agri-research and -business have brought about the development of virtual reality (VR), augmented reality, substitutional reality, mixed reality (MR), and cross or extended reality (XR) systems [3]–[8]. Many of these employ the latest VR tech-based trends in agriculture with a focus on human-engineering (kinematics) and other human-based sciences and technologies.

Several other closely related studies have targeted high-tech VR systems for scientific research and agri-business management [9]–[23].

Considering the aforementioned literature, technical backgrounds, and trends, in this study, we utilize the HTC Vive Pro Eye VR system (HTC Corporation Inc., Xindian, New Taipei, Taiwan, with technology by Valve). HTC and Facebook Inc. (California, USA) are the latest vendors of VR infrastructural sets that are strong candidates for the execution of diverse VR worlds and the collection of useful, meaningful data concerning objects and fields in a VR world.

We designed and constructed an original VR-based system with an eye movement tracking function for training agri-workers. The peripheral components such as base stations and controllers are included. Specifically, we developed and attempted to implement an eye-tracking-based gazing-point detection system utilizing visual records for traditional agri-workers and managers.

The proposed VR system has the following advantages: (1) Its cost is lower compared with past similar head-mounted-display (HMD) VR-based systems developed by academic researchers and manufacturers. (2) It is video-game-based and easily identifies and learns eye-tracking during harvest. (3) The development platforms Unity (Unity Technologies, San Francisco, USA) and Microsoft Visual C# are flexibly programmable and changeable. Furthermore, the standard Vive Pro Eye is the minimal set required for the product, which allows the users to program, arbitrarily expand, and handle diverse situations.

II. METHOD

A. Overview

This study consists of five phases:
1) designing and confirming VR simulation-based training systems and gazing-point tracking systems presented in Fig. 1;
2) constructing Unity 2019– and Visual C#–based programs to simulate a real outdoor farm (Fig. 2 and 3), and tuning the system parameters;
3) conducting experiments to verify the developed system;
4) observing and estimating the system’s performance, accumulating data concerning accurate responses;
5) discussing the data and the scope for future studies.

We verified which of the components of the system are essential for the operation of harvesting tomatoes. As shown in Fig. 4–7, to achieve the aim of this study, the system requires two fundamental functions; for the gaming time of the VR spaces, the initial components run sequentially; then, the subjects’ gaze starts and moves from position A (the start position) to B (the goal position), as indicated in Fig. 6.

B. Outline of VR-based System

In this study, we utilized and modified a rather recent VR kitset with an eye-tracking function. No agri-research has previously developed or utilized the system, although some studies have been undertaken into training methodology [9]–[23]. The fundamental components are the VR API sets provided by MSI Inc. (New Taipei City, Taiwan) and Unity 2019, which is a domain development environment. Unity is also a cross-platform game engine used for diverse game software development, such as in VR or XR spaces and for developing smartphone applications. It includes fundamental features (aspects) such as diverse target object notification
and including special effects like decorations, lightning, and pre-recorded sound effects.

The Unity environment-based projects generally run on various operating systems (e.g., Windows, Macintosh, Linux). We use the program code-developmental environment with the Visual C# software development kit (SDK) on Visual Studio 2019. The SDK contains general libraries, application programming interface (API) files, developer tools, documents, and demo samples.

We describe further specifications of the HMD-based VR system as follows:

1) HTC Vive Pro Eye HMD specs: The screen is dual OLED 3.5” diagonal with a resolution of 1440×1600 pixels per eye (2880×1600 pixels combined); the refresh rate is 90.0 Hz; and the field of view is 110.0°. The installed sensors are: 1) SteamVR Tracking, 2) a G-sensor, 3) a gyroscope, 4) a proximity sensor, 5) an IPD sensor, and 6) an eye-tracking sensor. Ergonomics involves eye relief with lens distance adjustment.

Vive Pro Eye is capable of supporting an area up to 32 ft × 32 ft (= 1024 ft²) using four SteamVR Base Station 2.0 Units. Here, two SteamVR Base Station 2.0 Units support an area of up to 22.11 ft × 22.11 ft (= 489 ft²) and two SteamVR Base Station 1.0 Units support an area of 11.5 ft × 11.5 ft (= 132 ft²).

2) Eye-tracking specifications: The gaze-data output frequency (binocular) is 120 Hz; the accuracy is 0.5°–1.1°; and there are five calibration points. The trackable field of view is 110° (eye surgery, eye disease, heavy makeup, and severe myopia may affect eye-tracking performance). Furthermore, the system can obtain data for the user’s gaze origin, gaze direction, pupil position, pupil size, and eye openness.

3) Tracking area requirements: For the cases of both standing and seated in one position, there are no minimum space requirements. For the room area requirement, the minimum play area is 6.6 ft × 4.11 ft (= 27 ft²), while the maximum area is 32 ft × 32 ft (= 1024 ft²) with four SteamVR 2.0 Base Stations.

C. Program

Several studies have used open-source codes for various developments (e.g., for direction indication in advanced and simple object identification). To elucidate the logics and theory of the VR-based systems, we present the relative spatial locations in this study in Fig. 4 [16].

Unity and the Unreal Engine are two of the most dominant platforms in data-based experimental vision computing. We select Unity owing to its simplicity in the directions programming of VR spaces and its rather simpler and neat VR objects (e.g., simple plates, cubes, and balls). We set it such that it outputs CSV files of experimental data in timeline form. HTC’s openly distributed applications, codes, calibration software for users’ eye characteristics, and so forth, reduce computational burden.

For the first phase, we imported essential factors (packages, libraries, etc.) such as “UnityEngine”, “System.Collections”.Generic”, etc.), “System” (“Globalization,” “Text,” “IO,” etc.). We developed a typically styled program code that includes “Start()” and “Update()” functions. We created five connected cubes (the size is 20×20×20 based on the unit of Unity spaces) of data in the simulated agri-farm (Fig. 5, 6, and 7) and utilized it with consideration for the current technical situation and trends in the fields of VR, MR, and XR.
Fig. 6. Overview of a version of the developed VR world in the Unity-based system for the trials. (Red circle: start point; orange circles: tomato points where the subjects are supposed to gaze in the trials).

Fig. 7. Schematic of the program code written in Python language.

As illustrated in Fig. 2, the subjects move around gazing at red tomatoes, and the system automatically reacts and records the gazing position on the tomatoes as well as the timestamp. The VR-based system makes alerts with recorded sounds to notify successful and missed (with different sounds) chances to gaze at the cube on which the numbers were printed and connected to the C# program (Fig. 5). We patched the agricircumstance printed-texture files on them (Fig. 6).

For graphic computing, we used an MSI Gaming laptop PC, GP65-9SE-067JP by MSI Inc. (New Taipei City, Taiwan) It has the following specifications: GPU: Core i7; graphic board: RTX 2060; display size: 15.6 inch (Full HD); refresh rate: 144 Hz; memory: 16 GB; SSD size: 512 GB.

D. Experimental Method
We conducted experiments to test the proposed system’s performance, especially in relation to possible future agri-training using this system. After the Unity program started, the system was automatically executed. The distance between the eyes (in millimeters) and latitude of the pupils were inputted semi-manually prior to this using the default calibrating software.

The tomato plants for harvesting (Fig. 3 and 6) were selected based on the opinions of agri-farmers and managers with 5–20 years’ experience as well as according to other information from sources such as academic paper and guidebook, etc.
Tomato plants approximately 140 cm long, 410 cm wide, and 150 cm high, excluding small branches, were selected for harvesting. For the first trial phase, the subjects’ gaze moved from position A (default starting point) to position B (target point) and for the second phase, the subjects gazed from B to A.

In this study, considering the characteristics of vegetables discussed in related studies in Japan, we intentionally selected a vivid and distinct vegetable, the tomato, as is has the best visual clarity for executing the harvesting task. In case of any unexpected problems concerning the computer settings during the trials, the experimenter stops the movement of the system by hand and turns off the main button of the system. Otherwise, the trial ends when the subject reaches position B.

III. RESULTS

Table I shows the experimental trial data (average and standard deviation (SD)). The data concerning the main results of the trials include subjects’ agri-experiences, description of the direction of the trials, trial times (number), and execution time (s).

As is usual in such studies with the VR-based system, in some cases, subjects could not execute the procedure mainly because of (A) the interruption of the experiment, (B) interruption of the program, or (C) interruption due to other reasons. Reason (A), for instance, could include the following three cases: (1) subjects could not understand the trials’ successive contents, (2) subjects could not locate the next harvest number within 10 s, and (3) other subtle human errors. In the case of (C), for instance, one subject experienced virtual reality sickness.

| Subjects           | Description of trial | Trial time (number) | Execution time (s) |
|--------------------|----------------------|---------------------|--------------------|
| Experienced subjects (n = 2) | (a) from position A to B (Fig. 7) | N = 4 | 28.2 (10.3) |
|                    | (b) from position B to A | N = 4 | 36.0 (7.71) |
| Inexperienced subjects (n = 2) | (a) from position A to B | N = 4 | 26.9 (13.9) |
|                    | (b) from position B to A | N = 4 | 28.8 (17.4) |

IV. DISCUSSION

We confirmed through experiments that we could design, develop, and apply the original VR-based system with a certain level of quality for outdoor agri-workers. In Section III, we obtained the datasets concerning execution times and described the characteristics of the system and the trial operations. We could find no similar approaches for agricultural applications in the literature or in commercial use, although such methodological approaches have been used in the industrial and sports fields. Thus, quantifying the data accuracy is difficult. Moreover, we experimented only under limited conditions.

Specifically, for the finding and gazering errors of the harvest or inaccurate eye movements, we can present the cases (1) if the subjects could not achieve target harvest numbers, because of the obscurity (ambiguity) of colors or shapes, and (2) if the subjects could not perform the trials fast enough and accurately, mainly because of their inexperience and poor touch and even the discomfort of using the VR system. For (2), it is possibly only a temporary problem due to the subject being unaccustomed to the VR technology and this could be overcome with practice. Nevertheless, there could be problems related to the quality of the Unity system in capturing the visual data for constructing the VR images; this offers scope for future research and improvement.

V. CONCLUSION AND FUTURE TASKS

We developed and verified the fundamental phases of training agri-harvest skills needed for finding and discriminating targets, the developed VR-based system, and the methodologies by experimenting in a virtual agricultural site for inexperienced agri-workers in Japan. We measured and analyzed subjects’ cognition and behavior indicators in VR environments that simulated a work site. We presented significant and practical data from four trials per subject under the limited conditions and methodology, including output-based suggestions for agricultural trainers, leaders, managers, etc. Moreover, we suggested analysis methods for processes from the cognition of targets to be gazed using the eyetracking system. We realized the aforementioned measures relating to specific agri-work without the need to use actual agri-work sites.

We have diverse plans to realize and spread a wholly Internet-connected integrated VR agri-training system using the developed VR system. Furthermore, we will attempt to implement improved versions of this system. Our results have not completely achieved a fusion of the latest VR-based techniques, existing agricultural informatics, and eye-movement analyses. By combining the proposed system with other agricultural electrical systems, agri-trainers and trainees can achieve holistic improvement.

Ideally, in subsequent studies, we expect to utilize other high-spec HMDs and other peripherals in experimental situations.

REFERENCES

[1] F. Karim, and F. Karim, “Monitoring system using web of things in precision agriculture,” Procedia Computer Science, vol. 110, pp. 402-409, Jan. 2017.
[2] S. Wu, X. Li, and X. Wang, “IoU-aware single-stage object detector for accurate localization,” Image and Vision Computing, arXiv:1912.05992v4, April 2020.
[3] H. Pandey, A. Amuya, R. Prajapati, A. Pandey, and V. Nagve, “Augmented Reality in Agriculture (No. 2752),” EasyChair Reprint, Feb. 2020.
[4] A. Pandey, P. Tiwary, S. Kumar, and S. K. Das, “A hybrid classifier approach to multivariate sensor data for climate smart agriculture cyber-physical systems,” Proceedings of the 26th International Conference on Distributed Computing and Networking, pp. 337-341, Jan. 2019.
[5] V. Ponnumamy and S. Natarajan, “Precision Agriculture Using Advanced Technology of IoT, Unmanned Aerial Vehicle, Augmented Reality, and Machine Learning,” In Smart Sensors for Industrial Internet of Things, Springer, Cham, pp. 207-229, Feb. 2021.
[6] L. Vigoroso, F. Caffaro, C. M. Micheletti, and E. Cavallo, “Innovating Occupational Safety Training: A Scoping Review on Digital Games and Possible Applications in Agriculture,” International Journal of Environmental Research and Public Health, vol. 18, no. 4, pp. 1868, Feb. 2021.

DOI: http://dx.doi.org/10.24018/ejfood.2021.3.3.264
C. M. Flores-Cayuela, R. González-Pereza, E. Camacho-Poyato, and P. Montesinos, “Verifiable Water Use Inventory Using ICTs in Industrial Agriculture,” *Proceedings in Water Footprint*, pp. 1-34, April 2021.

C. H. Wang, C. Y. Liu, P. N. Pan, and H. R. Pan, “Research into the E-learning model of agriculture technology companies: Analysis by deep learning,” *Agronomy*, vol. 9, no. 2, pp. 83, Feb 2019.

L. M. Müller, K. Mandon, P. Glesche, S. Weiß, and W. Heuten, ember, “Visualization of Eye Tracking Data in Unity3D,” *Proceedings in 19th International Conference on Mobile and Ubiquitous Multimedia*, pp. 343-344, Nov. 2020.

S. Narasimhan, “Evaluating the Accuracy of Gaze Detection for Moving Character (Doctoral dissertation),” URI http://hdl.handle.net/10106/29416, August 2020.

S. Kim, M. Billinghamurst, G. Lee, and W. Huang, “Gaze window: A new gaze interface showing relevant content close to the gaze point,” *Journal of the Society for Information Display*, vol. 28, no.12, pp. 979-996, July 2020.

S. Pastel, C. H. Chen, L. Martin, M. Naujoks, K. Petri, and K. Witte, “Comparison of gaze accuracy and precision in real-world and virtual reality,” *Virtual Reality*, pp. 1-15, April 2020.

A. K. Jogeshwar, G. J. Diaz, S. P. Farnand, and J. B. Pelz, “The Cone Model: Recognizing gaze uncertainty in virtual environments,” *Electronic Imaging*, vol. 2020, no.9, pp. 288-1-288-8, June 2020.

A. Erickson, N. Norouzi, K. Kim, R. Schubert, J. Jules, J. LaViola, and G. F. Welch “Sharing gaze rays for visual target identification tasks in collaborative augmented reality,” *Journal on Multimodal User Interfaces*, vol. 14, no. 4, pp. 353-371, Nov. 2020.

G. Park, H. Cho, U. Lee, and S. Chin, “Virtual figure model crafting with VR HMD and Leap Motion,” *The Imaging Science Journal*, vol. 65, no. 6, pp. 358-370, July 2017.

I. Kakinuka and S. Komiyama. “The effect of using a controller to adjust gaze pointing in VR space,” *Journal of Japan Human Interface Society*, vol. 23, no. 1, pp. 89-100, Jan. 2021.

R. Nakamura, “Behavioral Analysis and Automatic Content generation for utilization of Virtual Reality Data,” *Masashino University Academic Institutional Repository*, vol. 23, no. 1, pp. 55-63, Jan. 2021.

J. Hollis and J. Oliver, “The Development of Virtual Reality as a Tool to Investigate Eating Behavior,” *Current Developments in Nutrition*, vol. 4, no. 2, pp. 1308-1308, May 2020.

M. Hirota, H. Kanda, T. Endo, T. Miyoshi, S. Miyagawa, Y. Hirohara, and T. Fujikado, “Comparison of visual fatigue caused by head-mounted display for virtual reality and two-dimensional display using objective and subjective evaluation,” *Ergonomics*, vol.62, no. 6, pp. 759-766, March 2019.

R. Adachi, E. M. Cramer, and H. Song, “Using virtual reality for tourism marketing: A mediating role of self-presence,” *The Social Science Journal*, pp. 1-14, Feb. 2020.

M. Singh, A. K. George, W. Eyoh, R. P. Homme, D. Stansic, and S. C. Tyagi. “High-methionine diet in skeletal muscle remodeling: epigenetic mechanism of homocysteine-mediated growth retardation.” *Canadian Journal of Physiology and Pharmacology*, vol. 99, no. 1, pp. 56-63, Aug. 2020.

F. N. Schrama, F. Ji, A. Hunt, E. M. Beunder, R. Woolf, A. Tuling, and Y. Yang, “Lowering iron losses during slag removal in hot metal desulfurization without using fluoride,” *Ironmaking & Steelmaking*, vol. 47, no. 5, pp. 464-472, March 2020.

A. Boumezoued, “Improving HMD mortality estimates with HFD fertility data,” *North American Actuarial Journal*, URL: https://hal.archives-ouvertes.fr/hal-01276565/document, pp.1-25, Feb. 2016.

Shinji Kawakura Place of birth: Toyama Pref., Japan. Date of birth: July 14, 1978. Ph.D. in Environmentalology, University of Tokyo, 2015, Bunkyō-ku, Tokyo, Japan. B.A. in Control System Engineering, Tokyo Institute of Technology, 2003, Meguro-ku, Tokyo, Japan. M.A. in Human-Factor Engineering, Tokyo Institute of Technology, 2005, Meguro-ku, Tokyo, Japan. Career: Systems engineering, research for private companies. Development and verification of sensing systems for outdoor agricultural workers. Dr. Kawakura, Research Center for Artificial Photosynthesis (ReCAP) at Osaka City University/Osaka City, Osaka, Japan. IEEE senior member, Hong Kong Chemical, Biological & Environmental Engineering Society (HKCBEEES) senior member.

Ryosuke Shibasaki Place of birth: Fukuoka. Pref. Japan. Date of birth: March 1, 1958. Dr. in Engineering, the University of Tokyo, 1987, Bunkyō-ku, Tokyo, Japan. B.A in Engineering, the University of Tokyo, 1980, Bunkyō-ku, Tokyo, Japan. M.A. in Engineering, the University of Tokyo, 1982, Bunkyō-ku, Tokyo, Japan. Career: Professor at the Center for Spatial Information Science, University of Tokyo. Dr. Shibasaki, Center for Spatial Information Science (CSIS), The University of Tokyo, Meguro-ku, Tokyo, Japan, and Department of Socio-Cultural and Socio-Physical Environmental Studies, The University of Tokyo/Kashiwa-shi, Chiba, Japan.

DOI: http://dx.doi.org/10.24018/ejfood.2021.3.3.264
Vol 3 | Issue 3 | June 2021

126