

Article

Drone Operation with Human Natural Movement

Hironori Hiraishi *

Faculty of Engineering, Ashikaga University, Tochigi 326-8558, Japan

* Corresponding author. E-mail: hiraishih@gmail.com (H.H.)

Received: 24 March 2025; Accepted: 7 May 2025; Available online: 12 May 2025

ABSTRACT: This study proposes a method for operating drones using natural human movements. The operator simply wears virtual reality (VR) goggles. An image from the drone camera was displayed on the goggles. When the operator changes the direction of his or her face, the drone changes the direction to match that of the operator. When the operator moves their head up or down, the drone rises or falls accordingly. When the operator walks in place, rather than walking, the drone moves forward. This allows the operator to control the drone as if they were walking in the air. Each of these movements was detected by the values of the acceleration and magnetic field sensors of the smartphone mounted on the VR goggles. A machine learning method was adopted to distinguish between walking and non-walking movements. Compared with operation via conventional remote control, it was observed that the remote controller performed better than the proposed approach in the early stages. However, when the participants familiarized themselves with the natural operation, these differences became relatively small. This study combined drones, VR, and machine learning. VR provides drone pilots with a sense of realism and immersion, whereas machine learning enables the use of natural movements.

Keywords: Drone; Virtual reality; Human computer interaction; Natural user interface; Machine learning; Support vector machine



© 2025 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The technology for small flying drone vehicles has rapidly developed over the past decade. Some expected applications of drones include delivery services [1]; management of disasters such as earthquakes, floods, and forest fires [2]; agriculture [3]; and security patrols and monitoring [4,5]. The Global Positioning System (GPS) is the main technology used by drones for navigational control [5,6]. Therefore, the aforementioned outdoor applications are the primary focus of this research. However, some studies have been conducted on other topics, including manual controller design [7], localization methods [8–10], and landing and navigation methods [11,12] for autonomous flying in an indoor environment.

In general, drones are controlled using a smartphone or special remote controller, except for autonomous flights, which are controlled by software programming. In this study, a method by which a user can operate a drone using natural human movements was proposed. A drone operator wears a set of virtual reality (VR) goggles into which a smartphone is inserted. The operator can view the video feed captured by the drone camera in the VR display. The smartphone detects the movements of the operator based on the data collected by its acceleration and magnetic field sensors. When the operator turns their face or body to the right or left, the drone turns accordingly. When the operator turned their face up or down, the drone turned up and down, respectively. Moreover, the drone moved forward when the operator walked in place.

Some drones include a first-person view (FPV) mode, in which an operator can view a camera image in VR, as in this study. However, these drones are still operated using remote controllers. In the proposed approach, the flight of the drone is synchronized with the facial movements of the operator. Therefore, the operator can perceive the surroundings of the vehicle more immersively, as if they are onboard. This study is an extended version of [13] and is related to human computer interaction (HMI) and Natural User Interface (NUI) research for drone operation. Other studies have reported controlling drones using human body gestures [14], eye trackers, and hand gestures [15], and even using a VR environment, such as that in this study, to control the drone using voice and body postures [16]. This study differs from previous ones by not using special gestures, voice commands, or sensors to control the drone; instead, it relies on natural

human movements, which is a key feature of the study. The remainder of this paper is organized as follows. The architecture of the proposed system is presented in Section 2. Section 3 describes the method for detecting the facial movements of an operator. Section 4 presents the method used to detect stepping using machine learning. Section 5 presents the results of the experiments conducted to verify the effectiveness of the proposed approach, and Section 6 concludes the paper.

2. System Architecture

In this study, a Tello quadcopter (Ryze Technology) was used. The Tello is a small toy-type drone that is suitable for indoor flights and can be operated easily and safely. Tello focused on its educational use. Therefore, the programming library is public, which allowed us to develop our own software [17].

Figure 1 illustrates the drone operation scenario and the hardware setup. An operator wears VR goggles to watch camera images from the Tello drone and operates the drone by moving the face and body and walking on the spot. The smartphone shown in Figure 1b was inserted into the VR goggles. Therefore, the camera image from the drone was displayed in two parts: the left eye and the right eye.

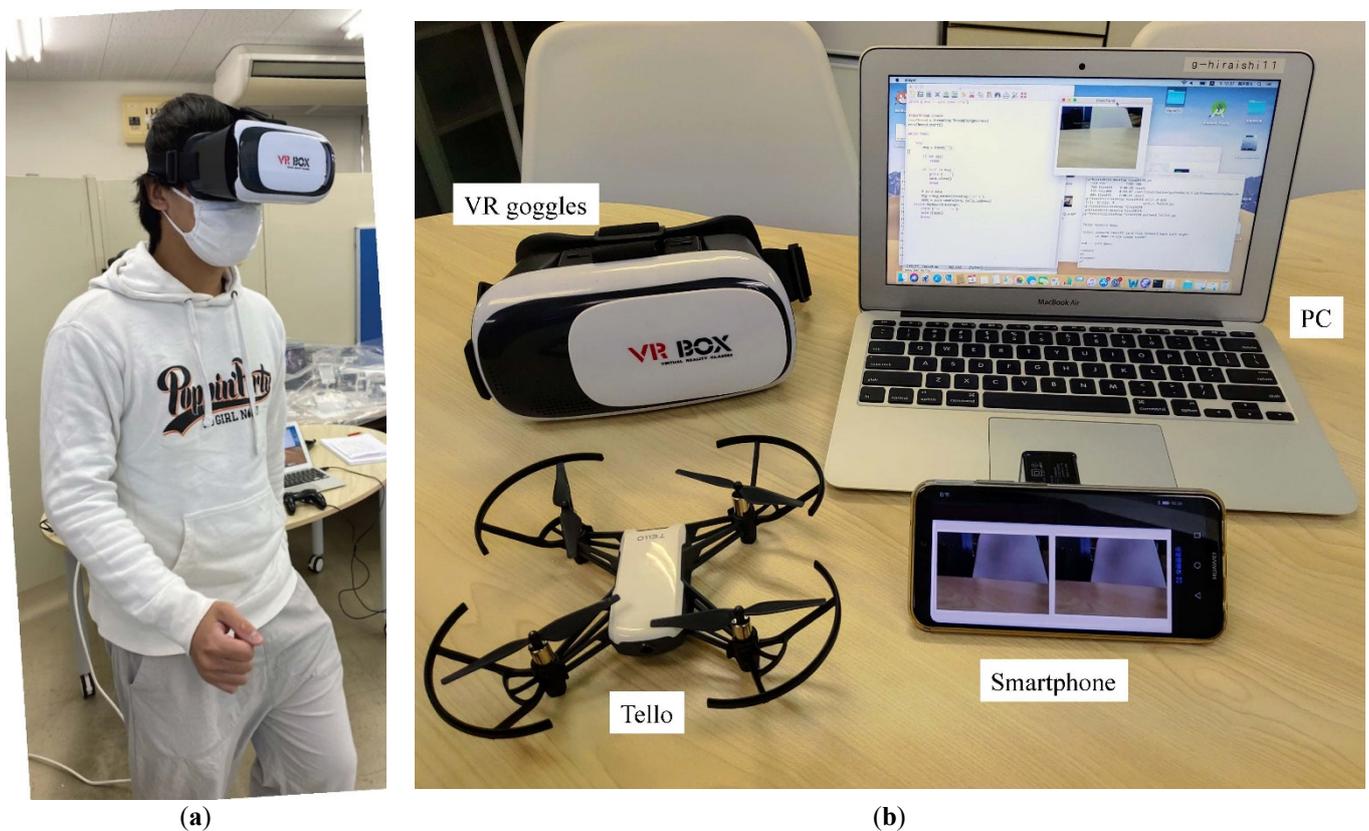


Figure 1. (a) Situation of drone operation; (b) Hardware set.

The main program of the system performs flight control based on user input and was developed to run on a PC (specifically, Apple MacBook Air). Flight control of Tello was executed using the main program. The PC communicates with Tello through Wi-Fi to send control commands and receive camera images.

The smartphone installed in the VR goggles was connected to a PC using a USB cable. Bluetooth is also available but is slower than USB. The smartphone in the goggles detects human movements. The direction of the smartphone was detected using acceleration and magnetic field sensors, and stepping was detected using the acceleration sensor. A support vector machine (SVM) is a powerful machine learning method adopted in a wide variety of applications. An SVM classifier [18] is used to distinguish stepping from other movements. The camera images were sent to the smartphone from the PC, and the direction of the smartphone and the stepping state were sent to the PC.

The Tello can also be operated using the included remote controls. The effectiveness of the operation with that of the proposed natural approach and remote controller was compared, and the results are described in Section 5.

3. Detecting Face Movement

Acceleration and magnetic field sensors in smartphones can be used to detect facial movements. A smartphone program was developed for the Android operating system. The program was developed to support a library that queries sensor data to detect smartphone orientation using the aforementioned sensors.

Figure 2 shows the orientation of the smartphone. Azimuth refers to the horizontal orientation, pitch indicates the vertical orientation, and roll indicates the tilt in the left and right directions. The smartphone was placed in the VR goggles. Therefore, azimuth corresponds to turning left and right, pitch corresponds to looking up and down, and roll corresponds to the user leaning their face in some direction. The Tello synchronizes with the user's facial movement, as detected by the smartphone in their VR goggles.

The main program on the PC detected the difference between the azimuth of the smartphone and that of the Tello drone when the operator turned. If the difference was greater than 15 degrees, the main program sent a command to turn toward the operator's azimuth. The system directs Tello to turn 15 degrees at each command. The main program continued to send commands until the difference became less than 15 degrees. If the pitch varies by 5 degrees or more from the horizontal direction toward the upward direction, Tello moves up by 20 cm. In contrast, if the pitch varied by 10 degrees or more from the horizontal to the downward direction, Tello moved down by 20 cm. In general, operators naturally face slightly down when wearing VR goggles. Therefore, the thresholds differed between the up and down controls. If the operator continues to look up or down, the command to move up or down is sent continuously. If the roll angle deviates by more than 15 degrees from the horizontal, the Tello is directed to move left or right by 20 cm.

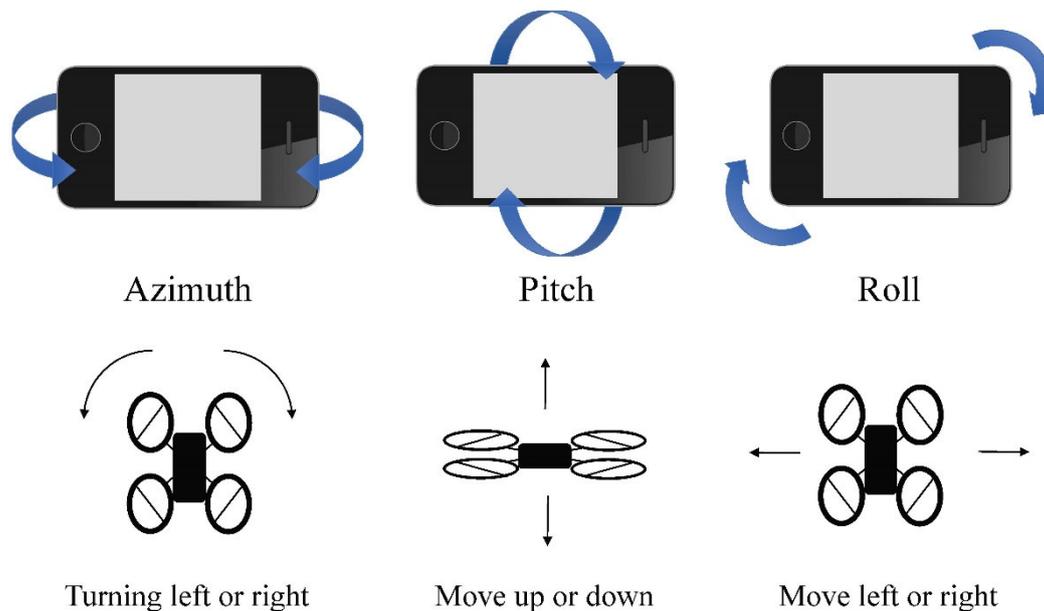


Figure 2. Hardware set operation overview.

4. Detecting Walking in Place

Detecting facial movements enables the direct control of the drone's facing direction. The user can walk to move the drone forward. Acceleration data were used to detect steps. These data comprise three dimensions, labeled the X-, Y-, and Z-axes. When facing the smartphone screen, the X-axis is in the horizontal direction, the Y-axis is in the vertical direction, and the Z-axis is in the front-back direction.

In this study, a SVM classifier was applied to distinguish stepping from other movements. The data from these two classes were collected to create the classification model. One class included data recorded while the user walked in place, whereas the other included data recorded while the user stood in place and turned their faces and bodies. Five acceleration data samples were collected per second. In this study, 1 s of data for each of the two classes were set to one dataset (three dimensions \times five data \times two classes = 30 data samples).

Fifty datasets for each of the three subjects (male students) were collected to train the model, and ten datasets were collected to evaluate its performance. Figure 3 shows the change in accuracy with an increasing number of datasets. ALL indicate accuracy when the data from the three subjects were merged. For the most part, the accuracy improved

with increasing amounts of data, peaking at 30 or 40 datasets. For Subject A, the accuracy gradually improved from 60% to 85% with 40 datasets. The accuracy decreased slightly for 50 datasets. This is thought to be due to overfitting, which is specific to machine learning. Next, for Subject B, the accuracy was initially 65%, but eventually reached 70%, with little improvement. For Subject C, the accuracy was initially low at 55% but reached nearly 100% for the 30 datasets and then stabilized at 95%.

The accuracy for Subject B was not as high as that of the other subjects. This may be attributed to the slower step speed compared to that of the other subjects. However, the ALL exceeded 80% even with only 10 datasets. The maximum accuracy was 88%, and it reached a stable level. Therefore, the model used in this study was created using all the collected data.

In the case of low accuracy, the stepping state was not recognized, and the user was judged as having stopped. There were no cases in which stopping was considered as stepping. Therefore, the drone never moved forward when the user stopped. Faster stepping is easier to recognize. The system was designed to make the drone move forward by 40 cm when it detected a walking step.

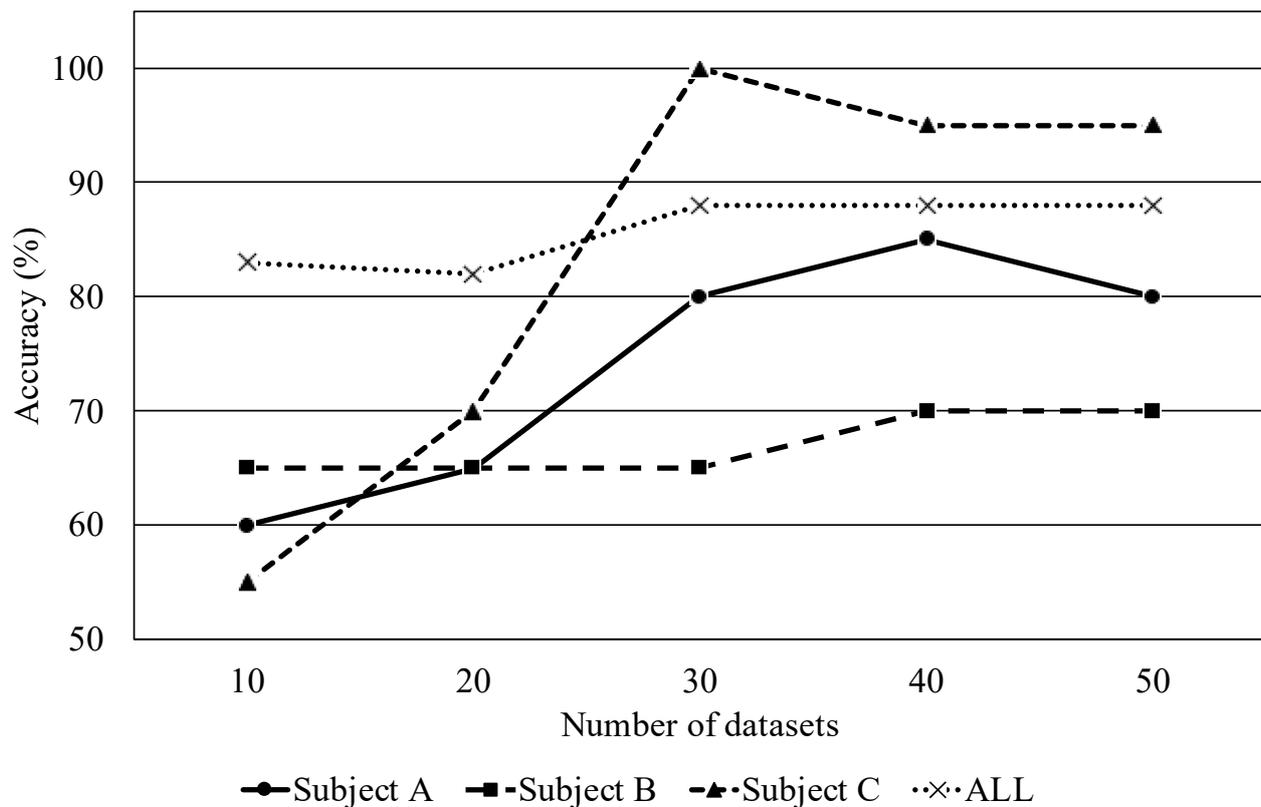


Figure 3. Accuracy of the classification model.

5. Experiments

To verify the effectiveness of the proposed approach, natural operation was compared with remote controller operation. During the remote controller operation, the subjects operated the controller while watching the camera image from the drone on the PC monitor, as shown in Figure 4a, rather than using VR goggles.

To eliminate the difference in control, the moving distance or angle assigned to one command of the remote controller was set to be the same as that of the natural operation. Figure 4b shows the assignment of each control. The cross keys on the left correspond to each movement, the top key corresponds to the walking step, and the left and right keys correspond to the left and right leaning of the head, respectively. The cross buttons on the right correspond to heading up and down, and turning right and left. In addition, to eliminate the difference in real-time performance, the image from the drone was sent to the smartphone, even during the remote controller operation. These settings allowed us to evaluate VR and natural operations.

Two types of experiments were conducted in this study. The first aim was to verify the operations in terms of face direction. The other was performed to evaluate the overall natural operation methods, including the stepping function.

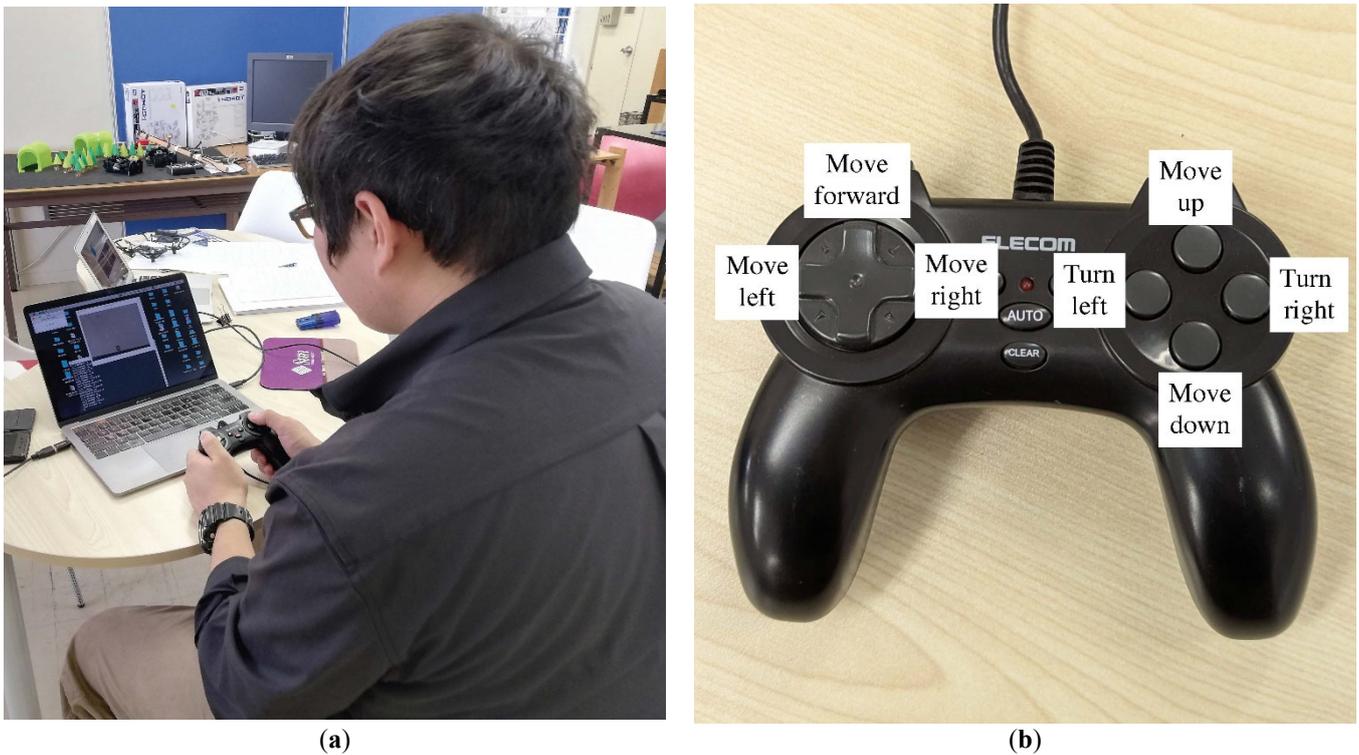


Figure 4. (a) The situation of the remote controller operation; (b) Control assignments for the remote controller.

5.1. Evaluation of the Face Operation

First, only facial operations were evaluated using the experimental environment shown in Figure 5. The environment was set up with two virtual cubes, each of which was 90 cm square. Four X-marks were placed on the walls at positions 1, 2, 3, and 4. The drone was located at the initial position and directed toward the first marker.

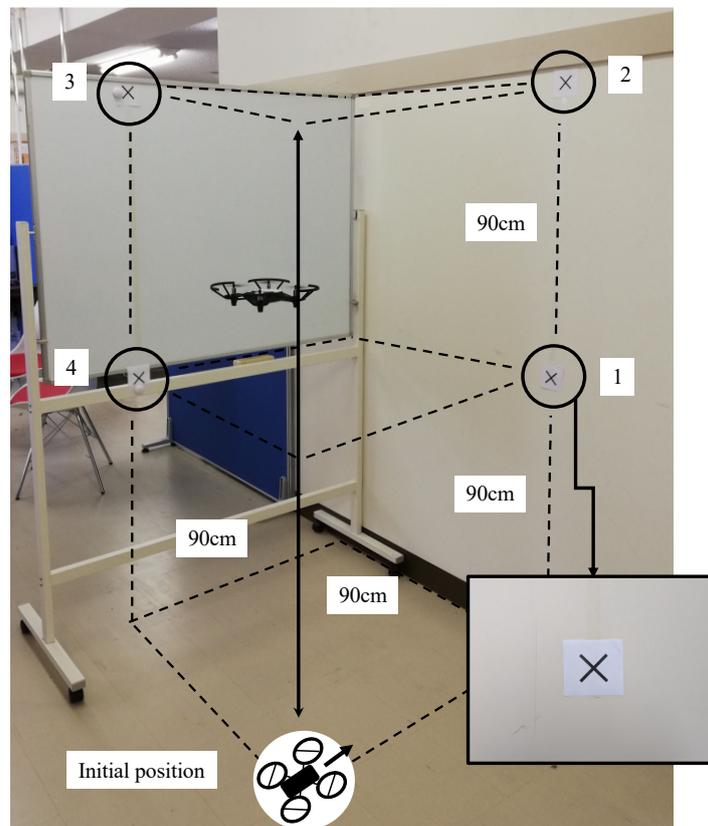


Figure 5. Experimental environment for the face operation.

After the drone performed takeoff, the operator moved the drone up until its camera caught marker No. 1 at the center of the image, as shown in Figure 5. This initiates the experimental data collection process. The operator moved the drone up until the camera found marker No. 2 at the center of the image. Next, the operator turned the drone to the left until the camera found the No. 3 marker at the center of the image. The operator then moved the drone down until the camera captured marker No. 4 at the center of the image. Finally, the operator turned the drone to the right until the camera found the marker No. 1 at the center of the image again. The experiment was then completed. For this experiment, the operator must face up until marker No. 1 is found and must still face up until marker No. 2 is found. After finding marker No. 2, the operator must turn left until marker No. 3 is found. Next, the operator faces down until maker No. 4 is found. Finally, the operator must turn right until maker No. 1 is found again. The execution time is measured in seconds. Figures 6 and 7 show the experimental results for the three subjects (male students other than those who participated, as noted in Section 4). Each subject executed the experimental process three times for both remote control and natural operations.

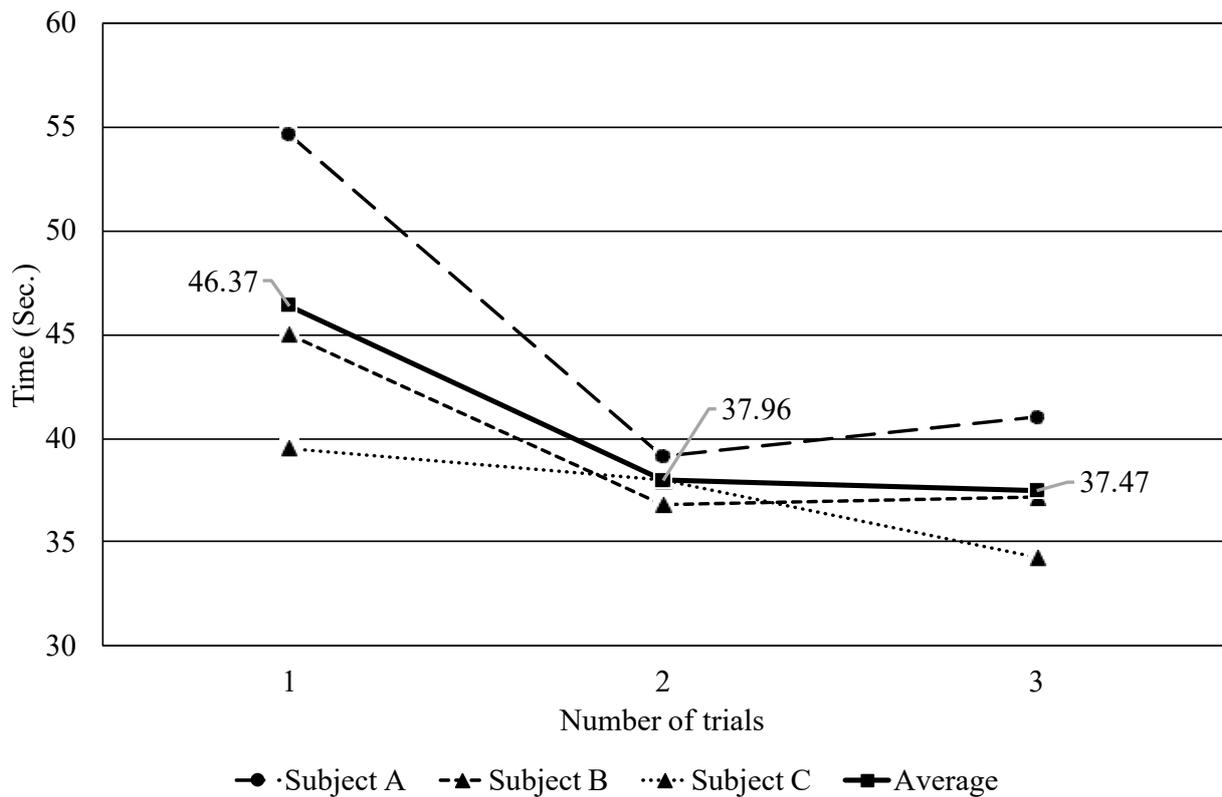


Figure 6. Results of the operation by remote controller.

For the remote control, as shown in Figure 6, the execution time for Subject A at first time was approximately 55 s. This took much longer than for the other subjects. However, the difference between the other subjects became much smaller the second time. The execution time for Subject B was approximately 45 s the first time. The time improved during the second attempt, just as it did for Subject A. In the case of Subject C, the execution time was less than 40 s even for the first time. The execution time gradually decreased and eventually dropped to less than 35 s. As indicated by the average time, the execution time for each subject was reduced to a relatively low level in the second trial. The remote controller is a general game controller. Because each subject was already familiar with this type of controller, a single trial may have been sufficient for them to learn the process well.

In the case of the natural operation method, as shown in Figure 7, the execution time decreased gradually. The execution time for Subject A improved linearly: 55 s for the first time, 48 s for the second time, and 41 s for the third time. The shape of the time for Subject B was similar to that of the remote controller. It was thought that Subject B understood how to operate it by the first trial. Subject C was fast for the first time, but worsened for the second time. It was assumed that the time of the first attempt was by chance, and the time improved gradually. As indicated by the average time, the execution time gradually improves. This indicates that users needed more time to familiarize themselves with the natural operation compared to the remote controller. However, this difference decreased as the number of trials increased.

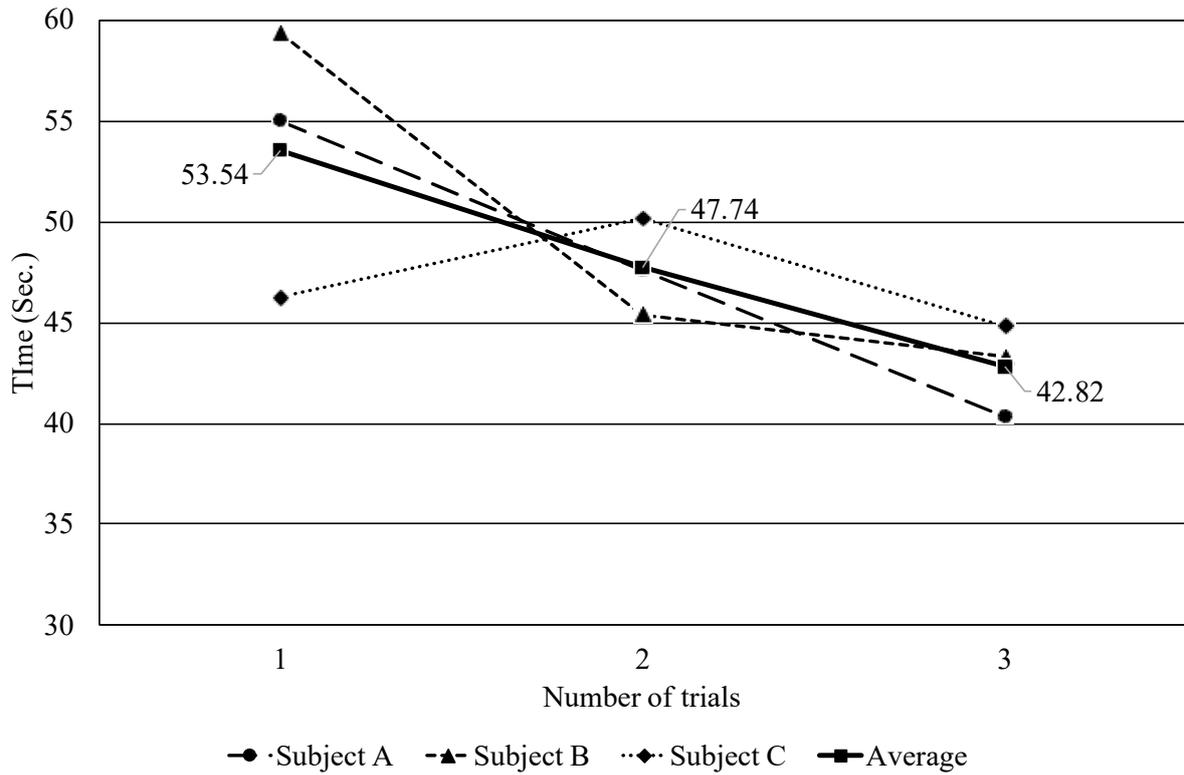


Figure 7. Results of the face operation.

5.2. Total Evaluation of the Natural Operation

Figure 8 shows the experimental environment used for the evaluation. An operator flies the drone along the course shown on the left of Figure 8. The drone was maintained automatically at a height of 1m. Therefore, height control was not considered in the experiment. The operator controlled the drone by stepping, leaning their face, and changing the direction they faced. The operator was directed to hover the drone over its initial position. The experiment began when the flight stabilized. The drone flew along the course. The experiment was completed when the drone crossed the goal line. The execution time was measured in seconds from the designated starting point to the time when the goal was reached.

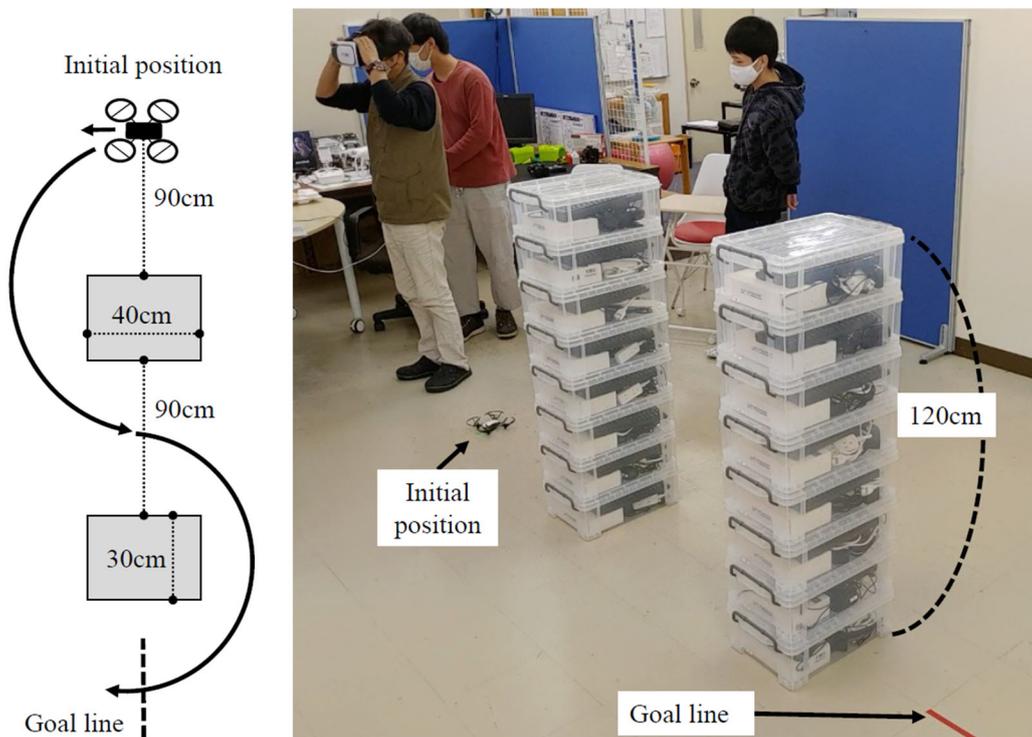


Figure 8. Experimental environment.

Three subjects (male students other than those who participated, as noted in Sections 4 and 5.1) performed the experiment with both the remote controller and natural operation. Table 1 lists the results of the operation using the remote controller, and Table 2 lists the results of the natural operation.

Table 1. Results of the remote controller (s).

Number of Trials	1	2	3
Subject A	108	103	88
Subject B	86	96	115
Subject C	110	102	88
Average	101.3	100.3	97.0

Table 2. Results of the natural operation (s).

Number of Trials	1	2	3
Subject A	153	118	116
Subject B	129	116	99
Subject C	116	97	88
Average	132.7	110.3	101.0

As listed Table 1, the execution time of Subject A decreased gradually. In contrast, the execution time of Subject B gradually increased. Subject B collided with an obstacle in the third trial. The execution time of Subject C gradually decreased, similar to that of Subject A. Subject B recorded the best time of all at first time. However, the execution time was not stable. Therefore, it was considered that Subject B could not learn the operation steadily after only three trials. However, the average time decreased as the number of trials increased.

As shown in Table 2, the execution time of Subject A was 153 s during the first trial, which significantly improved in the second trial. The execution time of Subject B gradually decreased during the three trials. The execution time of Subject C also gradually improved, similar to that of Subject B. Furthermore, Subject C achieved the same operation time as the remote controller with natural operation.

Comparing the average times for both results, the difference between the two operations decreased for each trial. Therefore, it has been considered that the proposed method can be used to operate drones as effectively as the remote controller after users become familiar with the process. Operating the drone with remote control took relatively more time in the third trial for Subject B, owing to a collision with an obstacle. The participants operated the device while watching the camera image on a PC monitor. They encountered some difficulty in understanding the distance to obstacles in relation to their flight speed. Natural operations with VR goggles provide a more immersive sense of distance. This result is similar to those of previous studies [14,15]. The proposed system allows users to control the viewing angle and movements in the remote space more intuitively.

6. Conclusions

This study proposes a method by which drones can be operated using natural human movements. The operator wears VR goggles to view the camera image from the drone and operated the drone by changing the face direction, leaning the head, and stepping on a spot. This natural operation is compared with that of a conventional remote controller. The remote controller performed better during the initial stages of the experiment. However, as the subjects became more familiar with the proposed approach, this difference became relatively small, and the proposed system allowed users to actively control the viewing angle and movements in the remote space more intuitively.

The software library included with the Tello Craft for general developers permits only step-wise control, such as moving a certain distance or turning to a specified degree. Because speed control cannot be performed in real-time, commands such as movements of 20 cm and 15 degrees must be sent continually. The movement of the Tello unit was irregular. Therefore, future research should apply the proposed approach to other drones with more precise control systems to achieve smoother operation.

This study combined drones, VR, and machine learning. VR provides drone pilots with a sense of realism and immersion, whereas machine learning enables the use of natural movements. This type of research could lead to intuitive operating experiences in entertainment and education, the ability to control drones with facial movements even when their hands are full during rescue missions and disaster sites, and assistance for people with disabilities in medical and rehabilitation applications by controlling the movements of specific body parts. To achieve these applications, issues

such as operational stability, adoption of differences between users, and ensuring safety need to be resolved. Human-centered design and user-adaptive interface design will become increasingly important in the future.

Acknowledgments

We would like to thank Editage (www.editage.jp) for English language editing.

Ethics Statement

Not applicable.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

There are no specific research data related to this paper. A video showing the results of this research can be accessed at https://youtu.be/SNwuBOT4yog?si=W-EoSh3KaT_RMjQ7 (accessed on 10 May 2025).

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Amazon Prime Air. Available online: <https://www.aboutamazon.com/news/transportation/amazon-prime-air-prepares-for-drone-deliveries> (accessed on 18 March 2025).
2. Restas A. Drone Applications for Supporting Disaster Management. *World J. Eng. Technol.* **2015**, *3*, 316–321.
3. Veroustraete F. The Rise of the Drones in Agriculture. *EC Agric.* **2015**, *2*, 325–327.
4. Ostojić G, Stankovski S, Tejić B, Đukić N, Tegeltija S. Design, control and application of quadcopter. *Int. J. Ind. Eng. Manag.* **2015**, *6*, 43–48.
5. Gowda M, Manweiler J, Dhekne A, Choudhury RR, Weisz JD. Tracking drone orientation with multiple GPS receivers. In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, New York, NY, USA, 3–7 October 2016; pp. 280–293.
6. Gadda JS, Patil RD. Quadcopter (UAVs) for border security with GUI system. *Int. J. Res. Eng. Technol.* **2013**, *2*, 620–624.
7. Hanafi D, Qetkeaw M, Ghazali R, Than MN, Utomo WM, Omar R. Simple GUI Wireless Controller of Quadcopter. *Int. J. Commun. Netw. Syst. Sci.* **2013**, *6*, 52–59.
8. Hussein A, Al-Kaff A, de la Escalera A, Armingol JM. Autonomous Indoor Navigation of Low-Cost Quadcopters. In Proceedings of the IEEE International Conference on Service Operations And Logistics, and Informatics (SOLI), Yasmine Hammamet, Tunisia, 15–17 November 2015; pp.133–138.
9. Mac TT, Copot C, De Keyser R, Ionescu CM. The development of an autonomous navigation system with optimal control of an UAV in partly unknown indoor environment, *Mechatronics* **2018**, *49*, 187–196.
10. Krajník T, Nitsche M, Pedre S, Přeučil L, Mejail ME. A simple visual navigation system for an UAV. In Proceedings of the 9th International Multi-Conference on Systems, Signals & Devices, Chemnitz, Germany, 20–23 March 2012.
11. Sani MF, Karimian G. Automatic navigation and landing of an indoor AR. drone quadrotor using ArUco marker and inertial sensors. In Proceedings of the International Conference on Computer and Drone Applications (IconDA), Kuching, Malaysia, 9–11 November 2017; pp. 102–107.
12. Polvara R, Sharma S, Wan J, Manning A, Sutton R. Towards autonomous landing on a moving vessel through fiducial markers. In Proceedings of the European Conference on Mobile Robots (ECMR), Paris, France, 6–8 September 2017; pp. 1–6.
13. Hiraishi H. Maneuvering a Drone with Natural Human Movement. In Proceedings of the 1st International Conference on Drones and Unmanned Systems (DAUS' 2025), Granada, Spain, 19–21 February 2025; pp. 8–10.
14. Gio N, Brisco R, Vuletic T. CONTROL OF A DRONE WITH BODY GESTURES. *Proc. Des. Soc.* **2021**, *1*, 761–770.
15. Di Vincenzo M, Palini F, De Marsico M, Borghi AM, Baldassarre G. A Natural Human-Drone Embodied Interface: Empirical Comparison With a Traditional Interface. *Front. Neurobot.* **2022**, *16*, 898859.

16. LO CY, HOU JH. A Natural Human-Drone Interface For Better Spatial Presence Experiences. In Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Sydney, Australia, 9–15 April 2022; pp. 99–108.
17. Tello SDK 1.3.0.0. Available online: <https://www.rzyzerobotics.com/jp/tello/downloads> (accessed on 18 March 2025).
18. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, 2, 1–27.