A NATURAL HUMAN-DRONE INTERFACE FOR BETTER SPATIAL PRESENCE EXPERIENCES

CHUN-YU LO¹ and JUNE-HAO HOU²
¹²Graduate Institute of Architecture, National Yang Ming Chiao Tung University.
¹chunyulo@arch.nycu.edu.tw, 0000-0002-7910-849X
²jhou@arch.nycu.edu.tw, 0000-0002-8362-7719

Abstract. As many remote construction projects increase in size and complexity, being able to manage personnel schedules, delegate tasks, and check work progress can improve work efficiency and productivity. Hence, video conferencing and remote monitoring software have been attempting to pursue an immersive and intuitive experience, but with limited developments. To better achieve that, we propose a system with a natural user interface (NUI) that can offer a vivid experience, facilitating AEC personnel who is novice drone operator to interact with the Unmanned Aerial Vehicle (UAV) by voice instructions and body posture to conduct remote site surveying, monitoring, and inspections instead of physical visiting. In addition, the proposed system is capable of on-demand path planning and camera movements for various tasks and enhances the spatial experience. We integrate these techniques to develop a human-drone interface, including a VR simulator and a haptic vest system, which offer a perceivable experience of spatial presence for different purposes. Compared with other relative works, the proposed system allows users to actively control the viewing angle and movements in the remote space more intuitively. Moreover, drones can augment human vision and let users gain mobile autonomy.

Keywords. Spatial Presence; Natural User Interface; Human-drone Interaction; Virtual Reality; Remote Working; Body Posture Recognition; Speech Recognition; SDG 9.

1. Introduction

This paper proposes a system with a natural user interface (NUI) for AEC personnel who is novice drone operator to interact with the Unmanned Aerial Vehicle (UAV) by voice instructions and body posture to conduct remote site surveying, monitoring, and inspections instead of physical visiting. In addition to NUI, the proposed system is capable of on-demand path planning and camera movements for various tasks and enhances the spatial experience. Besides, we develop a system including a VR
simulator for a plausible experience of spatial presence and training purposes.

The AEC industry has been impacted by the use of UAV (Albeaino et al., 2019). Yet it is still one of the domains with low remote working capabilities (Dey et al., 2020), which implies it has a high potential for developing remote communication technologies. Researches in telepresence suggested that human beings have been eager to participate in everything in-situ (Barfield et al., 1995). And drones are capable of augmenting human vision to enormous innovative applications (Erat et al., 2018). However, non-intuitive flight operation and camera control of UAVs not only becomes cognitively demanding during long-term operations but also discourages the use of drones as an experiential platform or physical avatar by novice users (Peschel and Murphy, 2012).

Fernandez et al. (2016) pointed out that the NUI can help the operator control the drone in an intuitive way. And the use of VR and first-person perspective can increase the user’s sense of presence in avatars (Macchini et al., 2021). Thus, this research aims to (1) allow UAV to respond to user’s intention by spoken instructions and body posture via proposed NUI; (2) simplify viewpoint control of UAV by semi-autonomous path planning; (3) improve the experience of drone operations in VR by enhancing spatial presence.

2. Research Background

2.1. REMOTE WORKING IN AEC INDUSTRY

Remote construction projects exist in many regions worldwide, such as deserts, polar regions, mountains, and sparsely populated areas; construction site monitoring of those areas can be a time-consuming task requiring users' presence for observations, decisions, and actions. In addition, since the outbreak of the epidemic, most industries have been forced to work remotely. However, many jobs cannot be performed remotely, and workers are required to be present in person. Thus, solutions for remote construction will save AEC personnel from visiting the site by letting them do the monitoring remotely. Furthermore, relevant research pointed out that the AEC industry is still one of the areas with low remote working capabilities (Dey et al., 2020), which means that it has great potential for the development of remote communication technology.

To provide this domain with the ability to work remotely, Cote et al. (2011) show that installing cameras on-site and assembling them into a 360 panorama can remotely view construction sites in an immersive way, which has been considered a very positive and effective method for remote monitoring and inspection.

2.2. FROM VIRTUAL REALITY TO SPATIAL PRESENCE

Many technology companies have been developing many immersive video conferencing software these years, such as Facebook Horizon Workroom, Spatial., which implies that human beings have been eager to experience space immersively. However, compared with the technique mentioned above, the AEC industry needs a remote communication technology that allows users to shuttle in the remote space freely.
Spatial presence is usually defined as the feeling of “being there”. The term “there” refers to a virtual location or a real remote location. Therefore, spatial presence includes the user's ability to experience a sense of presence in any environment where they are transported. According to the results of user testing, the remote environment unexpectedly has a higher sense of spatial presence or “hyper-presence” at certain times compared with the real environment (Khenak et al., 2020).

Moreover, drones can augment human vision and let operators gain mobile autonomy, and the use of VR and first-person perspective can increase the user’s sense of presence in avatars (Macchini et al., 2021). Therefore, combining drones with VR can provide users with a safe and fascinating flight environment for training to achieve spatial presence.

### 2.3. DRONE USER INTERFACE

Natural User Interfaces (NUIs) intend to make use of innate human features, such as speech, gesture, posture, and vision to interact with technology. Similarly, it is crucial for humans to interact and command drones in natural and efficient ways in Human-Drone Interaction (HDI) frameworks (Fernandez et al., 2016). Yam-Viramontes and Mercado-Ravell (2020) integrated body gestures into the control system of the drone. In the study, the proposed strategy is validated with different human users through a standard User Experience Questionnaire (UEQ), showing good results in usability and user experience. It proved that using NUI to control drones has good results and positive feedback.

### 3. System Design and Implementation

#### 3.1. SYSTEM STRUCTURE

This research developed a drone simulator that has the configuration of an actual UAV to test the feasibility in the future. Furthermore, we proposed a NUI with body posture and voice control for the drone according to the survey of the AEC industry's communication dialect.

Based on these two control principles, this research proposed three systems to control UAVs: (1) Responsive Control System (RCS), (2) Body Posture Control System (BPCS), and (3) Voice Control System (VCS). With these systems, the operator can receive the current flight status information from the simulator through vision and haptic feedback during the flight. Simultaneously, the recognition system will capture the body's response, transmit it to the workstation for data processing, and send the results to the simulator. Moreover, the operator can get the vision of the simulation from the screen or a head-mounted display (HMD), switch various scenes, perspectives, and give instructions through the NUI.
3.2. SYSTEM ANALYSIS AND DESIGN

At present, the applications of UAVs in the AEC industry are primarily used in monitoring, photography, scanning, and urban observation (Albeaino et al., 2019). According to the purposes of the above tasks, this research organizes these control systems into two UAV control modes: (1) Task mode and (2) Exploration mode. In addition, it is expected that the proposed NUI can automatically recognize the operator's intention to switch the control mode to achieve a natural use.

- **(1) The Task mode uses VCS as the primary input of the UAV, and the RCS serves as the emergency control. When performing simple tasks and flights such as photography and scanning, the drone only needs a simple flight path and functions. Hence, the operators can reduce their burden with the Voice control system and free their hands for other works or tasks. Moreover, if the body receives an environmental warning during the flight, the RCS will be activated and decide whether to terminate the ongoing mission based on the operator's behavior.**

- **(2) The Exploration mode uses BPCS as the primary input of the UAV, VCS and RCS serve as the function inputs and the emergency control. When performing remote city exploration or surveying, the drone needs to be fully controlled by the operators to follow their intention. Besides, the operators usually imagine themselves as the avatar of a drone during the flight. Thus, the operators can fully control the drone's flight through the BPCS in this control mode. In addition, VCS in this mode is used for taking off, landing, perspective switching, and other functions. Furthermore, if the body receives an environmental warning during the flight, the RCS will be activated and decide whether to terminate the ongoing mission based on the operator's behavior.**

To sum up, the VCS is mainly used to control UAV's take-off, landing, perspective-switching, path planning, image recording, etc. The BPCS is mainly used for controlling the flight (Roll, Yaw, Pitch). In addition, the RCS serves as emergency control; when the virtual drone perceives that the environment is facing an emergency,
the body reaction of the operator will be captured by the RCS to achieve obstacle avoidance.

For safety reasons, the RCS is designed as the highest-level system, covering any commands under any circumstances. Similarly, the operator can give instructions through voice, and if the BPCS is activated simultaneously, the previously spoken instructions will be overwritten.

![Hierarchy of control systems](image)

**Figure 2. Hierarchy of control systems**

### 3.3. THE SIMULATOR

This research uses Unreal Engine to develop a drone simulator in which the configurations of the actual drone are included, such as the flight status, distance, speed, and altitude. The virtual drone can detect the distances between itself and the surroundings in real-time and export the results to the external program for data processing. The ambient light and climate in the simulator can be changed accordingly.

In the simulator, users can conduct perspective-switching and path planning, as well as change scenes. It also simulates various environmental conditions. In order to show the typical scenes of remote construction, this article chooses “City Park” as the simulation field in this phase.

![Simulator images](image)

**Figure 3. (a) Distance detection system (b) Third-person perspective (c) Path planning**
3.4. NUI - BODY CONTROL AND SENSING

3.4.1. Responsive Control System (RCS) - Vest As a Passive Controller

When facing an emergency, humans will respond physically based on experience to avoid disasters (Dunsmoor and Murphy, 2015). Similarly, the operator will also perform body dodge actions when experiencing a dangerous state. Based on these phenomena, this research develops a Responsive Control System (RCS) that will be activated when the situation is severe to prevent accidents. Besides, it is difficult for the operator to clearly understand the actual distance between the aircraft and the surrounding obstacles through the visual display in dynamic flight.

Research has shown that haptic feedback can enhance environmental perception and effectively trigger the operator's emotions. In addition, jackets and vests can give users a multi-part perception of the body; the pressure's position, intensity, frequency, and temperature can correspond to various feelings defined in the emotional list of the social psychology model (Arafsha et al., 2012). Therefore, this paper applies a developed haptic vest, which can control the size of the airbag according to external information, and integrates it into the drone simulator to enhance the operator's understanding of the flight environment.

The haptic vest we used can activate specific airbags through inflating for pressuring on the pilot's body to achieve the function of direction warning when the environment is severe. The obstacle avoidance system in the simulator can instantly detect the environment around the aircraft; the detection range is set as five meters as a default, which can also be adjusted according to different aircraft types, velocities, and detection principles.

During the process, the RCS sends the detected distance value to the external program file for data processing. The processed results are sent to the specific motors on the vest through the Message Queuing Telemetry Transport (MQTT) protocol to inflate the specific airbags. Finally, the current airbag inflation rate has been optimized to 0.2 seconds to minimize the latency.

In order to quantify the swing of current body posture, this research sets a micro:bit on the user's shoulder, which is embedded in the vest. In addition, the micro:bit serves as the accelerometer for capturing body posture and the device for wireless communication. Thus, if the RCS detects a drastic change of the operator's body posture when the airbag is activated, it will immediately pause the drone's flight.

3.4.2. Body Posture Control System (BPCS) - Vest As an Active Controller

The vest acts as an active controller in the BPCS. It obtains the current data of body posture in real-time through the micro:bit set on the shoulder; sends the data to the external program file by setting a specific radio channel for data processing. The processed results are sent to the simulator for controlling the drone. After validation, the values of each axis captured by the accelerometer can be effectively mapped to the drone's control parameters (Roll, Yaw, Pitch) to change its flight path to achieve the BPCS.
It has also been tested that the captured value can effectively correspond to the body posture. Moreover, we have successfully optimized the update rate of the communication system to 30 data per second to minimize the latency.

3.5. NUI – VOICE CONTROL SYSTEM (VCS)

3.5.1. Speech Recognition System

In order to allow AEC personnel to carry out simple tasks and free up their hands easily, this research developed a Voice Control System (VCS) for drones. This system currently uses Google Speech Recognition as a standalone subsystem for Speech To Text. Besides, the operator can control the drone by simply speaking the requirements, such as flight parameters (Roll, Yaw, Pitch) adjustments, path planning, perspective mode-switching, screenshots; the system will automatically capture the keywords from the spoken instructions and execute them.

3.5.2. Speech Command Dictionary

The Speech Command Dictionary refers to the professional dialect in the AEC domain and on-site communication terms. Besides, this research also studies how the designers communicate with remote workers in simple, clear, and understandable terms (Christenson et al., 2013). At this phase, three types of functions have been developed, including basic controls, functions, and various types of path planning. The VCS can recognize sentences containing keywords; when the keywords are detected, the VCS will execute the corresponding commands.

During the whole process, the VCS can continuously recognize the spoken instructions during the flight and decide whether to execute them based on the weights of the instructions. For example, the system will overwrite the previous command if the drone receives a stop command while flying to the target; if the drone receives a screenshot command during the flight, the system will continue to fly and take a screenshot. The following examples illustrate several commands and actual functions.

- **Take off** = Move up 10 meters above the ground to stand by.
- **Move right** = Adjust the degree of the Yaw axis of the drone.
- **Circular movement on building B and screenshot** = Use the current distance between the drone and the target building as the radius to circle around and take a
screenshot.

- Switch to FPV = Switch the screen of the simulator to first-person perspective.

<table>
<thead>
<tr>
<th>Basic control</th>
<th>Function</th>
<th>Path planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take off</td>
<td>Screenshot</td>
<td>Linear movement to building A</td>
</tr>
<tr>
<td>Land</td>
<td>Start recording</td>
<td>Circular movement on building B</td>
</tr>
<tr>
<td>Turn right</td>
<td>Stop recording</td>
<td>SemiCircular movement on building C</td>
</tr>
<tr>
<td>Turn left</td>
<td>Switch perspective</td>
<td>Grid movement on building D</td>
</tr>
<tr>
<td>Move forward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move backward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Speech command dictionary

3.5.3. Path Planning and Photography

By designing a Speech command dictionary, the operator can use voice to give the drone flight commands. This paper refers to the current UAV software and proposes the following five types of movements in the preliminary stage.

Figure 6. Types of path planning

Figure 7. (a) Linear movement (b) Circular movement

4. Tests and Preliminary Feedback

This research plans a set of application scenarios for the AEC domain, which provides architects, planners, and practitioners abilities to perform remote operations with
intuitive body posture and voice to obtain a better spatial presence and experience. First of all, this research conducts a user test involving five professionals who work in the AEC field to verify the feasibility of using the drone NUI to assist remote tasks. Secondly, we described the goal of the test and show them how to use the NUI. During the process, the words and body postures they used are recorded to verify the system's availability deficiency, and the user's response and feedback will also become the focuses of the next stage.

The test requires users to fly around the building to take pictures. During the process, they need to set the origin, take off, approach the building, switch perspective, stop, screenshot, fly back to the origin, and land.

Finally, the users provided a lot of feedback and expectations, including that (1) it is expected operators in different regions can control the same UAV at the same time in the future; (2) it is expected the system can automatically perform tasks by observing the dialogue between users rather than by instructions; (3) it is expected the system can record the last flight path and perspective to execute tasks repeatedly; (4) it is expected the practitioners at the remote site can know who the drone operator is to increase interactivity; (5) it is expected this system can bring users to visit any indoor space, making the virtual tour experience more immersive.

5. Conclusion and Future Work

In summary, this research has developed a drone simulator and proposed a NUI strategy that combines RPS, BPCS, and VCS to allow operators to interact with the drone intuitively. Moreover, compared with the previous video conferencing and remote monitoring software, the proposed system allows users to actively control the viewing angle, which can help spatial designers improve the perception of remote space. Furthermore, users can perform multi-modal interaction with UAVs by switching the proposed NUIs to meet their application requirements. The use of NUI for HDI (Human-drone interaction), especially for tasks in the AEC industry, enhances spatial presence and expands its scope of application; We expect the system can recognize requirements automatically in the next step to achieve a more natural experience.

However, the current simulator still has room for improvements, including the efficiency of speech recognition and the limitation of the Speech command dictionary. Therefore, this research will apply natural language processing (NLP) to improve the VCS and apply the proposed system to the physical drone for validation in the next stage.

Lastly, the application of this research can be used in AEC industries, building and bridge inspections, progress monitoring, urban planning, architectural photography, and surveying, as well as taking guests and customers to any spaces for virtual tours and become a new type of telepresence robots in the future. In the future, if our simulated environment can precisely match the physical drone flight, the proposed NUI will be of great benefit to enhance the interactivity of the drone control system.
Acknowledgements

This research is supported by the Ministry of Science and Technology, and the grant number is MOST 110-2634-F-009-018; and thank the Architectural Informatics Lab (GIAAIL), National Yang Ming Chiao Tung University for technical support and assistance.

References


