Research Article

Immersive Gesture Interfaces for Navigation of 3D Maps in HMD-Based Mobile Virtual Environments

Yea Som Lee [ID] **and Bong-Soo Sohn** [ID]

School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea

Correspondence should be addressed to Bong-Soo Sohn; bongbong@cau.ac.kr

3D maps such as Google Earth and Apple Maps (3D mode), in which users can see and navigate in 3D models of real worlds, are widely available in current mobile and desktop environments. Users usually use a monitor for display and a keyboard/mouse for interaction. Head-mounted displays (HMDs) are currently attracting great attention from industry and consumers because they can provide an immersive virtual reality (VR) experience at an affordable cost. However, conventional keyboard and mouse interfaces decrease the level of immersion because the manipulation method does not resemble actual actions in reality, which often makes the traditional interface method inappropriate for the navigation of 3D maps in virtual environments. From this motivation, we design immersive gesture interfaces for the navigation of 3D maps which are suitable for HMD-based virtual environments. We also describe a simple algorithm to capture and recognize the gestures in real-time using a Kinect depth camera. We evaluated the usability of the proposed gesture interfaces and compared them with conventional keyboard and mouse-based interfaces. Results of the user study indicate that our gesture interfaces are preferable for obtaining a high level of immersion and fun in HMD-based virtual environments.

1. Introduction

Virtual reality (VR) is a technology which provides users with software-created virtual 3D environments that simulate physical presence of users to provide immersion [1]. A great deal of research has been performed to enhance the realism of VR by making the user's actual motion match the real-time interaction with virtual space [2, 3]. In 1968, Sutherland invented a head-mounted display (HMD), and other VR devices have since been developed to stimulate a user's vision and movement. The HMD, which is now a commonly used VR device, is a glasses-type monitor worn on the head. HMDs are currently attracting a huge amount of attention from industries and users since they provide the VR experience at an affordable cost. HMDs provide a high level of immersion through (i) a stereoscopic display, (ii) wide viewing angles, and (iii) head orientation tracking. Because of the above advantages, HMDs can be utilized in various fields such as education [4, 5], medical treatment [5–7], and entertainment.

3D maps [8] such as Google Earth [9] and Apple Maps (3D mode) allow users to see and navigate 3D models of real worlds in a map. With the recent development of automatic 3D reconstruction algorithm applied to satellite images and mobile environments, high-quality 3D maps of places have become accessible in a wide and ubiquitous way, such that any remote user can explore any place with great realism. However, in most of these cases, the 3D maps can be experienced on a two-dimensional flat screen. Research on virtual maps can also be used to visualize statistics regarding climate change and population density, or to display topographical maps, building drawings, and information in augmented reality. This means that methods that utilize HMDs, rather than conventional monitors, for 3D map navigation are valuable.

However, since there is a limit to the sense of reality imparted by the HMD device, it is necessary for the user to adopt a technique to explore and perceive a virtual space just like a real space. Virtual reality programs running on a PC

usually use traditional input devices such as a keyboard or a mouse, but this has the disadvantage that it does not match the behavior of the user in a virtual environment. Because computational speeds are limited, there is a time difference between a user's movements in physical space and movement in the virtual environment. In a VR environment, the time difference between the movement of the user and movement in the virtual space interferes with the immersion and causes dizziness [6, 10, 11], consequently reducing interest. For this reason, research has been needed to increase the immersion and interest in VR by adjusting the input methods to directly match body movements. Also, in order to maximize the satisfaction of the user, an intuitive interface method for the user in the virtual space is required.

Accordingly, development of devices related to the operation interface has been actively carried out in order to compensate for the disadvantage caused by the relative inability to use conventional input/output devices in the VR environment. A device that recognizes the user's motion is a haptic-type device [12], which is generally held in the hand, and the avatar in the virtual space has the ability to guide the user's desired motion naturally and without delay. For these reasons, joysticks or Nintendo Wiimote [13–16] controllers have been developed as control devices that replace the keyboard. As a result, there are more and more cases where appliances that recognize the movement of the user and improve the accuracy of the operation are utilized in the game or the virtual space. However, there is a limitation in maximizing the gesture recognition and immersion using the whole body of the user based on the position sensor of the haptic device in the virtual space. From this motivation, we designed and implemented various realistic gesture interfaces that can recognize user's gestures in real-time using Kinect to reflect the user's movements in an HMD-based virtual environment. In addition, we measured the usability of the proposed gesture interface and the conventional control interface based on the keyboard and mouse, and compared the advantages and disadvantages of each interface through a user study. Figure 1 shows a user wearing an HMD using customized motion recognition system, while experiencing a given virtual environment.

In this paper, we design and implement immersive gesture interfaces that are recognized in real-time using the Kinect depth camera. The position of each joint is identified and analyzed to allow the gesture of the user in the virtual environment to reflect the actual physical gestures. The degree of user satisfaction, including the degree of interest and ease of use, was checked according to the manipulation method. The main contributions of our paper are as follows.

(i) We designed and implemented immersive gesture interfaces with integration of flyover (bird, superman, and hand) and exploratory (zoom, rotation, and translation) navigation, which is recognized in real-time through the Kinect camera for HMD-based VR environments.

(ii) We evaluated the usability of the proposed gesture interfaces and conventional keyboard/mouse-based interfaces with a user study. Various usability factors
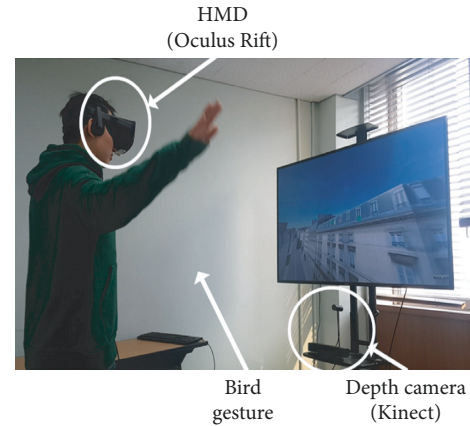

FIGURE 1: The virtual environment system implementation used in this paper. The user wearing the HMD (Oculus Rift CV1) performs an action similar to that of a bird, which the Kinect recognizes. In a virtual environment (Paris Town), the user feels immersed like a bird.

(e.g., immersion, accuracy, comfort, fun, nonfatigue, nondizziness, and overall satisfaction) were measured.

(iii) We analyzed the advantages and disadvantages of each interface from the results of the user study.

As a result of the user study, it can be demonstrated that the users prefer the gesture interface to the keyboard and mouse interface in terms of immersion and fun. The keyboard interface received high marks for accuracy, convenience, and unobtrusiveness. These results confirm that the method of manipulating a virtual environment affects the usability and satisfaction regarding the experience of the virtual environment.

The remainder of this paper is organized as follows. We discuss related papers in Section 2. The design and method of the proposed gesture interfaces are described in Section 3 and Section 4, respectively. Section 5 describes the user study design and the results of the user study. Section 6 discusses our conclusions.

2. Related Work

One of the main goals in VR research is to increase the sense of immersion. Mass-market HMDs are becoming popular because they can provide a high level of immersion at an affordable cost. With the emergence of a need for immersive movement control [17], companies that produce HMD devices have recently been introducing game controllers with auxiliary functions (e.g., Oculus Touch) [18]. The HMD was initially invented by Ivan Sutherland in 1968 [19], but it was initially difficult to commercialize for many reasons, including the high cost, heavy weight, space limitations for installation, and a poor display. The biggest problem was the limitation of the display technology [20]. HMDs are divided into two types, a desktop and a mobile VR, depending on the size of the image that can be processed and the complexity of the structure. Mobile VR is hosted and ultimately displayed on a mobile phone, and there is no real restriction on the range of movement because it is wireless.

Recently, IT companies have been developing a variety of products by studying and developing interfaces for HMDs that provide high immersion and allow for smooth and seamless user interaction. Desktop VR is widely used for research purposes. As the computational power and display resolution of smartphones increase, companies have developed diverse content using Mobile VR, rather than Desktop VR platforms. As HMD technology has progressed, HMDs have been used in various fields such as education [5], medical care [5–7], and architecture. Kihara et al. [21] conducted a study and experiment on laparoscopic surgery using an HMD and verified the feasibility of using HMDs in the medical field. It is now possible to use an HMD to minimize laparotomy incisions, instead of using abdominal laparotomy or high-cost robotic surgery systems, in which a large scar may remain, with an increased risk of infection. The surgeon wears an HMD, and the system provides a 3D image, depth map, and tactile feedback associated with the affected area, and performs a safe operation. In addition, varied research is being conducted to recognize the facial expressions of users using an HMD and to simulate these expressions in a virtual environment [22].

Research on the interaction between humans and computers has been studied in earnest as soon as personal computers became available. The HCI (human-computer interaction) [23] aims to allow people to use and communicate with a computer in a human-friendly manner. As the use of computers increases, HCI is carefully considered in the development of computer-user interfaces (UI) [24, 25]. For this purpose, a study has been conducted on an interface using body gestures rather than the conventional input devices [26]. The main difference from previous HCI-related researches is that our approach focuses on improving the level of immersion in an HMD-based virtual environment for designing navigation interfaces in addition to other important usability factors such as the level of accuracy, fun, and comfort.

Humans have the ability to make emotional expressions using the body and to allow meaningful behaviors to take the place of language [24, 25]. Gesture recognition can be applied to various fields, such as sign language [6, 12], rehabilitation [13, 15, 27], and virtual reality, and is easy to utilize in computer applications. In particular, a meaningful gesture using the body refers to expressible behavior related to the physical movement of a finger, hand, arm, leg, head, face, or body. The main purpose of human gestures is to communicate meaningful information or to interact with the surrounding environment. However, since the various operations used for this purpose may overlap or have different meanings, it is necessary to sufficiently study the development of interface technology based on gesture recognition. Unlike existing keyboard and mouse input devices, it is necessary to search the body part using sensors and to recognize the operation after tracking the position [24, 25, 28].

In particular, a device such as a joystick, which can be used as a substitute for a keyboard and a mouse, can be used to increase the user's sense of immersion. Since the effectiveness of the hand manipulation method has since been verified, controllers such as the Kinect [23, 26, 28, 29] and Leap motion [28, 30] have been released. As games and applications that can be experienced in a VR environment have been developed, it has been confirmed that the act of controlling the virtual space through the movement of the body plays an important role in making VR realistic and immersive. In addition, various methods for recognizing user's movements have been studied [31–33].

As mentioned, the haptic-type device has been developed, held in the user's hand, in order to reflect the user's gestures in such a manner that the user can easily forget the difference between the virtual reality and the real world [12]. The keyboard, mouse, joystick, and similar traditional input devices can be used to move around in virtual space by holding the device with a hand or by wearing it. However, these conventional devices have limitations. The haptic device increases the probability of accurately recognizing the user's motion, but it can limit the range of motion, and consistently wearing the haptic device can be troublesome [12]. In addition, it requires time to learn a formal haptic device operation method [34], and it is insufficient to realize the virtual reality realistically because it is manipulated while holding it in the hand or wearing it directly.

For these reasons, in this paper we have developed immersive and intuitive gesture interfaces to control the navigation in a virtual environment for HMD users. In particular, we deployed simple algorithms to recognize natural gestures in real time. Preliminary results of this paper have appeared in [35, 36]. The main differences are the integration of gestures for flyover and exploratory (e.g., zoom/rotation/translation) navigation and a detailed description of the formal user study results.

3. Design of Immersive Gesture Interfaces

As the need for immersive interfaces to replace traditional input/output devices for HMD-based VR navigation increases, related research has been actively conducted. For this purpose, Microsoft Kinect, which contains a low-cost depth camera, can be used to track and recognize the user's body gestures in real-time and control navigation in the VR environment while wearing an HMD. We developed a VR software system, in which a user can experience a virtual reality through the Unity3D Engine that supports the simultaneous utilization of the Kinect and Oculus Rift. We also defined two types of immersive gesture interfaces, as well as conventional keyboard and mouse-based interfaces. There are six types of gesture interface methods that are proposed in this study. The proposed gesture interfaces that are recognized using the Kinect can be seen in detail in Figures 2 and 3. The location of each joint and body skeleton segment that connects the joints are extracted using Kinect SDK, as shown in Figure 4. These are then used for the real-time recognition of gesture types and intensities.

Most people use their hands when accurately controlling objects, such as when driving a car or playing a PC game [25]. We considered a natural gesture interface that tracks the location and movement of hands since the keyboard and mouse are also hand-based input devices. Because the ratio of right-handed people is high in general, we defined gesture interfaces that primarily use a right hand [37, 38]. The
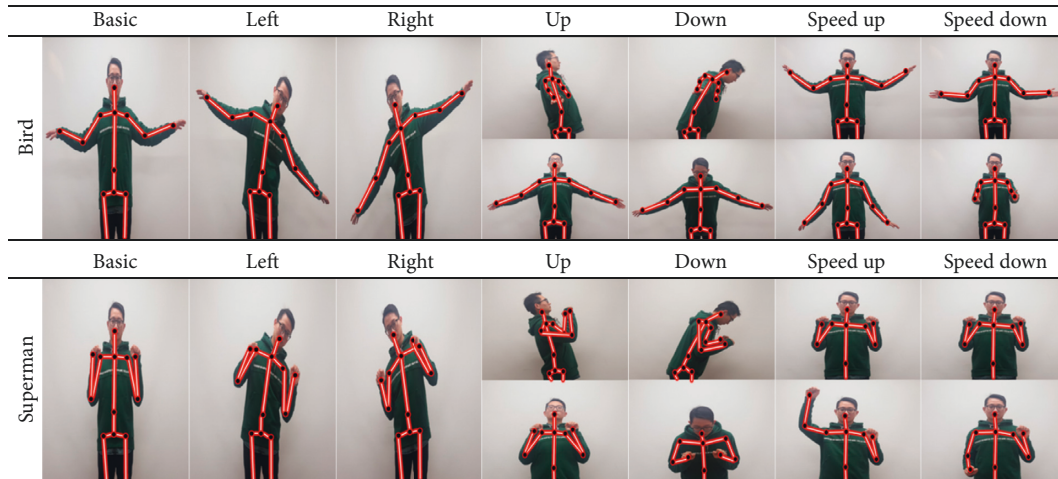
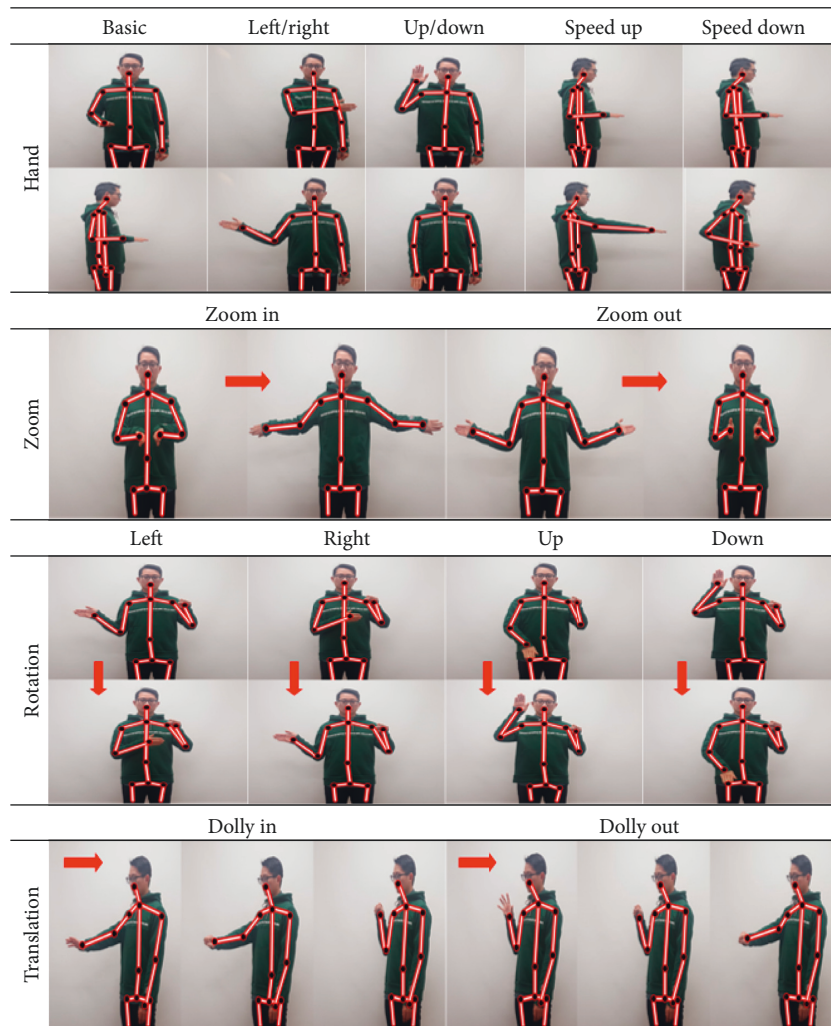Figure 2: Flyover gesture interfaces (bird and superman).

Figure 3: Exploration gesture interfaces (hand, zoom, rotation, and translation).

navigation interface implemented in this paper defines bird, superman, and hand gestures as flight mode operations [39] through the tracking of the user's movement with the Kinect. Our gesture interface also supports exploratory navigation

features that are provided in Google Earth, such as zoom, rotation, and translation.

For thousands of years, humans have dreamed of being able to fly like a bird. Rheiner developed a VR simulator,
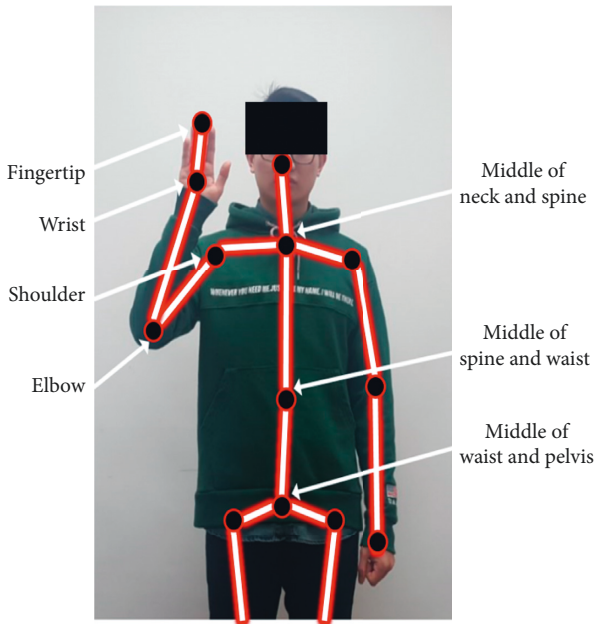
FIGURE 4: Location of joints and body skeleton segments that are recognized through the Kinect.

called *Birdly*, in which a user can experience flying through the 3D space with the Oculus Rift [40]. The user can navigate the Birdly simulator using hands and arms making a waving action that pantomimes the movement of bird wings in 3D. However, since this simulator is bulky and requires significant production costs, it is burdensome for a general user to possess it at home. Also, flying in the sky like a superman-like hero is hard to achieve. Therefore, we implemented a new and superman-like motion interface to implement a gesture interface that is difficult to otherwise experience, giving users a surrogate satisfaction.

3.1. Flyover Navigation. We aim to make certain that our gesture interfaces: (i) allow a simple and natural action for flyover control that is similar to actual flying behavior, (ii) are recognized in real-time by a low-cost motion sensor, such as the Kinect depth camera, and (iii) enhance the degree of immersion, which is unique to the HMD-based virtual environment. For this purpose, we designed three gesture interfaces (i.e., bird, superman, and hand) for the flyover navigation. The scales of these three gesture interfaces are different (i.e., bird > superman > hand) such that we can understand implicit relationship between usability properties and the scales. The detailed gestures for each interface are shown in Figures 2 and 3, and can be described as follows.

3.1.1. Bird. The user can adjust the direction by moving the body up, down, left, and right keeping the waist in the basic posture with both arms open, similar to bird wings. In the basic posture, both arms move up and down simultaneously to accelerate, and both arms can be stretched forward at the same time.

3.1.2. Superman. As shown in Figure 2, hold both hands on both sides of the face at the level of the shoulder line. Move the upper body in the direction to move. Move the body back and forth to go up and down, respectively. When a user wants to adjust the speed, the user can accelerate or decelerate by moving his or her right hand up or down, respectively.

3.1.3. Hand. Initially, the right hand is set as the reference point and the right hand is placed in the front of the body in a comfortable position, and then held at the initial reference position for 2-3 seconds. The user can manipulate the direction by moving his or her hands vertically or horizontally and can decelerate or accelerate the speed by moving the hand back or forth, respectively, as shown in Figure 3.

3.2. Exploratory Navigation. Figure 3 shows the proposed gesture interface for 3D map navigation. The defined gesture interface is based on Kinect recognition instead of using a keyboard and a mouse. It implements the operation of moving left/right/up/down, speeding up/down, zoom-in/out, rotation, and translation, which are typical features of the interface provided by Google Earth. The hand interface can be manipulated vertically and horizontally with the right-hand position as the reference point at the first execution, and the hand is moved back and forth to adjust the speed.

3.2.1. Zoom. The user can control zoom-in or zoom-out, which allows seeing objects either closer or farther away. For the zoom in motion, both arms are stretched straight ahead and then the arms are opened outward. This action gives the feeling of enlarging the space while maintaining symmetry about the body. In an opposite manner, for the zoom out motion, both arms start out to both sides, and are brought together in front of the body, keeping the symmetry as both arms are collected in front of the body.

3.2.2. Rotation. The user can rotate the screen in four directions. The user can think of the left hand as a globe and use the right hand to rotate it in the desired direction while holding the fist with the left hand.

3.2.3. Translation. This is an interface that allows one to move quickly to the desired location in the current VR environment and operates with the right hand only. The user has to move the right hand to the location to which he or she wishes to move and hold the fist at that position. The 3D map is enlarged or reduced as the user pulls or pushes the hand in the direction he or she wants to move, using the position of the right hand holding the fist as a reference point. The corresponding action of translating away from a location ends when the right hand with a fist is fully extended, and the fist is released.

The zoom and translation interfaces are similar but operate on different principles and differ from the actual moving subjects. Zoom is a function to zoom in or out of the current VR environment, and the translation interface moves the map such that the user is closer to or farther from the user's starting point in the map.

Table 1: Left/right and up/down angles of gestures.

Gesture interface	Left/right angle	Up/down angle
Bird	Angle between x-axis and the line connecting the left/right hands	Angle between y-axis and the line connecting the head and spine
Superman	Angle between y-axis and the line connecting the head and waist	Angle between y-axis and the line connecting the head and spine
Hand	Angle between the x-axis and the line connecting the right hand and right elbow	Angle between the z-axis and the line that connects the right hand and elbow

4. Recognition of Immersive Gesture Interfaces

In order to accurately recognize the meaningful behavior of the user, it is necessary to be able to track the position of the body features. Generally, there exist methods of learning to recognize body parts such as the face or hand in a photo, through Big Data Machine Learning [41]. However, it is difficult to recognize body parts in real-time because even using state-of-the-art algorithms optimized through machine learning, the classification of 3D body parts involves a nontrivial, potentially sickness-causing delay. The Kinect is a device that provides the ability to track a human joint using a depth camera. Skeleton points that are primarily used in this study include the human body parts of the hand, wrist, elbow, and shoulder. We also used a method to calculate the position of the center of the palm to accurately track the state of the hand (fist, palm, etc.) [42].

We utilized the depth map captured by the Kinect infrared projector sensor and Kinect SDK modules to track the location of feature points and extract a skeleton from the human body that was captured in a depth map. For the recognition of gesture types and intensity in bird, superman, and hand interfaces, we define the left/right and up/down angles as shown in Table 1.

We also utilized the Kinect to implement functions that Google Earth supports to navigate 3D maps. With Google Earth, one can perform zoom, rotate, and translate operations using the mouse and keyboard interface to navigate to the desired location in 3D models of buildings and terrain. While building a virtual environment for experiments, we implemented navigation functions that replace the traditional input devices, the mouse and keyboard functions.

Algorithm 1 describes the recognition of gesture interfaces and their magnitude defined in Figure 2.

In order to change the user's left and right direction, the angle between the x-axis (line 1) and the straight line between both hands is compared (line 2), such that the left and right movement is possible (line 3-4). To move up and down, it is necessary to calculate the angle between the line connecting the y-axis and the body part (line 5), and compare the angle (line 6), such that the line can be moved up and down (line 7-8). If the result obtained by calculating the difference from the previously measured distance from the current reference point distance is greater than the

```
A: left/right angle
D: up/down angle
HL: horizontal threshold
R: right hand
L: left hand
CA: left/right magnitude
VL: vertical threshold
CD: up/down magnitude
MD: distance threshold
 (1)    if A > HL
 (2)       if R > L
 (3)          CA = −A;
 (4)       else CA = A;
 (5)    if D > VL
 (6)       if R > L
 (7)          CD = −D;
 (8)       else CD = D;
 (9)    else CD = 0;
(10) if Distance - MD > SpeedUp
(11)      accelerate;
(12) else if Distance - MD < SpeedDown
(13)      break;
```

Algorithm 1: Recognition of navigation gestures and their magnitude.

acceleration threshold (line 10), then the speed is increased (line 11), and otherwise the speed decreases (line 12). We experimentally found that it was the best choice for setting horizontal and distance threshold to 0.4–0.7.

Algorithm 2 describes the rotation interface, defined in Figure 3 alongside the samples of zoom and translation interfaces. These operations basically consist of only the values of x and y subtracted by the z value, when the difference between the right hand and the right shoulder is smaller than a predefined threshold (line 1), and the z coordinate should be 0. When rotation or translation occurs (line 3), the degree of the change is shifted by the difference of the right hand, which is changed from the position of the right hand (line 4). When moving in the virtual space, the position of the current right hand becomes the position of the reference hand (line 6-7). When we rotate based on the horizontal and vertical lines (line 9), the values of the horizontal line and the vertical line are added respectively (lines 9-11). The current rotation position is 0 (line 12); only the x-value and y-value are converted at that position (line 13).

Figures 5–7 show details of the zoom, rotate, and translate interface algorithms for tracing joints of depth cameras. The red circle represents the state of the fisted hand, and the green circle represents the palm of the hand. The gray circle implies that some parts of the body may overlap, making it difficult to represent the exact position value.

5. User Study and Results

5.1. User Study Design. For evaluation of our proposed interfaces and for a comparison, we developed VR software based on a 3D map and investigated user responses. We used two 3D datasets, a Grand Canyon model and a French Town

```
R: right
L: left
H: hand
Rot: rotation
c: current
Ho: horizontal line
Ver: vertical line
 (1) if (LH == Fist and | LH.y–LSh.y | < 0.1f)
 (2)    RH.vector = RH.(x, y, 0);
 (3)    if Rotation
 (4)        add.Rot = RH.vector–cRH.vector;
 (5)        RotHV (add.Rot.x * 100, add.Rot.y * 100);
 (6)    else Rotation = true;
 (7)    cRH = RH;
 (8) else Rotation = false;
 (9)    RotHV (Ho, Ver)
(10)       cRotH += Ho;
(11)       cRotV += Ver;
(12)       Rot.(cRotV, 0, 0);
(13)       Rot.(0, −cRotH, 0);
```

ALGORITHM 2: Rotation interface.
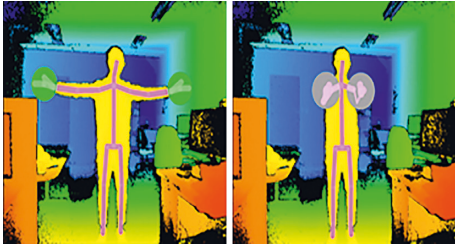
FIGURE 5: Kinect depth image with body skeleton representing zoom gesture interfaces.
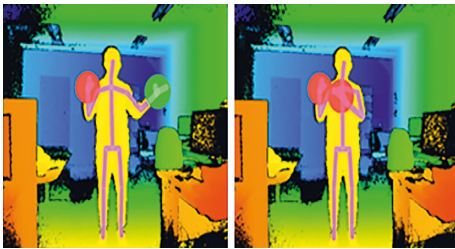

FIGURE 6: Kinect depth image with body skeleton representing rotation gesture interface.
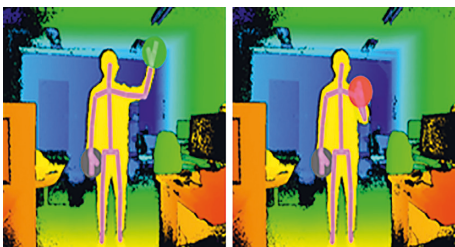

FIGURE 7: Kinect depth image with body skeleton representing translation gesture interface.

model as our test virtual environments (Figure 8). We chose the Oculus Rift (Consumer Version 1) and Microsoft Kinect (Version 2) as the test HMD device and motion sensor, which are relatively affordable for the general public. The VR environment was tested on a desktop PC equipped with an Intel i7 3.6 GHz CPU and 16 GB main memory.

The HMD-based VR software system for navigation was developed with Unity3D [43]. Our method for gesture recognition was developed using the Kinect SDK and Toolkit, distributed by Oculus and Microsoft. As a way to experience the environment for this user study, users could fly in the test virtual environments like birds and superman, and navigate using the right hand. We also made a scenario consisting of zoom, rotation, and translation navigations in the test.

The subjects were 23- to 31-year-old, 12 college students (10 males and 2 females) in the computer engineering department of our university. In order to confirm the clear difference between the existing interface and the proposed interfaces, we conducted a questionnaire to evaluate and quantify an experience index and usability score of each method. Experiments with HMDs were applied to the Grand Canyon model and French Town model, and experiments were conducted with six gesture interfaces and two interfaces based on the keyboard and mouse. For each participant who had never worn the HMD before or who complained of dizziness, we gave a rest period of 1 to 10 minutes between each experience depending on the degree of dizziness [11, 44].

The purpose of this study is to identify the necessity of the gesture interfaces that are needed to replace the existing keyboard manipulation method, through studying the development of technology that can enhance the satisfaction of experiencing a virtual space. We designed the user study to analyze advantages and disadvantages of the proposed interface compared to traditional interface and to verify the significance of the results.

5.2. Experimental Results. From the experiments, each of the 8 usability properties experienced in the two scenarios of the Grand Canyon and the French Town model (e.g., overall satisfaction, accuracy, ease of operation, comfort, immersion, and fun) were quantified in the questionnaire results. In Figure 9, we can see a picture of the average scores for the user's overall satisfaction with the above-mentioned 8 properties evaluated with scores ranging from 1 to 5. The graph starts from the middle (i.e., score 3) because it can better show whether it belongs to good (i.e., to the right from middle) or bad (i.e., to the left from middle) scores. Overall, the degree of fun was the highest, and the scores of other properties were generally good but subjects experienced significant dizziness when using the gesture interfaces.

The results obtained from the Grand Canyon and the Village model differed slightly. The bird interface scored high in the overall satisfaction, and the hand interface scored relatively high in the accuracy. The keyboard and mouse are the easiest to operate and can be redirected with fewer movements, resulting in greater convenience, nonfatigue, and nondizziness. The bird and hand interface is difficult to
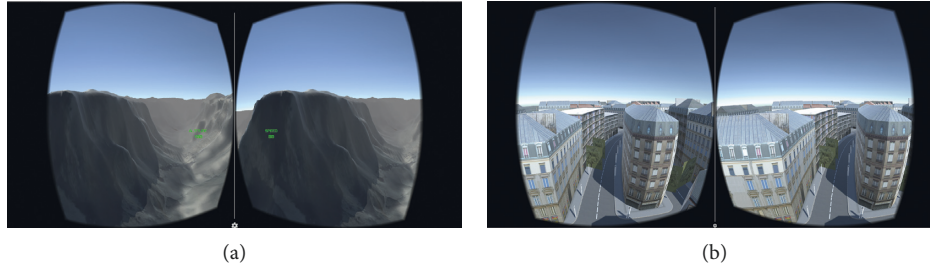
(a)   (b)
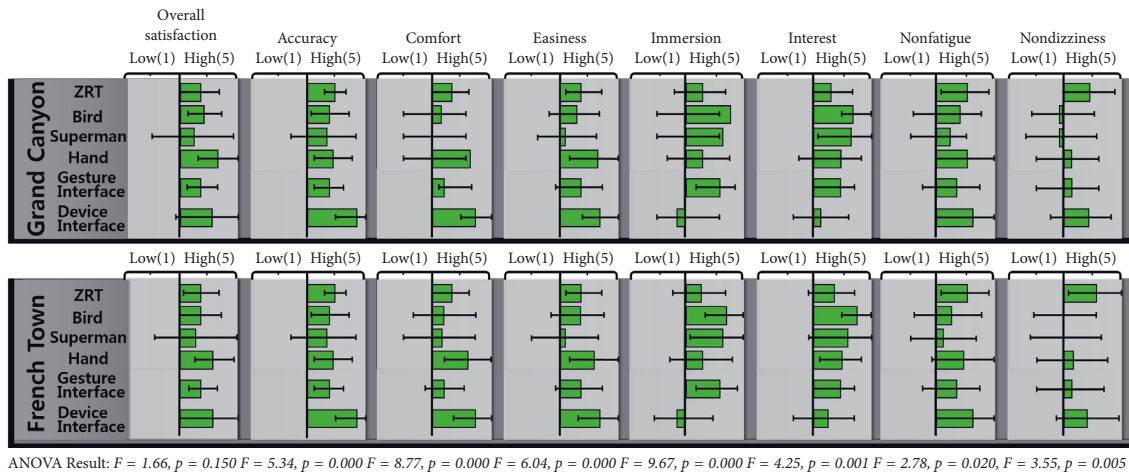
FIGURE 8: Test virtual environments. (a) Grand Canyon model. (b) French Town model.

ANOVA Result: *F = 1.66, p = 0.150 F = 5.34, p = 0.000 F = 8.77, p = 0.000 F = 6.04, p = 0.000 F = 9.67, p = 0.000 F = 4.25, p = 0.001 F = 2.78, p = 0.020, F = 3.55, p = 0.005*

FIGURE 9: Survey results of our proposed and conventional device interfaces for test virtual environments, (a) Grand Canyon and (b) French Town. The green bars represent the average scores and horizontal whiskers represent standard deviations. The middle vertical line means mid-score (3). The result of the statistical test (ANOVA) is marked below the graphs. Significant difference exists when $P(Sig.) < 0.05$.

manipulate, but has a high score on the degree of fun and immersion. Fifty-eight percent of students prefer to use gesture interfaces that use both hands at the same time, rather than to use one hand. Sixty-seven percent of students responded that it was better to use gesture interface rather than the keyboard and mouse interface. In addition to this, 92% of students liked to wear and experience the HMD instead of the monitor when asked what kind of screen offers better realism.

In order to verify the significance of the experiments conducted in this paper, a one-way ANOVA and Scheffé tests were performed, and the significance was verified in Figure 9. The significance level between each interface and the evaluation items was less than 0.05 for the remaining seven items except satisfaction. At the significance level of 5% (Sig. < 0.05), the null hypothesis was rejected and the alternative hypothesis was adopted. Thus, it is justifiable that the difference of usability between the proposed and existing interfaces is significant. As a result, there was a significant difference in accuracy between the device interface and superman (Sig. = 0.002), gesture interface (Sig. = 0.006), bird (Sig. = 0.017), difference between hand (Sig. = 0.007) and device interface (Sig. = 0.000). In the easiness factor, there was a difference between superman interface and hand (Sig. = 0.011) and device interface (Sig. = 0.000). The immersion factor showed significant

differences between device interface and gesture interface (Sig. = 0.001), superman (Sig. = 0.000), bird (Sig. = 0.000) and between ZRT (zoom, rotation, and translation) interface and bird interface (Sig. = 0.035). In the interest, there was a difference between device and bird interface (Sig. = 0.006).

We observed that the keyboard interface has a higher score in terms of accuracy, comfort, and easiness, compared to gesture interfaces. On the contrary, the difference in the gap of scores between the gesture and keyboard interfaces is very large in the factors of immersion and interest.

In Figure 10, we can see that the two virtual map environments, Grand Canyon and French Town, affect user preference. Overall, the user's score for the two virtual map environments did not appear to be significant, but the overall satisfaction of the hand interface was very high in the Grand Canyon, while the overall satisfaction of the Superman interface was the lowest. However, in the French Town, the overall satisfaction with the hand interface and the interface using the keyboard and mouse was the highest, and the overall satisfaction scores of the rest of the interfaces were similar.

Since the sample size (i.e., 12 participants) is relatively small and test scenarios are rather simple, further research
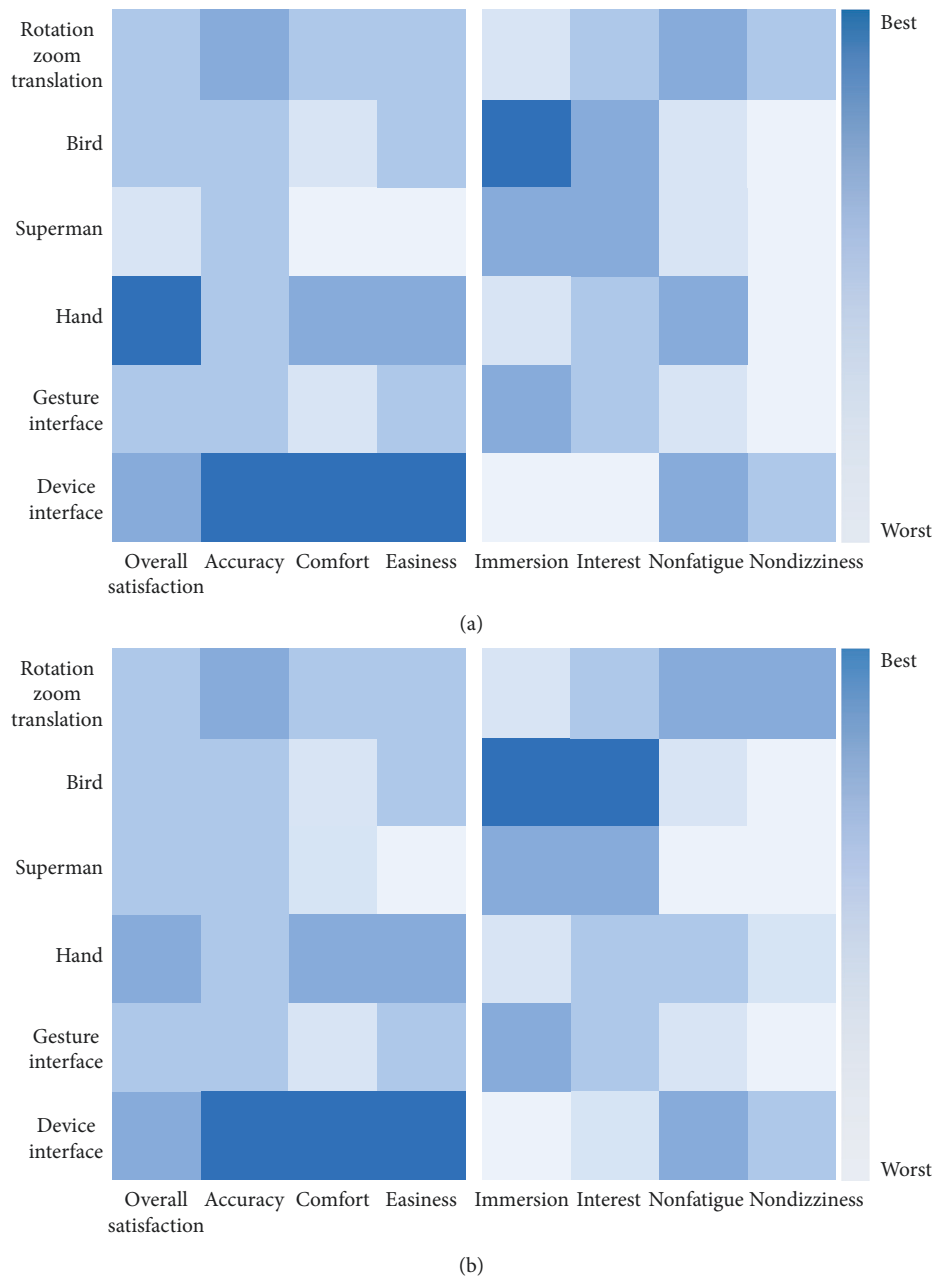
Figure 10: Visualization results of usability score distribution for each interface and usability properties for (a) Grand Canyon and (b) French Town model.

can be necessary to generalize and verify the usability of our method.

6. Conclusion

The results of this study indicate that the method of gesture recognition through body motion can provide a higher level of immersion than the conventional keyboard/mouse method. Since users experience an interface with which they are not familiar, it is necessary to learn the operation method and have time to adapt before the first execution. However, after a very short learning period, users were able to experience virtual reality more effectively. It is desirable to use the Kinect-based gesture interface for a higher level of immersion and fun. However, with long periods of VR use, users tend to become easily tired, and further research must be conducted to overcome this drawback. The results of this study show that it is more interesting and fun for the user to use his or her body to manipulate 3D space and navigate 3D environments, but the interface method can be different according to the type of scenario space. Considering the level of immersion and interest, it is necessary to research intuitive methods to perform operations that can easily make future human/computer VR interactions more easy and natural. Combination of gestures and speech recognition techniques can improve the usability of control interfaces.

Hence, we also consider the hybrid approach as a future research topic.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

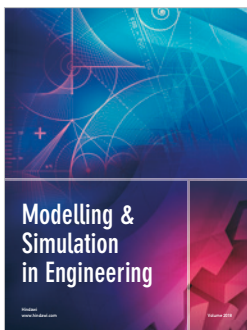The authors declare that there are no conflicts of interest regarding the publication of this paper.
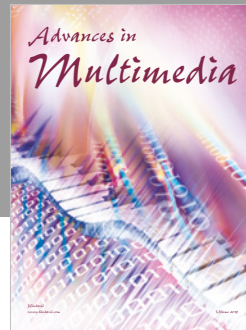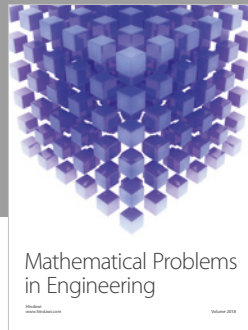
Acknowledgments

References

[1] F. Biocca and M. R. Levy, *Communication in the Age of Virtual Reality*, Routledge, Abingdon, UK, 2013.

[2] D. A. Bowman and R. P. McMahan, "Virtual reality: how much immersion is enough?," *Computer*, vol. 40, no. 7, pp. 36–43, 2007.

[3] J. Gregory, *Virtual Reality*, Cherry Lake Publishing, North Mankato, MN, USA, 2017.

[4] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, "A virtual reality dance training system using motion capture technology," *IEEE Transactions on Learning Technologies*, vol. 4, no. 2, pp. 187–195, 2011.

[5] H. H. Sin and G. C. Lee, "Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia," *American Journal of Physical Medicine and Rehabilitation*, vol. 92, no. 10, pp. 871–880, 2013.

[6] A. Henderson, N. Korner-Bitensky, and M. Levin, "Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery," *Topics in Stroke Rehabilitation*, vol. 14, no. 2, pp. 52–61, 2007.

[7] C. Pietro, S. Silvia, P. Federica, G. Andrea, and R. Giuseppe, "NeuroVirtual 3D: a multiplatform 3D simulation system for application in psychology and neuro-rehabilitation," in *Virtual, Augmented Reality and Serious Games for Healthcare*, pp. 275–286, Springer, Berlin, Germany, 2014.

[8] S. Houlding, *3D Geoscience Modeling: Computer Techniques for Geological Characterization*, Springer Science & Business Media, Berlin, Germany, 2012.

[9] L. Yu and P. Gong, "Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives," *International Journal of Remote Sensing*, vol. 33, no. 12, pp. 3966–3986, 2012.

[10] M. H. Draper, E. S. Viirre, T. A. Furness, and V. J. Gawron, "Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments," *Human Factors*, vol. 43, no. 1, pp. 129–146, 2001.

[11] J. J. W. Lin, H. B. L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Proceedings of the IEEE Virtual Reality*, Orlando, FL, USA, March 2002.

[12] T. R. Coles, D. Meglan, and N. W. John, "The role of haptics in medical training simulators: a survey of the state of the art," *IEEE Transactions on Haptics*, vol. 4, no. 1, pp. 51–66, 2011.

[13] F. Anderson, M. Annett, and W. F. Bischof, "Lean on Wii: physical rehabilitation with virtual reality Wii peripherals," *Studies in Health Technology and Informatics*, vol. 154, pp. 229–234, 2010.

[14] T. P. Pham and Y.-L. Theng, "Game controllers for older adults: experimental study on gameplay experiences and preferences," in *Proceedings of the International Conference on the Foundations of Digital Games*, Raleigh, NC, USA, May–June 2012.

[15] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60–71, 2011.

[16] B. Williams, S. Bailey, G. Narasimham, M. Li, and B. Bodenheimer, "Evaluation of walking in place on a Wii balance board to explore a virtual environment," *ACM Transactions on Applied Perception*, vol. 8, no. 3, pp. 1–14, 2011.

[17] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Transactions on Systems, Man, and Cybernetics: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[18] Oculus VR, LLC, *Oculus Touch*, Oculus VR, Irvine, CA, USA, 2016, https://www.oculus.com/rift/.

[19] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the American Federation of Information Processing Societies Conference (AFIPS 1968)*, vol. 33, p. 1, San Francisco, CA, USA, December 1968.

[20] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, "Evaluating display fidelity and interaction fidelity in a virtual reality game," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 626–633, 2012.

[21] K. Kihara, Y. Fujii, H. Masuda et al., "New three-dimensional head-mounted display system, TMDU-S-3D system, for minimally invasive surgery application: procedures for gasless single-port radical nephrectomy," *International Journal of Urology*, vol. 19, no. 9, pp. 886–889, 2012.

[22] H. Li, L. Trutoiu, K. Olszewski et al., "Facial performance sensing head-mounted display," in *Proceedings of the 42nd ACM SIGGRAPH Conference and Exhibition ACM Transactions on Graphics*, Kobe, Japan, August 2015.

[23] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Proceedings of the 8th International Conference on Information, Communications and Signal Processing (ICICS 2011)*, Singapore, December 2011.

[24] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, 2007.

[25] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.

[26] K. K. Biswas and S. K. Basu, "Gesture recognition using Microsoft Kinect®," in *Proceedings of the 5th International Conference on Automation, Robotics and Applications (ICARA 2011)*, Wellington, New Zealand, December 2011.

[27] D. Meldrum, A. Glennon, S. Herdman, D. Murray, and R. McConn-Walsh, "Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii® Fit Plus," *Disability and Rehabilitation: Assistive Technology*, vol. 7, no. 3, pp. 205–210, 2012.

[28] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2016.

[29] M. N. Kamel Boulos, B. J. Blanchard, C. Walker, J. Montero, A. Tripathy, and R. Gutierrez-Osuna, "Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation," *International Journal of Health Geographics*, vol. 10, p. 45, 2011.

[30] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and Kinect devices," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP 2014)*, Paris, France, October 2014.

[31] Q. Chen, N. D. Georganas, and E. M. Petriu, "Real-time vision-based hand gesture recognition using Haar-like features," in *Proceedings of the Instrumentation and Measurement Technology Conference IMTC IEEE*, Warsaw, Poland, May 2007.

[32] C. Keskin, F. Kıraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Consumer Depth Cameras for Computer Vision*, pp. 119–137, Springer, London, UK, 2013.

[33] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, HI, USA, January 2011.

[34] A. Shahroudy, J. Liu, T. T. Ng, and G. Wang, "NTU RGB+ D: a large scale dataset for 3D human activity analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Caesars Palace, NV, USA, June–July 2016.

[35] B.-S. Sohn, "Design and comparison of immersive gesture interfaces for HMD based virtual world navigation," *IEICE Transactions on Information and Systems*, vol. E99-D, no. 7, pp. 1957–1960, 2016.

[36] Y. Lee, W. Choi, and B.-S. Sohn, "Immersive gesture interfaces for 3D map navigation in HMD-based virtual environments," in *Proceedings of the 32nd International Conference on Information Networking (ICOIN)*, Chiang Mai, Thailand, 2018.

[37] M. C. Corballis, "From mouth to hand: gesture, speech, and the evolution of right-handedness," *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 199–208, 2003.

[38] J. R. Skoyles, "Gesture, language origins, and right handedness," *Psycoloquy*, vol. 11, p. 24, 2000.

[39] I. Yavrucuk, E. Kubali, and O. Tarimci, "A low cost flight simulator using virtual reality tools," *IEEE Aerospace and Electronic Systems Magazine*, vol. 26, no. 4, pp. 10–14, 2011.

[40] M. Rheiner, "Birdly an attempt to fly," in *Proceedings of the ACM SIGGRAPH 2014 Emerging Technologies*, Shenzhen, China, December 2014.

[41] J. Shotton, T. Sharp, A. Kipman et al., "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[42] J. L. Raheja, A. Chaudhary, and K. Singal, "Tracking of fingertips and centers of palm using Kinect," in *Proceedings of the Third International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM)*, Langkawi, Malaysia, September 2011.

[43] S. Wang, Z. Mao, C. Zeng, H. Gong, S. Li, and B. Chen, "A new method of virtual reality based on Unity3D," in *Proceedings of the 18th International Conference on Geoinformatics, 2010*, Beijing, China, June 2010.

[44] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman, "Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations," in *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, Espoo, Finland, September 2006.