Multimed Tools Appl (2016) 75:12331–12350
DOI 10.1007/s11042-016-3520-1

Application of precise indoor position tracking to immersive virtual reality with translational movement support

Jongkyu Shin[1] · Gwangseok An[1] · Joon-Sang Park[2] · Seung Jun Baek[3] · Kyogu Lee[1]

Received: 15 January 2015 / Revised: 31 March 2016 / Accepted: 5 April 2016 /
Published online: 12 April 2016
© Springer Science+Business Media New York 2016

Abstract In this study, we propose an application for immersive virtual reality experiences, which integrates three-dimensional (3D) head-mounted displays (HMDs) with a precise indoor position tracking algorithm based on ultrasound. Our method provides a natural virtual reality experience with interaction by precisely matching the physical movements in the real world with those in the virtual environment, unlike other methods that require external input devices to move around in the virtual environment. Users can move within the assigned indoor space while carrying a wireless client device with the HMD, without the risk of colliding with obstacles or structures. The system is designed to provide the accurate 3D X, Y, and Z coordinate values of translational movements in real-time as well as the pitch, roll, and yaw values of rotational movements supported by the HMD, resulting in the six degrees of freedom

✉ Kyogu Lee
kglee@snu.ac.kr

Jongkyu Shin
jqshin@snu.ac.kr

Gwangseok An
angoang@snu.ac.kr

Joon-Sang Park
jsp@hongik.ac.kr

Seung Jun Baek
sjbaek@korea.ac.kr

[1] Graduate School of Convergence Science and Technology, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 151-742 Seoul, Republic of Korea

[2] Department of Computer Engineering, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul 121-791, Republic of Korea

[3] Department of Computer and Communications Engineering, College of Information and Communications, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-701, Republic of Korea

required by immersive virtual reality. In addition, the system utilizes ultrasonic transducers in a grid format, which makes it simple to expand the position tracking coverage area, and supports simultaneous tracking of multiple users. Through experiments and a user study we show that the system obtains the accurate position of the moving objects and delivers a highly immersive virtual reality experience.

Keywords Mixed reality · Virtual reality · Immersive media experience · Real-time position tracking · Ultrasound

1 Introduction

The proliferation of three-dimensional (3D) head-mounted displays (HMDs) is reshaping the user experience in immersive virtual reality (VR) methods. These systems are used in a variety of different areas, e.g., entertainment and education applications, simulations, and training applications. Low-cost HMDs such as the Oculus Rift[1] or Sony HMZ series[2] are commonly used to provide virtual environments. Their simple setup procedures and the handy size of these devices make them easy to use, but they can also deliver immersive virtual vision to users anytime and anywhere. However, these applications require movements to change the position of the user in the virtual environment (e.g., first or third person shooting game applications), and thus traditional external input devices such as keyboards, joysticks, or game controller pads are still required to move around in the virtual environment. The use of external input controllers causes several problems. First, it reduces the level of immersion because the movement control action has to be performed outside the VR from a different perspective. In addition, most control devices are optimized for two-dimensional (2D) environments, which means that they are only capable of controlling two different axes, thereby making these traditional devices non-intuitive during 3D environmental control [18]. Second, they often cause negative effects in terms of simulator/virtual reality sickness, which is also known as cybersickness [16], where the common symptoms include nausea, oculomotor discomfort, and disorientation [11]. These symptoms are usually attributable to the fixed user position whereas the visual field is moving/changing inside the VR. These issues lead to a sensory gap with respect to the visual senses compared with the other senses, particularly the motor sense.

To address the limitations mentioned above, we propose a novel approach, which is capable of precisely tracking the translational movements of a user wearing a 3D HMD in real-time, while providing the 3D visual field via the HMD for the corresponding location. The system is designed to be implemented in indoor environments, with a high level of precision in position tracking. The system can also be installed outdoors with mounting structures, where it obtains more precise position estimates compared with global positioning system (GPS). We use Oculus Rift DK2 HMD as the 3D vision interface and we apply a position-tracking algorithm based on ultrasound, which delivers higher accuracy, stability and efficient usage of computing power compared with other techniques.

The aim of the proposed system is to provide an immersive VR experience to users, where the physical translation movements of users actually match their movements in the virtual world shown through the HMD device. In the given 3D space, the Cartesian coordinate system

[1] Oculus Rift. http://www.oculusvr.com

[2] Sony HMZ Series. http://www.sony.co.jp/Fun/design/activity/product/hmz_t1/01.html

values of X (forward and backward), Y (left and right), and Z (up and down) are measured using ultrasound, while the pitch, roll, and yaw values obtained from the user's head rotation movements are collected by the internal attitude and heading reference system (AHRS) sensor on the HMD. After combining the data from two different channels and sending the calculated values to the 3D engine, we can track most of the user's movements, including walking, running, jumping, ducking, and head leaning motions. The client device with the HMD communicates wirelessly with the server. Thus, the system is mobile and it does not require any complex setup procedure or connections. In addition, by differentiating the identity of the transmitter device, the system can track multiple users simultaneously in real-time.

Using this method, we provide an immersive virtual/mixed reality simulation platform that supports natural human movements because the system has six degrees of freedom in terms of motion. This avoids the requirement for an external input device when moving inside the VR and it reduces the risk of cybersickness by matching the feedback from the visual sense with other sensory feedback.

The remainder of this paper is organized as follows. In the next section, we describe previous research in this area and methods related to our study. In the following section, we describe our system in detail, including the position tracking method and its integration with the HMD. We then present the experimental results that we obtained in various conditions and the results of a user study, before giving suggestions for future research.

2 Related work

In this section, we describe previous works on user interfaces and controllers in an attempt to provide users with more natural and immersive interactions in virtual reality environments. In addition, we focus on previous studies on indoor position tracking to follow user's locations in real time.

2.1 Input interfaces in virtual environments

Many studies have attempted to develop methods that avoid the use of external input devices for movement control in virtual environments. These methods mainly aim to maximize immersion in the virtual experience by reducing the difference between the real world and the virtual world, as well as improving user interactions in these scenarios.

Mokka et al. used a stationary bicycle structure to create an exercise/fitness game known as the exergame [21]. In this "Virku" (virtual fitness center) interface, the virtual image projected in front of the system moved forward as the user rode upon a bicycle structure. However, this system did not provide any stereoscopic 3D perspectives and it was not possible to change the heading direction from going forward; thus, we assume that the system would not provide a realistic immersive experience. The more immersive PaperDude system proposed by Bolton et al. provides stereoscopic vision using a HMD [3]. This system resembles the Virku interface in terms of user interaction because it also employs a bicycle structure. However, in terms of the user's position, the system only provides head orientation and the exercises basically involve cycling on a fixed spot where the absolute position of the user does not change. This is not a viable solution for generating greater immersion in virtual environments.

Some previous methods employed optical-based head-tracking technologies for tracking the user's head position. Chow designed an HMD-tracking approach that used visual markers, which were detected by a camera [4]. The camera was mounted on the ceiling and it detected the movements of the user by tracking visual markers attached to the user. Thus, the user's location could be tracked in a limited space, which must be located within the range of the camera. The main disadvantage of this approach is that the line of sight must be maintained between the markers and the camera. If the marker becomes invisible to the camera, the position cannot be tracked. A similar method is used in the most recent release of the Oculus Rift Development Kit 2 (DK2), which features a head tracking capability that uses infrared (IR) LEDs and an external IR camera. In this system, an external camera detects the depths/distances of multiple signals obtained from the IR LEDs, which are installed on the front of the HMD, thereby tracking their movements. However, this approach still has the same limitations as Chow's design because tracking is only possible within the limited range of the camera, and the path must be clear from the camera to the targets (IR LEDs). These optical tracking methods have additional disadvantages because the systems tend to be highly complex and they require demanding computation as they are based on image processing. They are expensive because these systems require many cameras to cover a defined space and they have a limited range of detection.

Hodgson et al. developed an HMD-based wearable VR system with a location tracking capacity [15]. This system employs a similar concept to our system because it enables immersive VR simulation with human movement support. However, this system utilizes GPS, which means that it can only be operated in outdoor environments where the sky is visible but also has low accuracy, i.e., a resolution of 1 m, error of ±3 m and a refresh rate of 1Hz. This tolerance range is far greater than that required for a truly immersive experience.

In our previous work, we introduced a platform for experiencing VR with translational movement support, where an HMD was integrated with an indoor position tracking algorithm [24]. This system was designed to provide an immersive media experience to users, but it has several limitations. First, it only operated within a small limited space because only four signal receivers were employed. Second, it could be affected by radio interference due to the utilization of acoustic signal transmissions at 5.8GHz RF, and also required an external function generator device. Furthermore, the resolution of the VR image was not suitable for obtaining a high level of immersion and the system could only track one user at a time. In the present study, we propose an improved system that covers a larger space by using modular signal receiving grids with modulated synchronization signal transmission to provide greater robustness as well as better VR resolution, with the ability to track multiple users.

2.2 Indoor position tracking

Because GPS systems do not work in indoor environments, several studies have focused on indoor position tracking using various techniques, including. Wi-Fi [7], RF signals [1, 25], Bluetooth [5], IR vision [23], magnetic [12], optical tracking [28, 29] and acoustic-based methods, mainly using ultrasound [9, 13, 22, 27]. Each method has its own advantages and disadvantages in terms of its accuracy, coverage area, cost, infrastructure, availability, and privacy [2, 14, 20].

The proposed method has several advantages compared with other techniques by using the ultrasound technology. First, it has the lowest error rate and highest resolution/accuracy because the speed of ultrasound is known to be 343 m/s at room temperature, which is far slower than other radio signals used commonly for indoor position tracking. This allows us to measure the time of flight (TOF), even at very low sampling rates. Consequently, our approach is appropriate for precise indoor position tracking, which is the primary requirement of our system when matching physical movements with those in a virtual environment. Second, our method is easy to implement because it requires the installation of fewer devices to obtain the necessary infrastructure. Finally, it facilitates installation on other devices to allow tracking by "piggy-backing" the signal emitter, due to the small size of the ultrasonic transducers.

Our system differs from other ultrasound-based methods such as Cricket [22] and Dolphin [9] in terms of its aim, the transducer type, receiver/transmitter location and the synchronization process. Our system focuses on maximizing the accuracy when estimating the target location within an assigned grid space by using the multiple omnidirectional ultrasound transducers to cover a larger space where each sensor transmits/receives the encoded digital synchronization signal wirelessly. The InterSense 900,[3] which is a widely used commercial ultrasound-based position tracking system in VR environments, requires large ultrasonic sensor array strips to be installed (size varying from 60 to 180 cm), whereas our method utilizes small receivers (4 cm (W) × 4 cm (L) × 1.3 cm(H)) at each corner of a grid. For example, to track the movement in a 10 m × 10 m space, the InterSense 900 requires approximately 84 array strips, while our method needs a sparse installation of just 16 receivers.

3 System description

In this section, we explain our position tracking system in detail. First, we briefly describe the overview of the proposed method in the following section. In Section 3.2, we then explain the advantages of the proposed system over our previous work in two aspects: 1) expansion of coverage area and 2) multi-user tracking. Finally, we explain the implementation details for integrating the position tracking algorithm with HMDs in Section 3.3.

3.1 System overview

Figure 1 illustrates the overall architecture and components of the system. Our system is designed to be mobile, and thus it requires no external power or signal cables. The system comprises the client part and the server part. Each client device (shown on the left of Fig. 2) is paired with the HMD, which is carried by the user. The role of this device is to generate ultrasonic burst signals, to receive the estimated location values from the server, and to pass them to the 3D engine to provide the corresponding virtual visual input through the HMD. It is connected to the Oculus Rift DK2 HMD and has a UHF frequency-shift keying (FSK) signal transmitter modem to receive the synchronization signal, a Lenovo S440 laptop computer to run the 3D engine made based on Unity3D,[4] and a ZigBee[5] USB module to receive the estimated coordinate values from the server on the laptop. Unlike previously developed system which transmit raw acoustic signals via 5.8GHz RF [24], the current system encodes signals to ensure more robustness. An Altera FPGA board is also present inside the device to decode the synchronization signal, to check the client device ID, and to generate the pulse signal. An omnidirectional ultrasound transducer is mounted on the HMD, which generates a burst tone at 20Hz with a carrier frequency of 40 kHz. The maximum running time of the client device is

[3] InterSense 900. http://www.intersense.com/pages/20/14
[4] Unity3D. http://www.unity3d.com
[5] ZigBee Alliance. http://www.zigbee.org

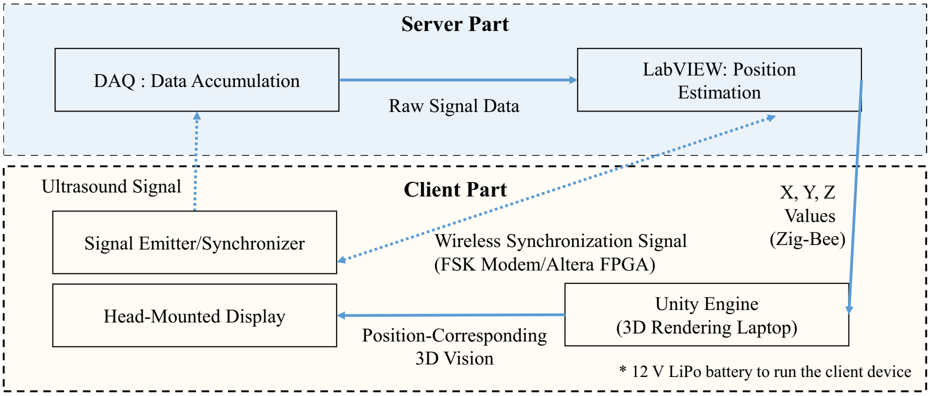

Fig. 1 System architecture and components

approximately 20 h with a 12 V 1200mAh lithium polymer battery. The size of the client device excluding the laptop and the HMD is 13 cm (W) × 13 cm (L) × 6 cm (H) (left of Fig. 2) and each costs about $100.

The server part comprises two sections; 1) the omnidirectional ultrasonic signal receivers mounted on the ceiling and, 2) a desktop PC for calculating the position by triangulating the signal. The size of each beacon is measured at 4 cm (W) × 4 cm (L) × 1.3 cm (H). The acoustic signals from several beacons are accumulated using data acquisition (DAQ) equipment and they are then processed by LabVIEW.[6] The calculated position values are then filtered to maintain constant delta values. This feature functions as a precaution to avoid sudden fluctuations in the values, which would translate into sudden teleport movements in the virtual environment.

Following this process, the refined translational movement values, i.e., X, Y, and Z, are transferred wirelessly to the client laptop using the ZigBee network. The user can then experience a "physically matched" virtual world while wearing the HMD with the client device. The right of Fig. 2 shows a user equipped with our system.

3.2 Position tracking

Omnidirectional ultrasonic transducers are used as signal receiving beacons and the ultrasound signal is collected by a National Instruments PXIe-6356 DAQ at a sample rate of 400 kHz. As mentioned above, two different types of wireless transmission channels are used between the client and the server: 1) 447.900 MHz UHF range to generate the signal for the ultrasonic transducer, which is installed on top of the HMD; and 2) a ZigBee network to transmit three types of location coordinate values from the server to the laptop.

By synchronizing the time of signal burst between the server and the client, the server organizes clients to generate ultrasonic burst signals at defined times. The control of the synchronization signal is a crucial feature of positioning systems that use ultrasound based on measurements of the TOF of emitted signals [14]. However, the use of complex digital modulation techniques for transmitting them would inevitably lead to delays. Thus, sending acoustic signals with a function generator device [24] or using other complex techniques, we use the Manchester coding scheme for signals with FSK modems that operate at 447.900 MHz

[6] National Instruments LabVIEW. http://www.ni.com/labview/

Fig. 2 The client device (*left*) and a user equipped with the system (*right*)

to transmit/receive synchronization signals between the server and the client. The server creates inherent IDs for each client in an 8-bit data format and transmits them to all of the client devices. The client only generates burst signals when it receives the corresponding ID. This also allows the simultaneous tracking of more than one client, as described in the following subsection. After each signal burst, the system measures the TOF of the ultrasound between the transmitter and receivers. The overall process is shown in Fig. 3.

The positioning algorithm is similar to Cricket [22] or the commercial system IS-900, but the proposed approach differs greatly because our system measures the TOF of each ultrasound beacon (minimum three) and transmits the encoded synchronization signal wirelessly. Second, the signal transmitter/receiver pair works in a reversed configuration, allowing the target (VR interface) to move freely in a space. In addition, our implementation focuses on high accuracy by utilizing limited ultrasound beacons within a pre-assigned area. Position tracking will only work within the range of these assigned spaces because our system is focused on maximizing the accuracy in this range rather than estimating random locations. As a result, the system has a resolution of less than 1 mm and a maximum error of 20 mm in optimal environments. The accuracy of the system in practical environments is described in the following experimental section. The calibration process is performed in the first initialization stage to minimize location errors which can be caused by the temperature variance (more than

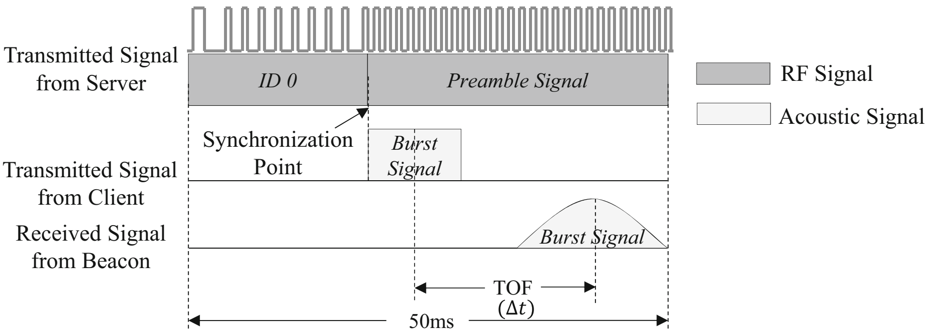

Fig. 3 Estimation of the TOF in a single user system

5° Celsius), humidity, and acoustic characteristics (extremely reflective materials and/or structures may cause the multi-paths of the ultrasound signal) of the environment.

We now describe in more detail the two most critical advances from our previous work – i.e., coverage area expansion and multi-user tracking – in the following sections.

3.2.1 Coverage area expansion

The structure of the receivers on the ceiling is designed to be modular, where four transducers create a rectangular shape per set to form a grid. Each rectangle measures 2.5 m × 3 m, which is an optimal size when considering the transmission power of ultrasound signals. The shape of the grid could be installed in a different form other than a rectangle, as long as at least three receivers are able to gather the signal from the transmitter.

In rectangular shape of the grid, the size of the area covered can be doubled by adding two extra receivers to one side. In this study, we tested up to eight channels by stacking three rectangle shapes in a row, which covered a length of 9 m, a width of 2.5 m, and a height of 2.6 m. In a real-world installation (discussed in more detail in Section 5), we successfully implemented the system to cover a 32 m (L) × 2.5 m (W) × 3 m (H) corridor space with 16 channels of beacons.

3.2.2 Multi-user tracking

In the current system, the refresh rate for signal acquisition is set to 20Hz. At this rate, the encoded synchronization signal (using the Manchester encoding scheme) is sent wirelessly to the client system using the previously mentioned FSK modem and the FPGA decodes the signal before deciding the time to generate the ultrasonic signal via the transducer, which is installed on top of the HMD. When tracking multiple users, this process differs slightly because the system only utilizes one channel in the UHF range for the synchronization signal. To differentiate the client device ID numbers while sharing the same channel, each ID data is transferred in different time slots of the 8-bit format signal data as shown in Fig. 4. This method is similar to time division multiple access [6]. Thus, because the assigned time slots are limited and multiple clients have to share the same frequency channel, the location estimation refresh rate inevitably

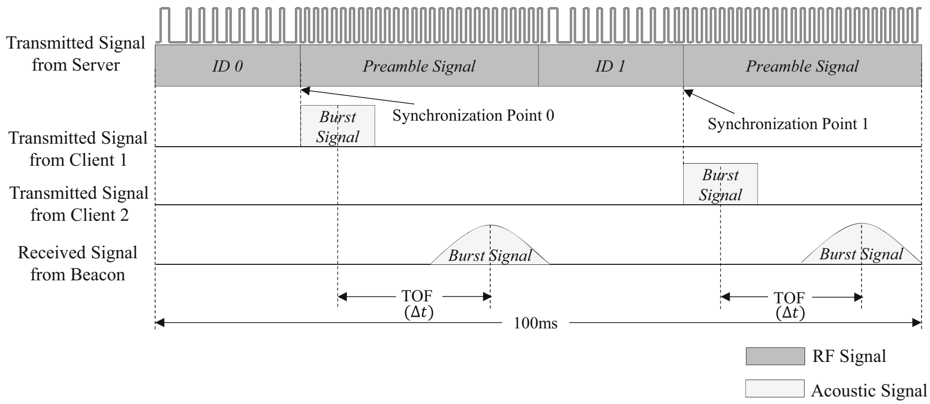

Fig. 4 Estimation of TOF in a multi-user system

declines as the number of clients increases. We experimented up to three clients simultaneously, each client having a refresh rate of 6Hz (bound by 20/number of clients).

3.3 Integration with HMDs

The Unity3D engine is used to render the 3D environments. A control script for Oculus Rift player control was also developed for receiving the user's coordinate values and placing the position of the user inside the virtual environment. To obtain six degrees of freedom in the VR, six different types of data are utilized in the system to represent the X, Y, and Z coordinates, and the pitch, roll, and yaw values. First, the translational movements (coordinate values) are measured with LabVIEW on the server, which are then transferred to the 3D engine of the laptop computer of the client using the ZigBee network via internal serial port communication. The ZigBee modules are utilized in the unicast mode instead of the broadcast mode, which maximizes the speed of communication. The head orientation data values are imported directly from the HMD to the 3D engine. The rotational movements (pitch, roll, and yaw values) are collected from the internal AHRS sensor of Oculus Rift DK2 instead of an additional external sensor, because the internal DK2 sensor delivers more reliable results when estimating attitude information compared with the DK1, which was utilized in previous system [24].

4 Experiment

We performed several experiments in various conditions to examine the feasibility of indoor position tracking and the level of immersion with the proposed system. We set up eight ultrasonic receivers in an open office room to create three grids in a row, which covered an area of 9 m (L) × 2.5 m (W) × 2.6 m (H). The X, Y, and Z coordinate values of the virtual environment were designed to match the real environment. Three types of experiment were conducted.

1) Measurement of fixed position points
2) Real-time position tracking estimation
3) User movement trajectory observations/user feedback

Each experiment and the results are explained in more detail in the following sections.

4.1 Position stability measurement

In the first experiment, we aimed to measure the steadiness of the system in terms of estimating immovable, fixed positions. Highly fluctuating values when estimating static situations can be represented as unexpected movements such as teleportation in a virtual environment; thus, it is essential to obtain steady estimates in these situations. The experiment measured the standard deviations of all the X (forward/backward), Y (left/right), and Z (height) coordinates. We developed a filter to consider constant delta values as the estimated values for these scenarios, as mentioned above, but it was removed in this experiment and the raw data values were used instead.

In the experimental space of 9 m × 2.5 m × 2.6 m, we selected 10 random positions and we differentiated all three types of coordinate values. After detaching the signal emitting transducer from the HMD and placing it in the assigned spot, the system estimated 10 fixed positions with 20 samples for each test set.

Figure 5 shows the estimated standard deviations, i.e., X axis = 0.8 mm, Y axis = 0.79 mm, and Z axis = 2.94 mm.

In contrast to X and Y, the standard deviation was greater for height or Z, where the calculated TOF variances of the height values were noticeably lower compared with the other values because the ultrasound receivers were mounted on the ceiling. No particular sweet spot was found in the assigned area.

4.2 Position tracking measurement

The second experiment measured the estimated position tracking values in real-time using model trains and rails. This experiment was also performed without the HMD attached because the aim of this test was to determine the error rate of the movement position estimates. The rail was set up in the center of the aforementioned area of 9 m × 2.5 m × 2.6 m.

The total distance of the rail structure was 612 cm and it measured 168.6 cm × 137.4 cm. The ultrasonic transducers on the ceiling were at a height of 260 cm, where the floor was at 0 cm. The model train moved at a constant speed and the ultrasound signal was emitted 20 times per second. To measure the height (Z) of the target as well as its position in an X-Y plane, the rail was placed on a table at a height of 73 cm with one side tilted at 22.3° (Figs. 6 and 7).

Two conditions were specified with different train velocities and one round of travel was recorded for each condition. Figures 6 and 7 show the position tracking estimation results obtained in the two different conditions. The red lines in the figures represent the actual rail whereas the overlapping blue lines denote the sum of the measured positions.

In the first condition, the average velocity was 6.3 cm/s, the maximum estimation error was 1.27 cm, and the root mean squared (RMS) error was 0.181 cm. In the second condition, the average velocity was 17.8 cm/s, the maximum error was 1.4 cm, and the RMS error was 0.238 cm. For both conditions, the average data acquisition delay on position tracking did not exceed 10 ms.

Because we tested the train at different speeds, the results tend to have larger errors with the increasing velocity. However, in normal situations when the transmitter was moving within the range of the signal receivers, the maximum distance estimation error did not exceed 20 mm regardless of height.

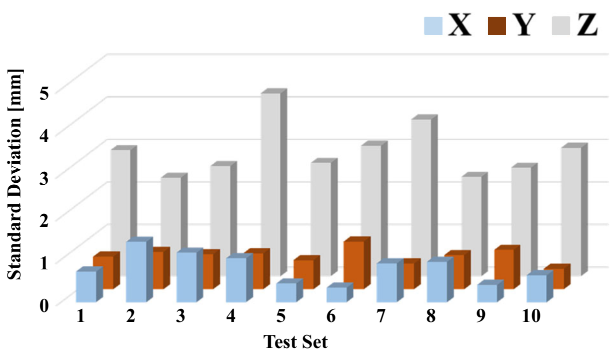

Fig. 5 Standard deviations of the fixed location measurements

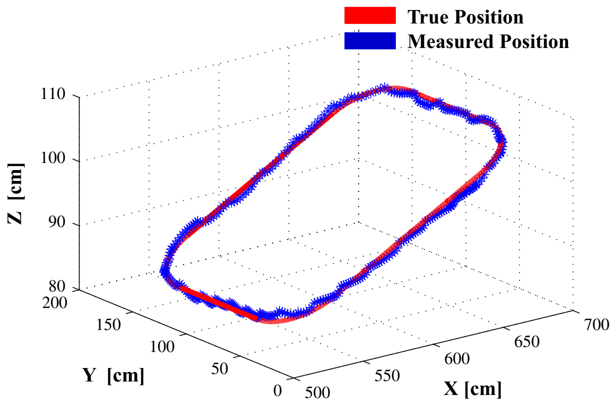

Fig. 6 Position tracking estimation results when the train velocity was 6.3 cm/s

4.3 User study

As the final experiment, we performed a user study in order to evaluate the qualitative performance of the proposed system, and compared with a keyboard interface in completing a simple navigation task.

4.3.1 User movement trajectory tracking

To validate our approach in practice, we performed a user study with 15 participants, i.e., nine men and six women, who were aged between 20 and 34 years. All of the participants were familiar or had an experience with 3D-based VR in computer games. All the participants had (corrected) normal vision and responded that they do not suffer from any health issues. The system concept was briefly explained before experiments. For comparison with a conventional system, we performed another experiment under a different condition where the subjects are directed to perform the same task using a keyboard for translation movement.

In the first condition, the participants tested our system by moving around freely in the assigned space while carrying the client device and wearing the HMD. In the second condition, the participants sat on a chair and used the keyboard to move in the virtual environment. In the second condition, the directions faced in the virtual perspective were controlled by tilting the

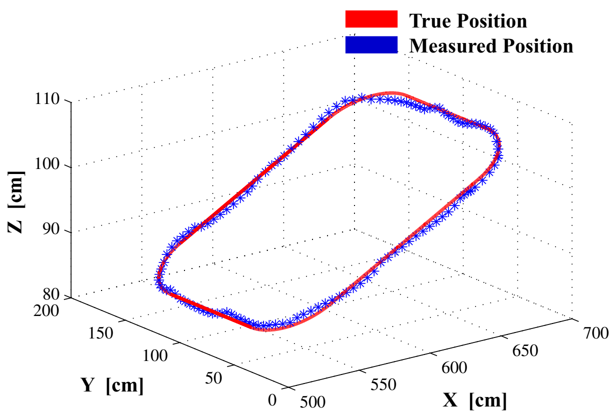

Fig. 7 Position tracking estimation results when the train velocity was 17.8 cm/s

head using the AHRS sensor on the HMD because the keyboard inputs only allowed four movement options (forward, backward, to the left, and to the right). The participants experienced both systems for approximately 10 min each, with 5 min of break in between.

The same virtual image was shown to the participants in both conditions. As shown in Fig. 8, the environment was rendered as a 9 m × 2 m rectangular stone cliff with the two large pillars of fire placed at 3 m and 6 m from the starting point, which represent the obstacles that exist in the actual space (plastic barrels with a diameter of 80 cm). It was designed to allow the subjects to fall off the cliff if they cross the boundaries (the edges of the cliff) to examine the degree of immersion in each condition. Each subject was given 6 to 8 min to explore the VR environment to get used to each condition before the experiment. Finally, we asked the subjects to walk from the start to the end of the cliff without catching fire (i.e., avoiding colliding with the obstacles) nor falling off the cliff. The movement trajectories of the subjects were recorded individually.

In Figs. 9 and 10, the trajectories of each user in two conditions are drawn in different colors. As shown in Fig. 9, all users in the first condition accomplished the task successfully: they avoided neither bumping into the fires nor falling off. In the second condition, all the participants also successfully avoided the obstacles. However, the two participants fell off the cliff (one at 4 m point and the other at the end point of the space, indicated by dotted squares in Fig. 10). These participants claimed that they fell off unintentionally as they felt keyboard control was not intuitive to feel the sense of space. Another interesting observation from the figures is that the trajectories made by the keyboard interface tend to contain many piecewise "straight lines" while those made by the proposed system do not. This in turn is one of the evidences that the proposed method provides a closer-to real VR experience to users.

4.3.2 User feedback

After experiencing each system, a survey including the Simulator Sickness Questionnaire (SSQ) [19] and questions adapted from the part of the MEC Spatial Presence Questionnaire (MEC-SPQ) [26] was conducted to collect user feedback about the possibility of cybersickness and spatial

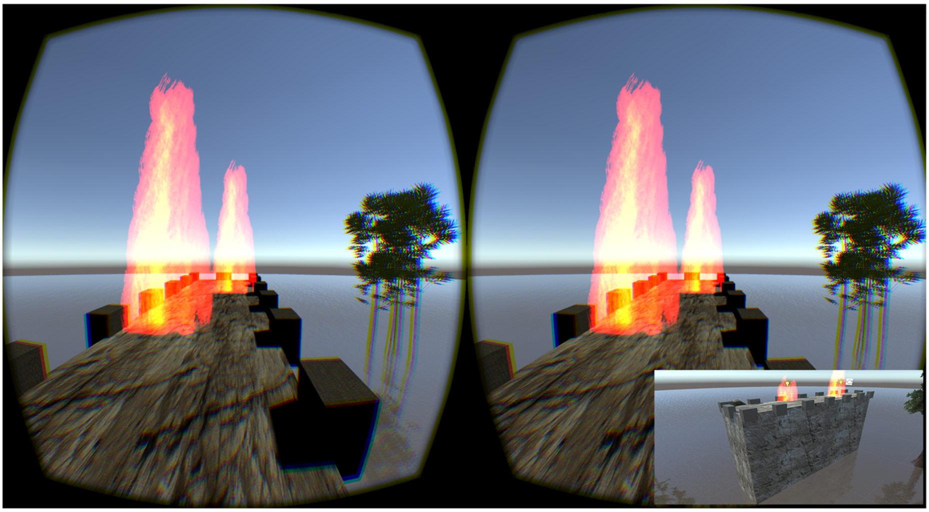

Fig. 8 The virtual image shown through the HMD. The overall view of the environment is shown in lower right

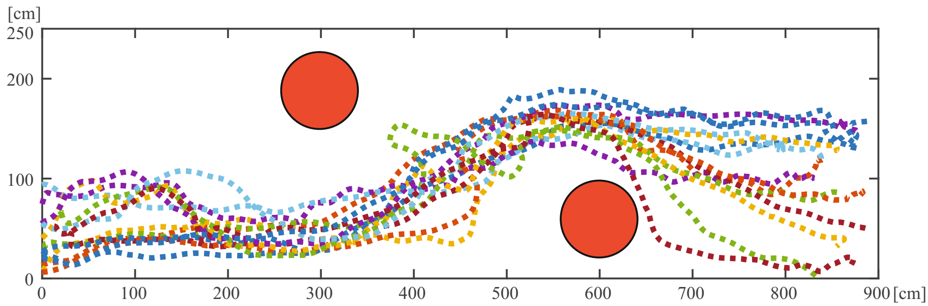

Fig. 9 Recorded user movement trajectories (proposed method). The red circles represent the obstacles (fires) in the VR

presence. Since our sample size was too small ($N = 15$) and the distribution of responses was not normally distributed, we used a Wilcoxon Signed Rank Test to analyze the data.

Developed by Kennedy et al., the SSQ consists of 16 symptoms in three distinct clusters: Nausea, Oculomotor, and Disorientation [19]. Participants rated each symptom on a 4-level Likert scale ("none = 0", "slight = 1", "moderate = 2" and "severe = 3"). The three clusters were then combined to compute the overall SSQ score. Only post-immersion data was collected in this test.

The SSQ results are shown in Fig. 11. It shows that the mean value of the total score is significantly lower for our system (mean = 16.20, SD = 14.32) than for the keyboard interface (mean = 29.92, SD = 17.26), Z = −2.273, $p < 0.05$, r = −0.587. Furthermore, the results are similar on every subscale, yielding significantly lower values for the proposed method: the mean values for each condition were 9.54 (SD = 10.19) and 17.80 (SD = 10.11) (Z = −2.1889, $p < 0.05$, r = −0.565) on Nausea, 17.18 (SD = 14.17) and 30.32 (SD = 20.05) (Z = −2.0785, $p < 0.05$, r = −0.537) on Oculomotor, 14.85 (SD = 21.98) and 29.69 (SD = 21.61) (Z = −2.0664, $p < 0.05$, r = −0.536) on Disorientation.

To compare the spatial presence of each condition, we adapted questions from the Spatial Presence: Self Location (SP:SL) part of MEC-SPQ. We used the four-item scale version of SP:SL questions for the user study since our current system focuses on precise location tracking in a virtual environment and yet could not provide any additional interaction (e.g. gesture interaction). The participants were asked to rate four items (marked as Q1-Q4 in this survey) on a 5-level Likert scale ranging from 1 ('I do not agree at all') to 5 ('I fully agree'). Figure 12 shows the results of the partial MEC-SPQ.

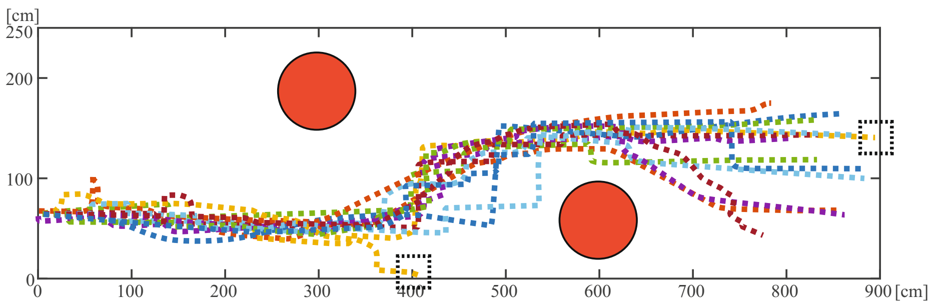

Fig. 10 Recorded user movement trajectories (keyboard interface). The red circles represent the obstacles (fires) in the VR. The dotted squares indicate the two participants who fell off the cliff

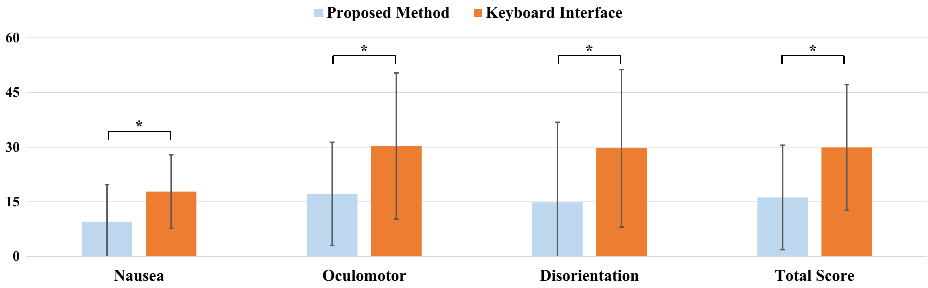

Fig. 11 SSQ score comparison between the two conditions (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

As shown in Fig. 12, the results confirm that our system achieves significantly higher SP:SL scores than the conventional keyboard interface on all four items. In response to the first question "I felt like I was actually there in the environment of the presentation", our system obtained a higher mean value (mean = 3.6, SD = 0.63) than the traditional system (mean = 2.27, SD = 0.59) (Z = 3.465, $p<0.001$, r = 0.895). For the second question "It was as though my true location had shifted into the environment in the presentation," the mean values were 4.13 (SD = 0.74) and 1.93 (SD = 0.79) (Z = 3.360, $p<0.001$, r = 0.868). In response to the third question, "I felt as though I was physically present in the environment of the presentation," the mean values were 3.67 (SD = 0.72) and 2.27 (SD = 0.70) (Z = 3.401, $p<0.001$, r = 0.878). For the final question "It seemed as though I actually took part in the action of the presentation," they were 3.47 (SD = 0.92) and 2.13 (SD = 0.74), respectively (Z = 3.051, $p<0.05$, r = 0.788).

Additional feedback about the system from the participants includes: "The movement support system seemed very intuitive and natural," "The movement support system provided more immersion," "Keyboard control was difficult, as precise movement control was not possible while moving my head. Thus I fell off the cliff by mistake," "It would be better to see myself (referring to body, hands, and feet) in the VR." We accept that awkwardness due to missing body parts in the virtual environment could be a distraction to immersion. A fundamental solution to this problem would be the addition of extra devices to track the user's body parts, such as his/her feet or hands, but this would require more computational power and make the system even more complex. As a workaround, we adjusted the initial user height by making the line of sight 45 cm lower than it is in reality to reduce the "floating in the air" feeling.

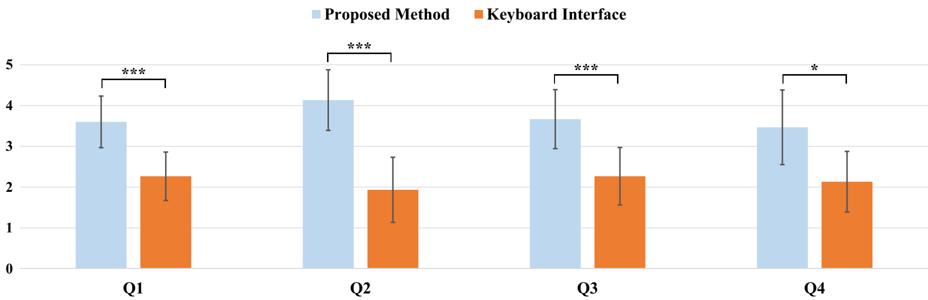

Fig. 12 SP:SL (Partial MEC-SPQ) score comparison between two groups (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

5 Application in interactive media installation

For the real-world application of the proposed system, we implemented it at the exhibition named "Just Like the Road Across the Earth" [17], as an interactive media installation. The exhibition was conceived by a Korean photographer Yeondoo Jung at the contemporary art gallery in Art Tower Mito,[7] Mito, Ibaraki, Japan. Among the several sections of the exhibition, the system was implemented in the "Blind Perspective" artwork section. The installation lasted for 3 months, from November 8, 2014 until February 1, 2015. A total number of approximately 6200 visitors experienced the interactive media installation, and during the period the system operated flawlessly without any major technical issues.

The concept of the artwork was to provide the visitors with the two contradictory perspectives while they walk down a long corridor of 32 m long; one representing the "chaos" (real-world installation) and the other illustrating the "dreamland" (virtual environment rendered through the HMD). The visitors were able to easily switch between the two perspectives while they were viewing the exhibit. Figure 13 shows the two different perspectives used in the installation.

It was critical to precisely track the location of a user's head in real-time in order to seamlessly align the perspectives in the two different environments which allows deeper immersion. Our system covered a corridor area of 32 m (L) × 2.5 m (W) × 3 m (H) with 16 beacons or ultrasound receivers installed on the ceiling in a zigzag shape to maximize the coverage area with a limited number of channels. As we mentioned in Section 3.2.1, the zigzag shape did not affect the system accuracy/stability due to the fact that the transmitter signal from the top of the HMD is always detectable by at least three beacons at any point in the assigned area. Two users were tracked simultaneously in the area using a refresh rate of 10Hz each. We believe that the successful deployment and operation at an actual exhibition, in addition to the promising results from the experiments and a user study from the previous section, confirm that the proposed system can be used in practical applications for mixed or virtual reality that require immersion.

6 Limitations and future work

The current system has some limitations. First, a clear line of sight must be maintained between the ultrasound transmitters and receivers to operate the system. Any objects that obstruct the path between the ceiling and the transmitter on top of the HMD would create acoustic shadows, thereby preventing the system from estimating the locations. Although we utilized omnidirectional units and implemented an algorithm to avoid these types of problems, they were not sufficient in some scenarios (e.g., extreme body leaning motions or the user's hands blocked the path) and they often caused slight blinks or slight teleportation in the virtual image. We consider that these issues can be addressed utilizing multiple transducers and by controlling the emission power.

Second, during the experiments, we also found that fluctuations often occurred when estimating the location in transition sections where grids meet each other and when the room temperature is different from that at the calibration stage. A deeper investigation on this issue led to a finding that temperature variance more than 5° Celsius may affect the system to cause

[7] Art Tower Mito. https://arttowermito.or.jp/gallery_en/gallery02.html?id=413

Fig. 13 The real environment (*left*) and the VR environment (*right*)

a relative error or an "offset" in location estimation (approximately 5 cm at 5° Celsius), which requires modifications of the speed of sound variable in the server to prevent errors in the grid transition sections. It is therefore important to perform an accurate pre-calibration that takes into account several factors, including the temperature, humidity, and acoustic characteristics of the space.

In addition, as mentioned in the user study, some participants noted that they felt awkward in the virtual image because they could not see their whole body as well as their hands inside the HMD. If the body is rendered in VR so the users' feet and hands become traceable, a better immersion and presence result may be obtained. In addition, some participants claimed that it would be more immersive if a sound feature was added. Spatial audio, also known as HRFT binaural 3D audio [8, 10], could be a solution to this problem. This 3D audio function is widely known and it can be implemented directly in 3D rendering engines. Thus, we suggest that including a few additional devices in our system would allow this feature to give users a better VR experience.

Furthermore, the current system has a spatial ratio of 1:1 between the physical environment and the virtual space. This was set to prevent the risk of cybersickness, by minimizing the sensory gap between the visual and motor senses. But it also limits the virtual perspective to the range of the actual space.

7 Conclusion

In this study, we successfully demonstrated the feasibility of an immersive VR application using the 3D HMD with an indoor position tracking ability, which provides six degrees of freedom in VR. Our method avoids the use of traditional input devices such as joysticks or keyboards to move freely in the virtual environment. The major advantage of the proposed system is that it allows exact matching of the physical movements in the real world with those in the virtual environment, thereby providing a truly immersive and engaging VR experience than the existing methods.

We developed and implemented the prototype by installing the modular signal receiving grids to cover a larger space. We used a modulated synchronization signal to improve the system's reliability, and designed the system to provide images with a higher resolution and the capability to track multiple users. Through various experiments we demonstrated that the system achieves the accuracy and robustness of real-time position tracking that are sufficient for immersive VR experience. Our user study also verified that the proposed system achieved qualitatively better performance than the conventional keyboard interface in a task with translational movements to allow natural interactions. The users also preferred the proposed system to the existing method in terms of spatial presence and visual immersion. Real-world implementation of the proposed system in an interactive media installation further confirmed the potential of the system in various applications with immersive virtual or mixed reality.

There are several directions for future research. First, we plan to improve the system so that it can track multiple users while preserving the refreshing rate which is currently limited by the speed of sound and the number of users. This will open far more possibilities for deeper immersion with user-to-user interactions. In addition, we aim to develop a device for user body tracking that will allow more natural interactions, as suggested in our user study. Another direction for future work is to develop a very small-sized transmitter so that it can be placed on any moving object like a baseball or bat, for example. This will enable user-to-object interactions as we experience in our everyday lives.

Acknowledgments This research was partly supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2015-H8501-15-1016) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

1. Bahl P, Padmanabhan V (2000) RADAR: an in-building RF-based user location and tracking system. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE 00
2. Bhatnagar DK (1993) Position trackers for Head Mounted Display systems: a survey. University of North Carolina, Chapel Hill TR93-010 1–22
3. Bolton J, Lambert M, Lirette D, Unsworth B (2014) PaperDude: a virtual reality cycling exergame. CHI'14 Extended Abstracts on Human Factors in Computing Systems 475–478
4. Chow Y (2009) A cost-effective 3D interaction approach for immersive virtual reality. Int J Recent Trends Eng Technol 1:527–529
5. Cruz O, Ramos E, Ramirez M (2011) 3D indoor location and navigation system based on Bluetooth. In: CONIELECOMP 2011 - 21st International Conference on Electronics Communications and Computers, Proceedings, pp 271–277
6. Falconer DD, Adachi F, Gudmundson B (1995) Time division multiple access methods for wireless personal communications. IEEE Commun Mag IEEE Commun Mag 33:50–57
7. Ferris B, Fox D, Lawrence N (2007) WiFi-SLAM using gaussian process latent variable models. IJCAI: 2480–2485
8. Frauenberger C, Noisternig M (2003) 3D audio interfaces for the blind. Proceedings of the 2003 International Conference on Auditory Display 1–4
9. Fukuju Y, Minami M, Morikawa H, Aoyama T (2003) DOLPHIN: an autonomous indoor positioning system in ubiquitous computing environment. Proceedings IEEE Workshop on Software Technologies for Future Embedded Systems WSTFES 2003. doi:10.1109/WSTFES.2003.1201360
10. Gardner W (1999) 3D audio and acoustic environment modeling. Wave Arts, Inc 1–9
11. Hale KS, Stanney KM (2014) Handbook of virtual environments: design, implementation, and applications. CRC Press
12. Haverinen J, Kemppainen A (2009) Global indoor self-localization based on the ambient magnetic field. Robot Auton Syst 57:1028–1035. doi:10.1016/j.robot.2009.07.018
13. Hazas M, Ward A (2002) A novel broadband ultrasonic location system. In: UbiComp 2002: Ubiquitous Computing, Springer, pp 264–280
14. Hightower J, Hightower J, Borriello G, Borriello G (2001) Location systems for ubiquitous computing. Computer 34:57–66. doi:10.1109/2.940014
15. Hodgson E, Bachmann ER, Vincent D, Zmuda M, Waller D, Calusdian J (2014) WeaVR: a self-contained and wearable immersive virtual environment simulation system. Behav Res Methods 296–307. doi:10.3758/s13428-014-0463-1
16. Jr JL (2000) A discussion of cybersickness in virtual environments. ACM SIGCHI Bulletin 32

17. Jung Y (2014) Just Like the Road Across the Earth [Exhibition Catalogue]. 04 Nov 2014–01 Feb 2015, Contemporary Art Gallery, Art Tower Mito, Mito, Ibaraki, Japan
18. Katzourin M, Ignatoff D, Quirk L, LaViola JJ, Jenkins OC (2006) Swordplay: innovating game development through VR. IEEE Comput Graph Appl 26:15–19. doi:10.1109/MCG.2006.137
19. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3:203–220. doi:10.1207/s15327108ijap0303_3
20. Mautz R (2012) Indoor positioning technologies. Thesis submitted to ETH Zurich. doi:10.3929/ethz-a-007313554
21. Mokka S, Väätänen A (2003) Fitness computer game with a bodily user interface. Proc Second Int Conf Entertain Comput 1:3–5
22. Priyantha NB, Chakraborty A, Balakrishnan H (2000) The Cricket location-support system. Proceedings of the 6th annual international conference on Mobile computing and networking - MobiCom'00 32–43. doi:10.1145/345910.345917
23. Satyavolu S, Bruder G, Willemsen P, Steinicke F (2012) Analysis of IR-based virtual reality tracking using multiple Kinects. In: Proceedings - IEEE Virtual Reality, pp 149–150
24. Shin J, An G, Lee K (2014) Integration of a precise indoor position tracking algorithm with an HMD-Based Virtual Reality System. Proceedings of the 2nd ACM International Workshop on Immersive Media Experiences 23–26. doi:10.1145/2660579.2660582
25. Uddin M, Nadeem T (2013) SpyLoc: a light weight localization system for smartphones. Proceedings of the 19th annual international conference on Mobile computing & networking, pp 223–225
26. Vorderer P, Wirth W, Gouveia FR, Biocca F, Saari T, Jäncke L, Böcking S, Schramm H, Gysbers A, Hartmann T, Klimmt C, Ravaja N, Sacau A, Baumgartner T, Jäncke P (2004) MEC Spatial Presence Questionnaire (MEC-SPQ). Short Documentation and Instructions for Application. Report to the European Community, Project Presence: MEC (IST-2001-37661)
27. Ward A, Jones A, Hopper A (1997) A new location technique for the active office. IEEE Pers Commun 4:42–47. doi:10.1109/98.626982
28. Yuan Q, Chen IM (2013) 3-D localization of human based on an inertial capture system. IEEE Trans Robot 29:806–812. doi:10.1109/TRO.2013.2248535
29. Yuan Q, Chen IM, Lee SP (2011) SLAC: 3D localization of human based on kinetic human movement capture. In: Proceedings - IEEE International Conference on Robotics and Automation, pp 848–853

Jongkyu Shin is a Ph.D. student at the Music and Audio Research Group (MARG) in the Graduate School of Convergence Science and Technology at Seoul National University, Seoul, Korea. He received his B.S. degree in Business Administration from Hongik University Seoul, Korea, in 2009 and his M.S. degree in Engineering from Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea in 2014. His main research interest is in audio-based sensor networks and its application to human computer interaction.

Gwangseok An is a Ph.D. student at the Music and Audio Research Group (MARG) in the Graduate School of Convergence Science and Technology at Seoul National University. He received his B.S. degree in Electronic Engineering from the Myongji University in 2012. His main research interest is in digital audio signal processing for embedded systems.

Joon-Sang Park received the M.S. degree in computer science from the University of Southern California in 2001 and the Ph.D. degree in computer science from the University of California, Los Angeles (UCLA), in 2006. He is currently an Associate Professor with the Computer Engineering Department, Hongik University, Seoul, Korea. His research interests include various aspects of mobile computing and sensor networks.

Seung Jun Baek received the B.S. degree from Seoul National University in 1998, and the M.S. and Ph.D. degrees from the University of Texas at Austin in 2002 and 2007, respectively, all in electrical and computer engineering. From 2007 to 2009, he was a Member of Technical Staff with the Communications Systems Lab, DSPS R&D Center, Texas Instruments, Dallas, TX. In 2009, he joined the College of Information and Communications, Korea University, Seoul, Korea, where he currently is an associate professor. His research interests include modeling and analysis of broadband access networks, wireless ad hoc and sensor networks.

Kyogu Lee received the B.S. degree in Electrical Engineering from Seoul National University, Seoul, Korea, in 1996, the M.M. degree in Music Technology from New York University, New York, in 2002, and the M.S. degree in Electrical Engineering and the Ph.D. degree in Computer-based Music Theory and Acoustics from Stanford University, Stanford, CA, in 2007 and 2008, respectively. He worked as a Senior Researcher in the Media Technology Lab at Gracenote from 2007 to 2009. He is now an associate professor in the Graduate School of Convergence Science and Technology at Seoul National University, Seoul, Korea and is leading the Music and Audio Research Group. His research focuses on signal processing and machine learning applied to music/audio.