

Article

Full-Body Motion Capture-Based Virtual Reality Multi-Remote Collaboration System

Eunchong Ha ¹, Gongkyu Byeon ¹ and Sunjin Yu ^{2,*}

¹ Department of Culture and Technology Convergence, Changwon National University, Changwon 51140, Korea; chong3286@naver.com (E.H.); bgk9517@gs.cwnu.ac.kr (G.B.)

² Department of Culture Technology, Changwon National University, Changwon 51140, Korea

* Correspondence: sjyu@changwon.ac.kr

Abstract: Various realistic collaboration technologies have emerged in the context of the COVID-19 pandemic. However, as existing virtual reality (VR) collaboration systems generally employ an inverse kinematic method using a head-mounted display and controller, the user and character cannot be accurately matched. Accordingly, the immersion level of the VR experience is low. In this study, we propose a VR remote collaboration system that uses motion capture to improve immersion. The system uses a VR character in which a user wearing motion capture equipment performs the same operations as the user. Nevertheless, an error can occur in the virtual environment when the sizes of the actual motion capture user and virtual character are different. To reduce this error, a technique for synchronizing the size of the character according to the user's body was implemented and tested. The experimental results show that the error between the heights of the test subject and virtual character was 0.465 cm on average. To verify that the implementation of the motion-capture-based VR remote collaboration system is possible, we confirm that three motion-capture users can collaborate remotely using a photon server.

Keywords: motion capture; remote collaboration; multi-user virtual reality; virtual reality



Citation: Ha, E.; Byeon, G.; Yu, S. Full-Body Motion Capture-Based Virtual Reality Multi-Remote Collaboration System. *Appl. Sci.* **2022**, *12*, 5862. <https://doi.org/10.3390/app12125862>

Academic Editors: Chang-Hun Kim and Soo Kyun Kim

Received: 6 May 2022

Accepted: 6 June 2022

Published: 9 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Because of the prolonged COVID-19 pandemic, non-face-to-face technology is rapidly developing in all industries [1]. Accordingly, the demand for non-face-to-face remote collaboration technology has increased significantly [2]. In general, remote collaboration systems use webcam-based video conferencing systems such as Google Teams and Zoom [3]. However, these video collaboration systems provide limited visual information [4]. To overcome the limitations of existing video collaboration systems, virtual reality (VR)/augmented reality (AR)-based realistic collaboration systems are emerging [5]. According to the Immersive Technologies Market Research Report published by Transparency Market Research, the revenue of the global immersive technology market in 2020 was valued at 81.82 Bn. By 2031 and is expected to exceed 2.6 Tn. Moreover, it is expected to expand at a compound annual growth rate of approximately 38% during the projected period [6].

After a VR head-mounted display (HMD) was initially developed by Ivan Sutherland in 1968, he began to research input-processing technology to control virtual physical movements [7]. Subsequently, graphic processing and 3D stereoscopic sound technologies were developed. To expand the multi-sensory experience, various companies have developed data gloves and haptic suits [8]. When a user's illusion of body ownership occurs through a multi-sensory experience, the sense of immersion in the virtual environment increases. Body ownership is the sense of belonging to the body in the field of cognitive neuroscience [9]. A representative experiment is the "Rubber Hands Illusion" experiment published by Botvinick and Chen published in "Nature" in 1998 [10]. This experiment proved that when we transmit complex stimuli such as sight and touch stimuli to the

brain, we can mistake other objects for our own bodies. In 2008, Petkova and Ehrsson demonstrated through a study using a camera and VR that users can feel the body of a character in the virtual world as if it were their own [11]. When the user can see the virtual body from a first-person perspective, the sense of body ownership increases [12,13]. To create an avatar from a first-person view, using VR technology is necessary to build a virtual environment that can be simulated. A realistic collaboration system with VR technology can enhance the effects of collaboration through information sharing, an immersive sense of presence, and natural interactions [14]. However, most VR immersive collaboration systems use the inverse kinematic (IK) method with an HMD and controller [15–17]. In this approach, the joints of the upper skeleton of the elbow or shoulder follow depending on the position of the lower skeleton, such as the hand [18]. Because the actual elbow or shoulder value is calculated and used as the median value, the user's behavior cannot be the same as that of the VR character. In this study, we propose a motion capture-based VR remote collaboration system for improving the IK method. This system implements a virtual environment that can synchronize users and characters in VR in real-time through full-body motion capture. Body motion data are obtained via a motion capture device and applied to a virtual character in the Unity development engine. An environment is created from the user's first-person perspective so that the user can observe from the character's perspective in the virtual environment. When the user accesses the virtual environment, the height of the HMD is measured to synchronize the sizes of the user and character. The accuracy of the character size synchronization system is measured by obtaining the error value between the actual user's height and a character key through an experiment. Photon Unity Network 2 (PUN2) was used to remotely collaborate between virtual characters. When three people wearing motion capture equipment collaborated remotely with a remote partner, each user was given a different network ID. The feasibility of the remote collaboration system utilizing the proposed motion capture was verified. The proposed system provides users with a virtual reality experience that maximizes telepresence.

2. Literature Review

2.1. Immersive Collaboration Platform

An immersive collaboration system enhances the effectiveness of collaboration through rich information sharing and immersive interactions using VR and AR technologies [19]. Realistic collaboration technology can be used to share the work to be solved by reducing the ambiguity of communication. [20]. Problems can be solved accurately and quickly, even in non-face-to-face conditions. An immersive collaboration system supports various collaboration environments, such as idea visualizations or meetings, model design and evaluation environments, and presentation services [21,22]. Table 1 analyzes 28 immersive collaboration systems as classified into four main items. All 28 immersive collaboration platforms can be accessed through VR. Among them, five platforms can be accessed through AR, and 22 platforms can be accessed through a PC. In addition, 13 platforms can be accessed through mobile devices such as smartphones and tablets. In the item analyzing the avatar type in the VR environment, four platforms have avatar types of head and hands. There is only one platform with the avatar type of only showing the head and upper body. There are 13 platforms where avatars have a head, hand, and upper-body shapes. Four platforms implement the avatar's head, arms, hands, and upper body. Five platforms use only the HMD and controller and use IK to implement the movements of the entire avatar body. One platform naturally implements the movement of the entire avatar body according to motion-capture equipment. Of the 28 platforms, 22 can freely convert the face shape of the avatar. Among them, spatial can generate an avatar by photographing a user's face with a camera. At least six people can collaborate in one room of the network on all platforms. Platforms such as FrameVR can accommodate up to 100 people.

Table 1. Immersive collaboration platform analysis.

Immersive Collaboration Platforms														
Number	System	Cross-Platform				Avatar Type (VR Environment)						Custom Avatars	Maximum User in a Room	
		VR	AR	PC	Mobile	Head	Arm	Hand	Torso	Full-Body				
										IK	Motion Capture			
1	Softspace [23]	O				O		O						12
2	Dream [24]	O				O		O						6
3	Sketchbox [25]	O		O		O		O						16
4	VIZIBLE [26]	O		O		O		O						6
5	FrameVR [27]	O		O	O	O			O			O		100
6	Rumii [28]	O		O	O	O		O	O			O		40
7	AltspaceVR [29]	O		O		O		O	O			O		50
8	Glue [30]	O		O	O	O		O	O			O		10
9	Immersed [31]	O		O		O		O	O			O		12
10	meeting Room [32]	O		O	O	O		O	O			O		12
11	Foretell Reality [33]	O		O		O		O	O			O		50
12	WondaVR [34]	O		O	O	O		O	O			O		50
13	BigScreen [35]	O				O		O	O			O		12
14	VSPatial [36]	O				O		O	O			O		16
15	MozillaHubs [37]	O		O	O	O		O	O			O		25
16	RecRoom [38]	O		O	O	O		O	O			O		40
17	Mesh [39]	O	O	O	O	O		O	O			O		8
18	MeetingVR [40]	O	O	O		O		O	O			O		32
19	Horizon Workrooms [41]	O				O	O	O	O			O		16
20	Arthur [42]	O		O		O	O	O	O			O		70
21	The Wild [43]	O	O	O	O	O	O	O	O					8
22	Spatial [44]	O	O	O	O	O	O	O	O			O		50
23	VR Chat [45]	O		O						O		O		40
24	Neos VR [46]	O		O						O		O		20
25	ENGAGE [15]	O		O	O					O		O		70
26	Vive sync [16]	O		O	O					O		O		30
27	MOOTUP [17]	O	O	O	O					O		O		50
28	V-Armed [47]	O									O			10

Figure 1 summarizes the avatar types found in the virtual world as six categories. Each realistic collaboration platform uses an avatar type based on its characteristics. Unlike hand-only avatars, full-body avatars can improve a user's presence in a virtual environment [48]. Rendering the user's body in VR increases immersion and the illusion of being in the virtual world [49]. Full-body tracking technology using full-body motion capture equipment can match body motion data with virtual character movements, thereby enabling specific body information transfer [50]. In addition, in previous studies, users felt that it was easier to visually track full-body characters and understand their behaviors than characters with only hands or upper bodies [51]. As an immersive VR environment is experienced by the user from a first-person perspective, it is essential to have a virtual character body. Therefore, it can be confirmed that a full-body full-tracking system provides higher user

immersion than a system using only hand information or tracking only hands and arms using HMD information.

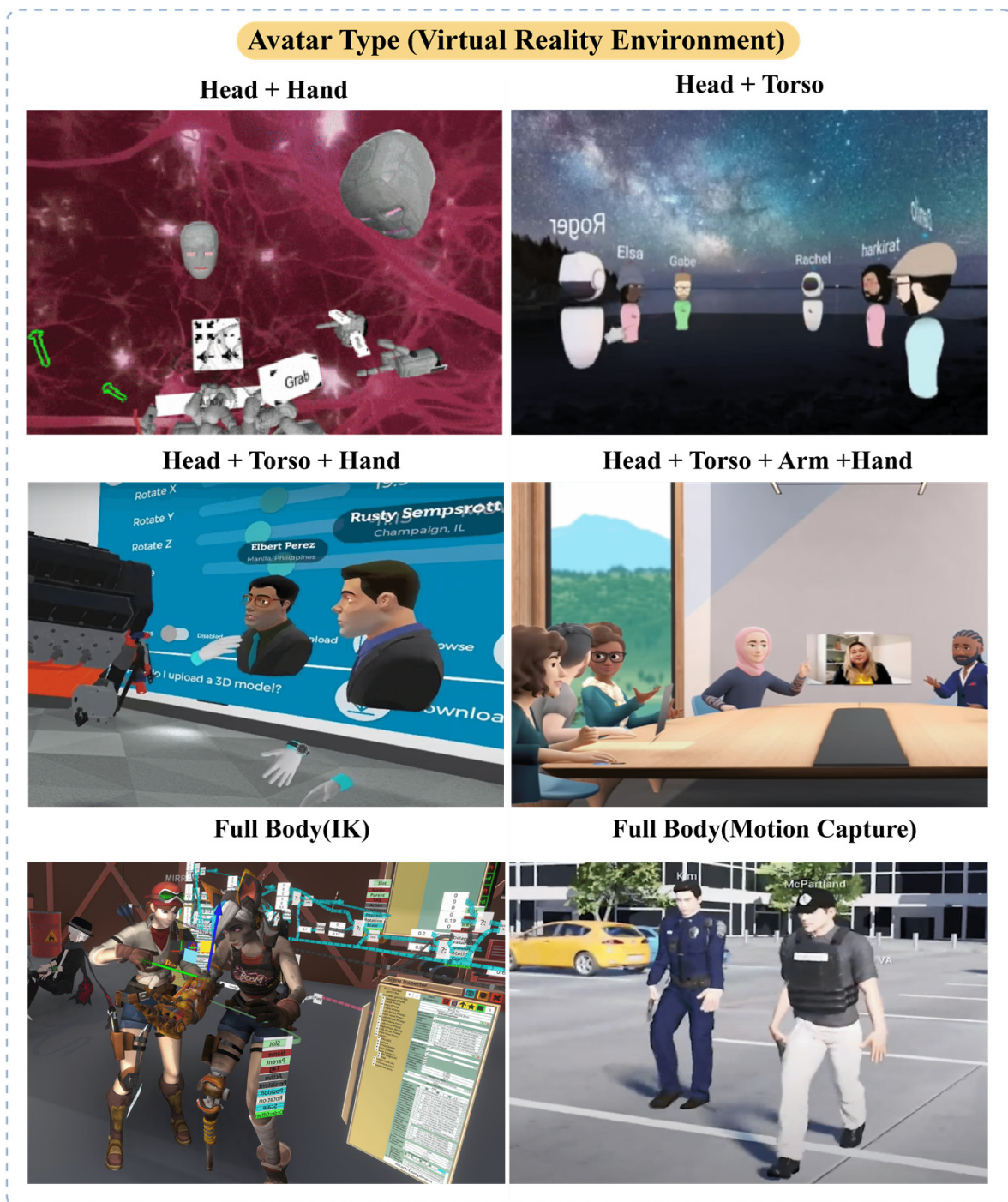


Figure 1. Avatar type (virtual reality environment).

2.2. Motion Capture

Motion capture refers to a technology for extracting specific information using an algorithm using data concerning the movement, position, speed, and direction of objects, such as moving people and animals [52]. Motion capture is used not only in movies, but also in video games, sports, medicine, robotics, and defense. An early form of motion capture was the Rotoscoping technique devised by filmmaker Max Fleischer in 1915. In this method, the movement of an actor was recorded with a camera and then moved and

drawn one-by-one for each frame. Disney's first feature film, "Snow White and the Seven Dwarves," was produced using the Rotoscoping method. This motion capture method provided a fluidity and realism not seen in previous animations [53]. Beginning in the 1970s, filmmakers designed new methods for capturing actors' movements using 3D optical technology. A computer was used to calculate the movements of an actor marked with a marker using a camera. Hollywood applied this approach to a 3D character named Jar Jar Binks from Star Wars: The Phantom Menace in 1999. Since then, motion capture has developed as a camera performance method based on capturing motion data, and new motion capture methods have emerged [54]. Currently, motion capture technology is being used in many fields such as games, healthcare, and broadcasting. EA (Electronic Arts) used motion capture and artificial intelligence technology to capture the characteristics of a soccer player's actual movement to obtain data. By applying this hyper-motion technology to FIFA 22, realism and realism were increased [55]. Home-based virtual PT systems that use motion tracking sensors to monitor patient motions and provide guidance services are also of increasing interest [56,57]. Motion capture makes it easy to capture the movement and expression of a subject. In the field of performing arts, motion capture technology is used to document, store, and learn the movements of dance, theater and martial arts [58]. Virtual influencers who perform virtual character activities using motion tracking equipment, face motion capture, and artificial intelligence technology are also spreading around the world [59].

As shown in Table 2, motion capture technologies can be classified into optical marker-based, marker-less, and inertial measurement methods [60]. The optical marker-based method marks a capture target with a marker to improve the measurement accuracy. This method requires dozens of cameras, as all markers on the subject must be always tracked. Although the accuracy of the body data is high, the process is time-consuming and expensive because it can only be shot in a fixed indoor studio. Optical marker-based imaging equipment companies include OptiTrack, Motion Analysis, and VICON.

Table 2. Three types of motion capture.

Title	Optical Marker-Based Method	Markerless Measurement Method	Inertial Measurement Unit
Company	OptiTrack, Motion Analysis VICON, ART	Microsoft—Kinect	Xsens, Noitom, Nansense, Rokoko
Capture method	Optical camera data	Kinect camera data	IMU(Inertial Measurement Unit) inertial sensors data
Cost	Very High cost	Low cost	High Cost
Available places	Indoor studio	Indoor	Indoor, Outdoor

Microsoft Kinect is an optical markerless device. The Kinect V2 equipment can analyze a target using a camera sensor, infrared sensor, and depth sensor using 25 skeleton joint point algorithms [61]. Using this method, motion data can be obtained without attaching markers. However, the image processing method is inaccurate relative to those using markers or sensors because there is no clear reference point. Moreover, if the number of tracking personnel increases or the object is obscured, errors occur in the measurement values [62].

The inertial measurement method acquires motion data by attaching a sensor to the body joints of the user. Unlike marker-based methods, it is used for outdoor shooting or action scenes owing to its high portability and spatial freedom. However, when the sensor is used for a long time, the error range of the data widens. Companies related to inertial measurement equipment include Xsens, Noitom, NanSense, and Rokoko.

Figure 2 shows the differences between the motion capture methods. Motion capture methods should be adopted and used with consideration of the characteristics and user environment of the platform to be manufactured.



Figure 2. Three types of motion capture.

3. Proposed System

The system proposed in this study uses inertial measurement unit (IMU) inertial measurement equipment for capturing motion in both indoor and outdoor locations. Several studies have demonstrated that IMUs can provide precise tracking of upper and lower bodies [63–65]. Accordingly, we developed an improvement plan for character movement appearing in the existing IK method system. First, we implemented a VR remote collaboration system using motion capture equipment and matched the user's body data with the body data of the virtual character. Second, three researchers collaborated to verify that the proposed system was practically usable.

3.1. Proposed System Development Environment

We used a PC with the following specifications and software for our development environment.

- OS: Window 10
- CPU: Intel® Core™ i9-10980HK CPU @ 2.40 GHz 3.10 GHz
- RAM: 32.0 GB
- GPU: NVIDIA GeForce GTX 2080 Super
- Framework: Unity 2019.4.18f, Visual Studio 2019
- Language: C#
- Hardware: Perception Neuron Studio, HTC Vive Pro Eye, Oculus Rift S
- Software: Axis Studio Software, Unity

The proposed system used the C# script language and Visual Studio 2019 to create a VR remote collaboration environment with Unity (2019.4.18f). Unity is a development engine used in various applications, including PC, VR, AR, and mobile devices. It supports a cross-system that can be distributed to various devices through the development of one application [66,67]. The XR Interaction Toolkit plugin was used for the VR interaction. This was used to create a multi-variety VR platform able to use both the HTC Vive Pro Eye and Oculus Rift S. For remote collaboration, a Unity-only Photon PUN2 server was used. For real-time motion capture, the Perception Neuron Studio (PNS) IMU inertial measurement equipment and gloves were used. The PNS uses an inertial body sensor, an independent small sensor with an inertial measurement method equipped with a gyroscope, accelerometer, and magnetometer mounted on a strap. The user's movements can be measured in real time using embedded data fusion, anthropometric, and physical engine algorithms embedded in the sensor [68]. Ryan Sers tested the accuracy and effectiveness of the perception neuron motion capture device in terms of its ability to measure the range of motion (ROM) of an upper body posture [50]. According to the study, the device had an average ROM difference of less than 5°. Root mean square deviation (RMSD) is suitable for expressing the precision, as it is used when dealing with the difference between a predicted result value and the value observed in the real environment. In Ryan Sers' experiment, the RMSD was properly below 4°, providing that the error range was low, and that the measurement precision was valid for evaluating upper-body motion [68]. In addition, the PNS shows better precision and performance than the PN3 from Noitom in 2021. PN3 is

capped at a 60 fps capture rate, whereas PNS can track fast movements at up to 240 fps. PN3 can track 3 performers simultaneously, whereas PNS can track up to 5.

3.2. System Development

The motion capture-based VR multi-remote system is shown in Figure 3. The users wore motion capture equipment and HMDs. The motion data from the motion capture devices were received by the Axis Studio software through data conversion. In addition, Axis Studio Software acted as a hub to receive and store data from the motion capture equipment. The procedure was as follows. (1) Connect the motion capture transceiver to the computer. (2) Link the IP address of the motion capture transceiver with the computer IPV4 settings. (3) Place the PNS Socket Type to TCP and Local Address IP/Port number. (4) Add the SDK used by integrating Axis Studio Software and Unity to the Unity project. In this project, we used the Neuron Unity Integration 0.2.19 Plugin. (5) Attach a script that can receive motion capture device data communication provided by the Neuron Unity integration Plugin to the PNS model to be used. Unity uses TCP/IP communication to link the motion capture data and characters. Each user's character data were network-synchronized using a photon server. The users wearing motion capture equipment could collaborate remotely using a character with the same size as their body size in a virtual environment.

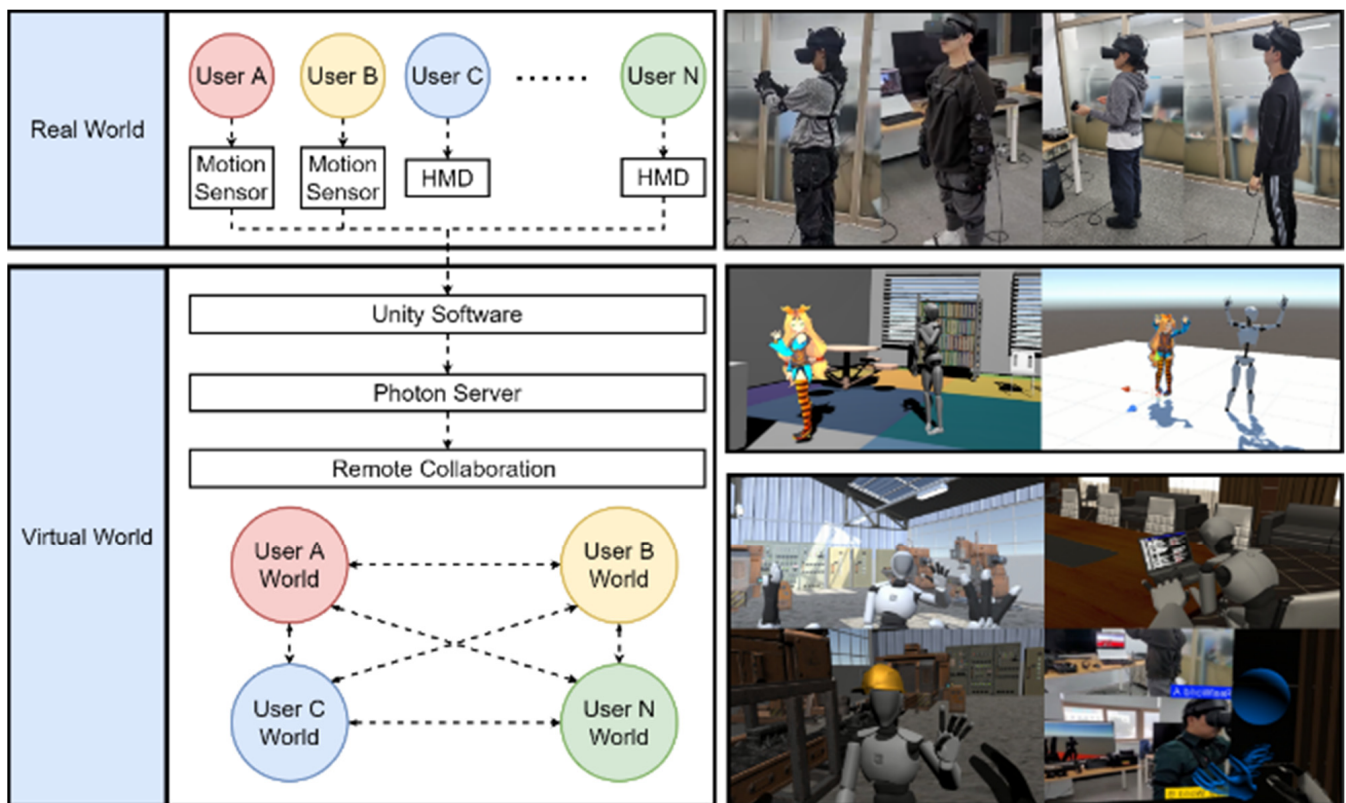


Figure 3. Motion capture-based multi-remote realistic collaboration system process.

As shown in Figure 4, the motion data of the PNS equipment performed the user motion data initialization process and posture calibration process using Axis Studio. The motion data were converted to BVH and transmitted to Unity in real time through TCP/IP communication.



Figure 4. Perception Neuron Studio (PNS)/Axis Studio connect.

To apply the motion data sent to Unity to a character, a character-retargeting process was required. Character retargeting is a method for minimizing the error between an applied character joint and the original data joint [69]. Figure 5 shows a scene in which one piece of motion data is retargeted to two Unity characters of different sizes.

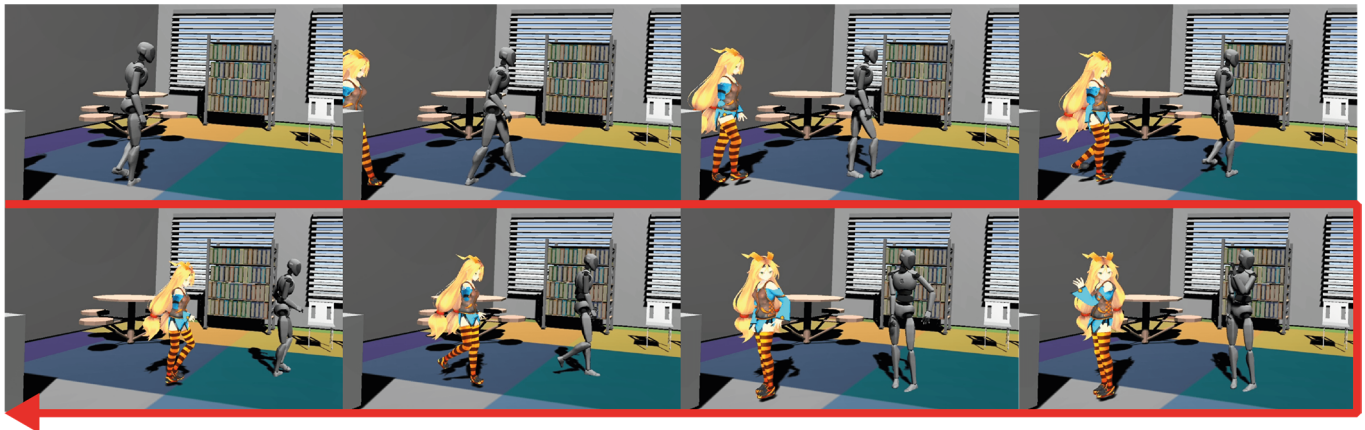


Figure 5. Real-time motion capture in Unity.

The XR Interaction Toolkit was used to create the VR environment in Unity. If an HMD and Unity camera are connected to the XR Rig of XR Interaction, the user can see the Unity VR space on the HMD screen. To make the virtual character's body move according to thoughts like one's own in a VR environment, the virtual character is implemented in a first-person view while the user is wearing an HMD. When integrating and synchronizing the HMD with PNS, there are three core rules. First, the rotation values of the PNS model's head cannot be used, as the HMD's rotation values must be given priority when the camera follows the character's eyes. In addition, the transform value of the HMD cannot be used. Finally, the HMD camera node cannot be set as the parent of the PNS model skeleton [54].

Figure 6 shows a method for performing a remote multiplayer experience through a photon server. In the picture there are users named A, B, C, and D and characters a, b, c, and d that each user can use. The first room creator, Master A, creates a world. In general, it is easy to be mistaken for a form in which characters a, b, c, and d all participate in World A. In fact, each user creates his/her own world and transmits information to the clones of each world. When the remaining users connect to the network, Worlds B, C, and D are created. In that world, all characters except your own are clones from other worlds. The remote character is different from the local character, but the same identifier is used. For example, remote character A of World B has the same identifier as local character A of World A, so the information of local character A can be synchronized. In this way, local and remote characters are created in every world, and information is synchronized with characters in other worlds through the server. Therefore, if there are four users connecting to the network, four identical worlds will be created, and 16 characters will be created.

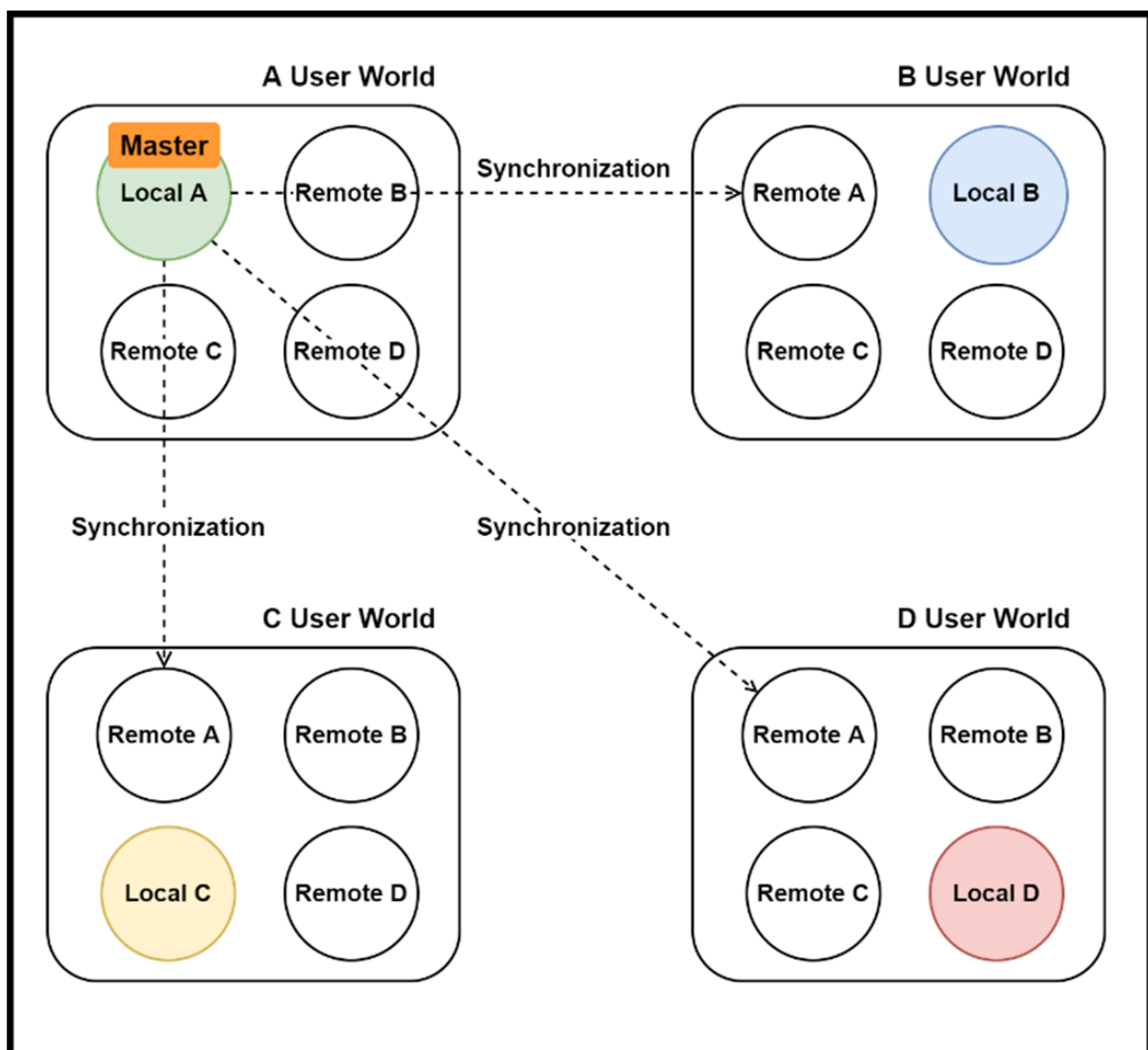


Figure 6. Photon network synchronization.

Figure 7 shows whether two users of motion capture equipment are remotely linked in a VR network environment using the photon server. Users enter a room through the photon server. Each user is created with a prefab character corresponding to the stored unique ID. At this time, a script is activated to receive motion capture data only for characters created in their own world to avoid network errors.



Figure 7. Remote collaboration system demo.

4. Experiments and Results

In the manufactured remote collaboration system, the size of the prepared basic character is predetermined, so there is a difference in body height between the real user and virtual character. Owing to this difference, the actions of the user and the character do not match, and the feeling of immersion is reduced. To improve this, a method was implemented to reduce the gap between the real and virtual environments by matching the height of the virtual character to that of the user. Using the implemented experimental environment, the error value for the height of the character, insofar as matching the actual user's height, was measured. The experimental environment was the same as the development environment. The experiment was conducted using an Oculus Rift S HMD.

4.1. Character Scale Synchronization

In Figure 8, A stands for User A's Height, B stands for the character height, C stands for the HMD height, and D stands for the character eye height. The value of A is calculated from the values of C, B, and D using a proportional formula. The basic eye height value D of the virtual character is divided from the height value (Localposition.y) C of the HMD worn by the user. Through this, the ratio difference between the HMD's height value and the character's eye height value is calculated. This ratio difference value is defined as the scale value (C/D). After that, the HeadEnd component is applied to the end of the character's head. If measuring the HeadEnd LocalPosition.y, the character's height B is measured. Then, by multiplying the scale value by the height value B of the character's head, the actual user's height value A can be obtained. If the size of the character is adjusted up to the defined height of A, the heights of the user and character are synchronized. There are two further issues to consider. First, it is necessary to adjust the character size to match the actual camera position of the HMD (user's eye level) and character's eye level. Second, after synchronization, the virtual camera should follow the position of the character's eyes.

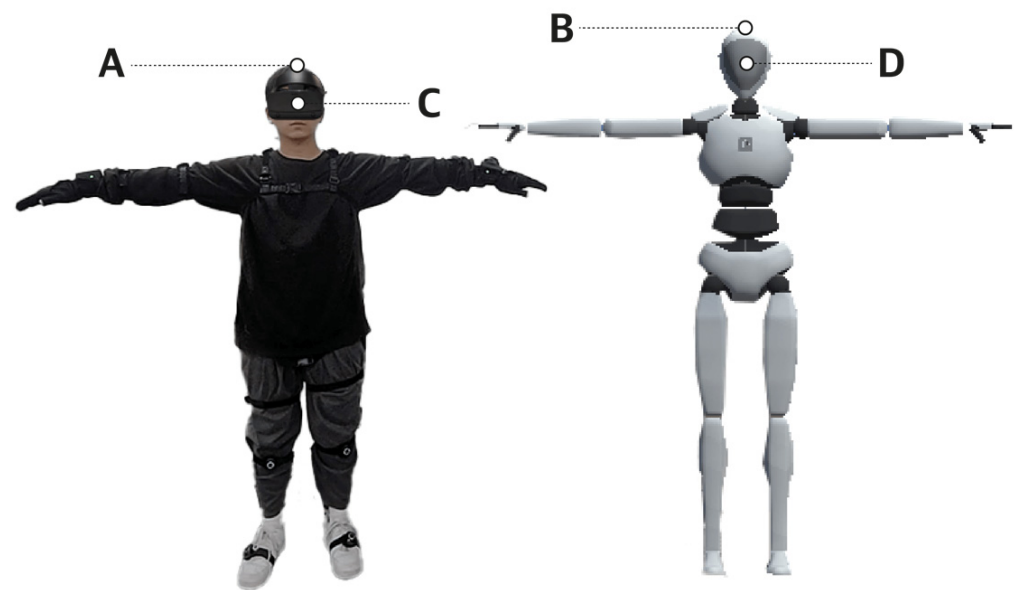


Figure 8. How to synchronize character scale according to user body information.

Unity's basic unit is 1 m. The height of the basic character is 1.8 units, so when converted to m, it becomes 1.8 m. As shown in Figure 9, if the height of the user wearing the HMD is 1.7 m, the user's gaze rises by as much as the gaze of the character (who is 1.8 m) by the retargeting process. Because a user with a height of 1.7 m corresponds to a first-person view of a 1.8 m character, there is a difference between the real and virtual movements. Therefore, actions such as grabbing an object in the virtual environment cannot be precisely performed.

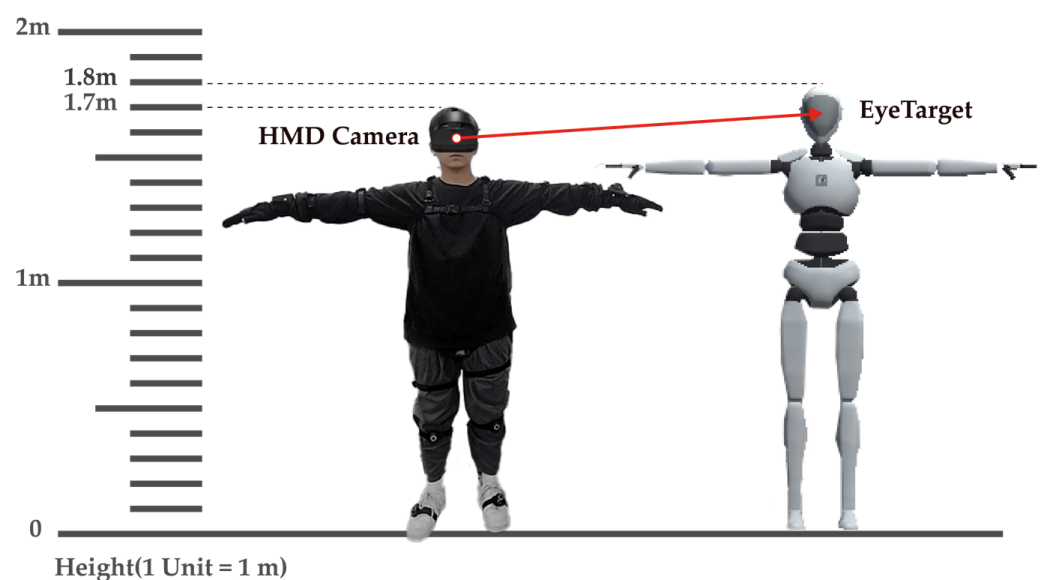


Figure 9. Existing character size synchronization method.

After the retargeting process, the size of the character in VR should be converted to the same size as the real user. Figure 10 illustrates the task of adjusting the character to have the same height as the user's body.

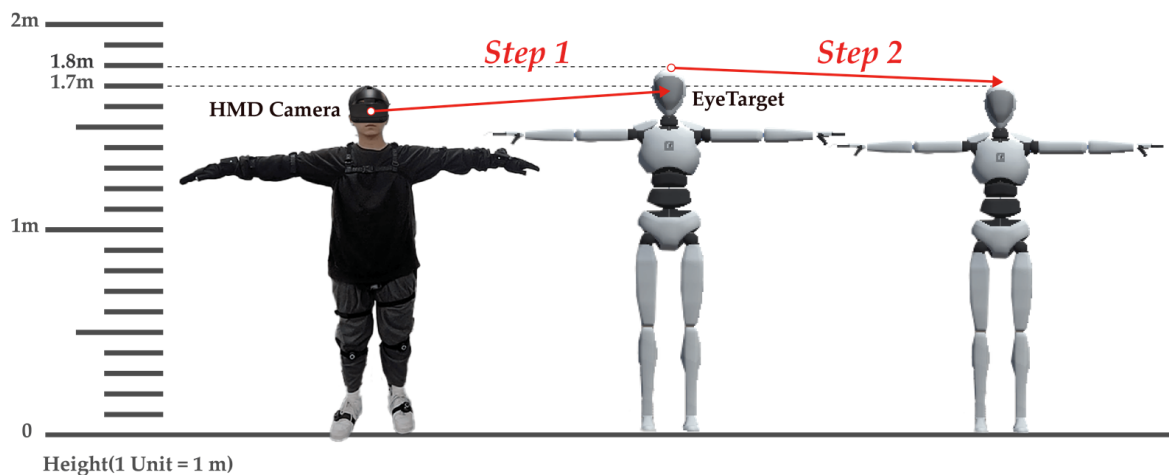


Figure 10. Suggested character size synchronization method.

Experiments were performed with an Oculus Rift S Perception Neuron motion capture device. Eight participants experimented at the same location. The experiment first measured the actual height of the participant on the wall using a T-pose motion with an automatic extensometer. Second, the user wore motion capture equipment and went through a calibration process. Then, they put on the HMD. Next, they followed instructions to connect to the virtual system (connect to the photon server to connect to the network, and then connect through avatar selection and room selection). Subsequently, the character height was synchronized using the same T-pose as when measuring the height from the wall. The participants repeatedly performed the same motion for measurement ten times. After wearing the motion capture device, they faced the front of the measurement system and stayed in the T-pose posture for 3 s. Then, the height of the HMD was used to calculate the height of the user.

The experimental participants consisted of six males and two females in their 20s and 30s (six in their 20s and two in their 30s). All participants had VR experience but no experience with wearing motion capture equipment and HMD equipment simultaneously.

Table 3 shows the average error values obtained by subtracting the key value of the character as synchronized and changed when connecting 10 times in VR from the actual key. The average error of the experimenters was 0.465 cm. The experimenter with the largest difference showed a difference of 0.89 cm, whereas the experimenter with the smallest difference showed a difference of 0.15 cm. Using the proposed system, it can be confirmed that the error in the character's body size was less than 1 cm on average when the user's body scale was synchronized.

Table 3. Character size synchronization error value calculation [Unit: cm].

Test Group	Real World Height	Virtual Character Average Height	Difference Value
subject 1	163	162.85	0.15
subject 2	163	162.12	0.88
subject 3	164.3	164.07	0.23
subject 4	164	163.11	0.89
subject 5	167.5	167.36	0.14
subject 6	173	172.87	0.13
subject 7	173.5	173.09	0.41
subject 8	180	179.11	0.89
Mean			0.465

4.2. Multi Remote Collaboration System

Table 4 shows the experimental environment. Experiments were conducted in different buildings using different networks. Considering convenience of movement, all experiments were conducted with three laptops. To support a variety of models, HTC Vive Pro Eye, Oculus Rift S, and Oculus Quest 2 were used as VR devices. There was a risk of colliding with real objects during the experiment, so the surrounding environment was cleaned up and tested.

Table 4. Experiment environment.

	PC 1	PC 2	PC 3
CPU	Intel® Core™ i9-10980HK CPU @ 2.40 GHz 3.10 GHz	Intel® Core™ i7-10870H CPU @ 2.20 GHz 2.21 GHz	Intel® Core™ i7-10870H CPU @ 2.20 GHz 2.21 GHz
RAM	32.0 GB	16.0 GB	16.0 GB
GPU	NVIDIA GeForce GTX 2080 Super	NVIDIA GeForce RTX 3070	NVIDIA GeForce RTX 3070
VR Device	HTC Vive pro Eye	Oculus Rift S	Oculus Quest 2

All three users wore VR HMD devices and accessed the proposed system. Before entering the room, it could be chosen if the motion capture device was connected or only the HMD was connected. All users selected an avatar and entered the desired room. As shown in Figure 11, a total of three environments were created for collaboration: factory, meeting room, and an exercise room. When users entered the system, they could meet users from other places connected at the same time. Users could communicate through Photon Voice and collaborate while listening to each other's voices.

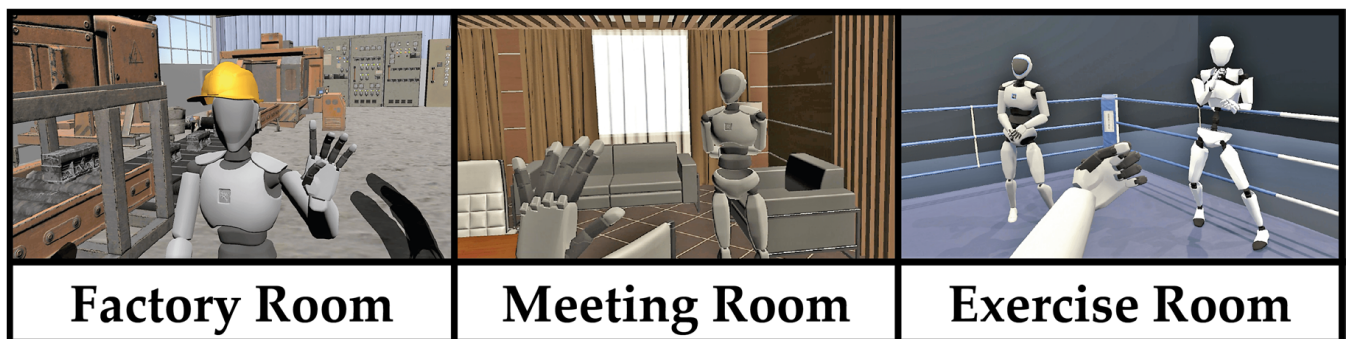


Figure 11. Multi-remote collaboration.

Using the following system, detailed training is possible because all body movements can be expressed in a virtual environment. In this experiment, remote collaboration was performed in an exercise room using three motion capture devices. Figure 12 shows the VR View of User1. Training of experts and beginners is also possible in this system. User 1 is an expert who has learned boxing for more than 3 years, and User 2 and User 3 are beginners who have never tried boxing. An expert at Use 1 teaches beginners the jab and straight punch. Using motion capture, we were able to train the wrist and the position of the foot movement.

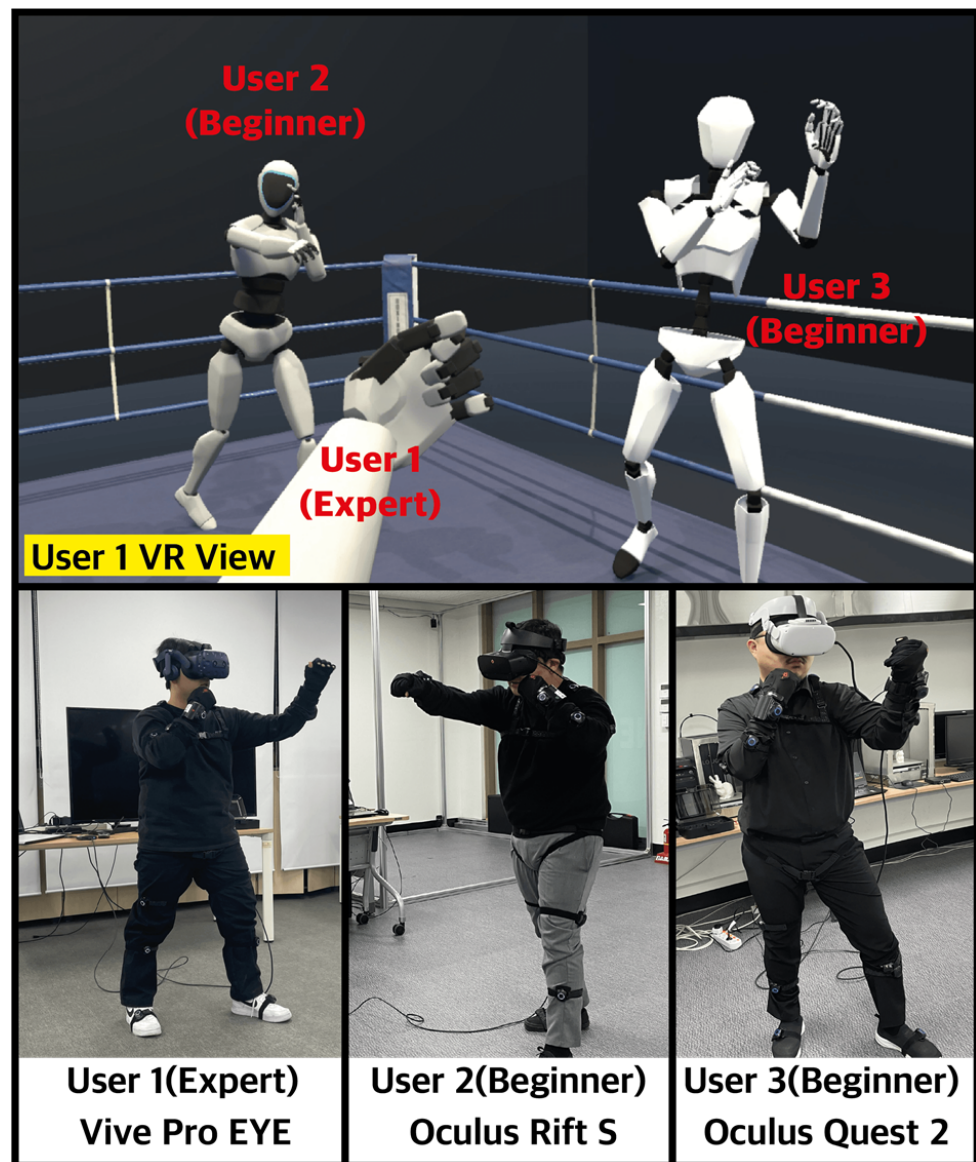


Figure 12. Multi-remote collaboration Test.

5. Discussion

As mentioned above, a character-size matching experiment was conducted to improve the immersion of a VR remote collaboration system using motion capture. The experiment was conducted ten times by eight people each. When the error value was obtained by subtracting the key value of the synchronized character in VR from the actual height, an error value of 0.465 cm was obtained. By using the improved system, the error between the user's height and virtual character's key was reduced, thereby increasing the immersion. It was confirmed that the system could be run by three people remotely accessing the system. Our study had some limitations. First, owing to the nature of non-optical motion capture equipment, the longer the capture time, the lower the precision. As the inconvenience caused by wearing the motion capture device and VR HMD device simultaneously was not resolved, the low usefulness of long-term training needs to be addressed. Second, it is necessary to compensate for the possibility of network delays depending on the Internet situation during remote collaboration. If the network environment is improved by applying 5G in the future, it will be possible to work smoothly without network delays in remote collaboration situations. Third, we limited the number of trial participants owing to COVID-19.

6. Conclusions

Owing to the impact of COVID-19, the importance of non-face-to-face remote collaboration technologies is growing, and VR/AR immersive content-based immersive collaboration platforms are emerging in various companies. An immersive collaboration platform is more expressive than a 2D video remote platform and has the advantage of being able to accurately share a common goal. When using a full-body avatar to conduct realistic collaboration in a virtual environment, the sense of ownership of the body increases and immersion improves. However, as a result of examining 28 existing immersive collaboration platforms, it was noted that 22 did not use full-body characters. Five of the six platforms that used full-body avatars used the IK method. However, because the IK method processes the rest of the arm by computer calculations according to the movement of a controller, errors such as bent arms can occur. Whole-body character tracking using motion capture has the advantage of being able to deliver more information when performing remote collaboration than characters with only hands or the upper body. In this study, motion capture-based full-body avatars were used in a virtual environment and remote collaboration was attempted through simultaneous access by three or more people. A script activation method accepted motion capture data only for one character when three or more motion capture devices were connected remotely, and the error of the motion data overlapping with that of other characters was improved.

Future research will plan scenarios with collaborations, such as in telemedicine or sports rehabilitation, to further analyze motion capture data. In addition, we will attempt an approach using digital twin technology for transferring motion capture data to a robot such as a manipulator.

Author Contributions: Conceptualization, E.H., G.B. and S.Y.; methodology, E.H.; software, E.H.; validation, E.H. and G.B.; formal analysis, E.H. and G.B.; investigation, E.H.; data curation, E.H.; writing—original draft preparation, E.H.; writing—review and editing, E.H. and G.B.; visualization, E.H.; supervision, S.Y.; project administration, S.Y.; funding acquisition, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2020R1F1A1073866). It was also supported by a Manufacturing Process Innovation Simulation Center Establishment Project funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Institutional Review Board Statement: Not applicable.

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

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Byrnes, K.G.; Kiely, P.A.; Dunne, C.P.; McDermott, K.W.; Coffey, J.C. Communication, collaboration and contagion: “Virtualisation” of anatomy during COVID-19. *Clin. Anat.* **2021**, *34*, 82–89. [CrossRef] [PubMed]
2. Lee, D.; Jeong, I.; Lee, J.; Kim, J.; Cho, J. VR content for Non contact job training and experience. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 19–21 October 2021; IEEE: New York, NY, USA, 2021; pp. 486–489.
3. Rahmat, A. An Analysis of Applying Zoom Cloud Meeting towards EFL Learning in Pandemic Era COVID-19. *Br. J. Bhs. Dan Sastra Ingg.* **2021**, *10*, 114–134. [CrossRef]
4. Syaharuddin, S.; Husain, H.; Herianto, H.; Jusmiana, A. The effectiveness of advance organiser learning model assisted by Zoom Meeting application. *Cypriot J. Educ. Sci.* **2021**, *16*, 952–966. [CrossRef]
5. Mahmood, T.; Fulmer, W.; Mungoli, N.; Huang, J.; Lu, A. Improving information sharing and collaborative analysis for remote geospatial visualization using mixed reality. In Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 14–18 October 2019; IEEE: New York, NY, USA, 2019; pp. 236–247.
6. Transparency Market Research. Available online: <https://www.transparencymarketresearch.com/immersive-technologies-market.html> (accessed on 10 April 2022).

7. Sutherland, I.E. A head-mounted three dimensional display. In Proceedings of the AFIPS Fall Joint Computer Conference, Part I, San Francisco, CA, USA, 9–11 December 1968; pp. 757–764.
8. Han, S.; Kim, J. A study on immersion of hand interaction for mobile platform virtual reality contents. *Symmetry* **2017**, *9*, 22. [CrossRef]
9. Gallagher, S. Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn. Sci.* **2000**, *4*, 14–21. [CrossRef]
10. Botvinick, M.; Cohen, J. Rubber hands ‘feel’ touch that eyes see. *Nature* **1998**, *391*, 756. [CrossRef]
11. Petkova, V.I.; Khoshnevis, M.; Ehrsson, H.H. The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol.* **2011**, *2*, 35. [CrossRef] [PubMed]
12. Peck, T.C.; Seinfeld, S.; Aglioti, S.M.; Slater, M. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* **2013**, *22*, 779–787. [CrossRef]
13. Egeberg, M.C.S.; Lind, S.L.R.; Serubugo, S.; Skantarova, D.; Kraus, M. Extending the human body in virtual reality: Effect of sensory feedback on agency and ownership of virtual wings. In Proceedings of the 2016 Virtual Reality International Conference, New York, NY, USA, 23–25 March 2016; pp. 1–4.
14. Truong, P.; Hölttä-Otto, K.; Becerril, P.; Turtiainen, R.; Siltanen, S. Multi-User Virtual Reality for Remote Collaboration in Construction Projects: A Case Study with High-Rise Elevator Machine Room Planning. *Electronics* **2021**, *10*, 2806. [CrossRef]
15. Engage. Available online: <https://engagevr.io/> (accessed on 15 April 2022).
16. Sync. Available online: <https://sync.vive.com/> (accessed on 15 April 2022).
17. Mootup. Available online: <https://mootup.com/> (accessed on 15 April 2022).
18. Jeong, D.C.; Xu, J.J.; Miller, L.C. Inverse kinematics and temporal convolutional networks for sequential pose analysis in vr. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality, Virtual/Online, 14–18 December 2020; IEEE: New York, NY, USA, 2020; pp. 274–281.
19. Tea, S.; Panuwatwanich, K.; Ruthankoon, R.; Kaewmoracharoen, M. Multiuser immersive virtual reality application for real-time remote collaboration to enhance design review process in the social distancing era. *J. Eng. Des. Technol.* **2021**, *20*, 281–298. [CrossRef]
20. Young, J.; Langlotz, T.; Cook, M.; Mills, S.; Regenbrecht, H. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 1908–1918. [CrossRef] [PubMed]
21. Diaz-Kommonen, L.; Vishwanath, G. Accessible Immersive Platforms for Virtual Exhibitions involving Cultural Heritage. In Proceedings of the Electronic Visualisation and the Arts, Florence, Italy, 14 June 2021; Leonardo Libri Srl: Firenze, Italy, 2021; pp. 125–130.
22. Saffo, D.; Di Bartolomeo, S.; Yildirim, C.; Dunne, C. Remote and collaborative virtual reality experiments via social vr platforms. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–15.
23. Softspace. Available online: <https://soft.space/> (accessed on 15 April 2022).
24. Dream. Available online: <https://dreamos.com/> (accessed on 15 April 2022).
25. Sketchbox Design. Available online: <https://design.sketchbox3d.com/> (accessed on 15 April 2022).
26. Worldviz. Available online: <https://www.worldviz.com/virtual-reality-collaboration-software> (accessed on 15 April 2022).
27. Frame. Available online: <https://framevr.io/> (accessed on 15 April 2022).
28. Rumii. Available online: <https://www.dogheadsimulations.com/rumii> (accessed on 15 April 2022).
29. Altspace VR. Available online: <https://altvr.com/> (accessed on 15 April 2022).
30. Glue. Available online: <https://glue.work/> (accessed on 15 April 2022).
31. Immersed. Available online: <https://immersed.com/> (accessed on 15 April 2022).
32. meetingRoom. Available online: <https://meetingroom.io/> (accessed on 15 April 2022).
33. Foretell Reality. Available online: <https://foretellreality.com/> (accessed on 15 April 2022).
34. Wonda. Available online: <https://www.wondavr.com/> (accessed on 15 April 2022).
35. Bigscreen. Available online: <https://www.bigscreenvr.com/> (accessed on 15 April 2022).
36. vSpatial. Available online: <https://www.vspatial.com/> (accessed on 15 April 2022).
37. hubs moz://a. Available online: <https://hubs.mozilla.com/> (accessed on 15 April 2022).
38. Rec Room. Available online: <https://recroom.com/> (accessed on 15 April 2022).
39. Microsoft Mesh. Available online: <https://www.microsoft.com/en-us/mesh> (accessed on 15 April 2022).
40. MeetinVR. Available online: <https://www.meetinvr.com/> (accessed on 15 April 2022).
41. MetaQuest. Available online: https://www.oculus.com/workrooms/?locale=ko_KR (accessed on 15 April 2022).
42. Arthur. Available online: <https://www.arthur.digital/> (accessed on 15 April 2022).
43. The Wild. Available online: <https://thewild.com/> (accessed on 15 April 2022).
44. Spatial. Available online: <https://spatial.io/> (accessed on 15 April 2022).
45. VRChat. Available online: <https://hello.vrchat.com/> (accessed on 15 April 2022).
46. Neos. Available online: <https://neos.com/> (accessed on 15 April 2022).
47. V-Armed. Available online: <https://www.v-armed.com/features/> (accessed on 15 April 2022).
48. Mendes, D.; Medeiros, D.; Sousa, M.; Ferreira, R.; Raposo, A.; Ferreira, A.; Jorge, J. Mid-air modeling with boolean operations in VR. In Proceedings of the IEEE Symposium 3D User Interfaces, Los Angeles, CA, USA, 18–19 March 2017; pp. 154–157.
49. Schwind, V.; Knierim, P.; Tasci, C.; Franczak, P.; Haas, N.; Henze, N. “These are not my hands!” Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 1577–1582.

50. Sers, R.; Forrester, S.; Moss, E.; Ward, S.; Ma, J.; Zecca, M. Validity of the Perception Neuron inertial motion capture system for upper body motion analysis. *Measurement* **2020**, *149*, 107024. [\[CrossRef\]](#)
51. Wang, T.Y.; Sato, Y.; Otsuki, M.; Kuzuoka, H.; Suzuki, Y. Effect of Full Body Avatar in Augmented Reality Remote Collaboration. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; IEEE: New York, NY, USA, 2019; pp. 1221–1222.
52. Gong, S.J. Motion capture and video motion analysis. *Sports Sci.* **2019**, *146*, 40–45.
53. Agarwala, A.; Hertzmann, A.; Salesin, D.H.; Seitz, S.M. Keyframe-based tracking for rotoscoping and animation. *ACM Trans. Graph.* **2004**, *23*, 584–591. [\[CrossRef\]](#)
54. Perception Neuron. Available online: <https://www.neuronmocap.com/> (accessed on 22 April 2022).
55. EA. Available online: <https://www.ea.com/ko-kr/games/fifa/fifa-22/hypermotion/> (accessed on 22 April 2022).
56. Wei, W.; Kurita, K.; Kuang, J.; Gao, A. Real-Time Limb Motion Tracking with a Single IMU Sensor for Physical Therapy Exercises. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 1–5 November 2021; pp. 7152–7157.
57. Wei, W.; Carter, M.; Sujit, D. Towards on-demand virtual physical therapist: Machine learning-based patient action understanding, assessment and task recommendation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1824–1835. [\[CrossRef\]](#)
58. Herrow, M.F.M.; Azraai, N.Z. Digital Preservation of Intangible Cultural heritage of Joget Dance Movement Using Motion Capture Technology. *Int. J. Herit. Art Multimed.* **2021**, *4*, 1–13. [\[CrossRef\]](#)
59. Conti, M.; Gathani, J.; Tricomi, P.P. Virtual Influencers in Online Social Media. *IEEE Commun. Mag.* **2022**, 1–13. [\[CrossRef\]](#)
60. Tee, K.S.; Low, E.; Saim, H.; Zakaria, W.N.W.; Khialdin, S.B.M.; Isa, H.; Awad, M.I.; Soon, C.F. A study on the ergonomic assessment in the workplace. *AIP Conf. Proc.* **2017**, *1883*, 020034.
61. Cho, Y.; Park, K.S. Design and development of the multiple Kinect sensor-based exercise pose estimation system. *J. Inf. Commun. Converg. Eng.* **2017**, *21*, 558–567.
62. Naemabadi, M.; Dinesen, B.; Andersen, O.K.; Hansen, J. Influence of a marker-based motion capture system on the performance of Microsoft Kinect v2 skeleton algorithm. *IEEE Sens. J.* **2018**, *19*, 171–179. [\[CrossRef\]](#)
63. Podkosova, I.; Vasylevska, K.; Schoenauer, C.; Vonach, E.; Fikar, P.; Bronederk, E.; Kaufmann, H. Immersededeck: A large-scale wireless VR system for multiple users. In Proceedings of the IEEE 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems, Greenville, SC, USA, 13–23 March 2016; pp. 1–7.
64. Malleson, C.; Kosek, M.; Klaudiny, M.; Huerta, I.; Bazin, J.C.; Sorkine-Hornung, A.; Mine, M.; Mitchell, K. Rapid one-shot acquisition of dynamic VR avatars. In Proceedings of the IEEE Virtual Reality, Los Angeles, CA, USA, 18–22 March 2017; pp. 131–140.
65. Kishore, S.; Navarro, X.; Dominguez, E.; de la Pena, N.; Slater, M. Beaming into the news: A system for and case study of tele-immersive journalism. *IEEE Comput. Graph. Appl.* **2018**, *38*, 89–101. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Unity. Available online: <https://unity.com/kr> (accessed on 24 April 2022).
67. Kučera, E.; Haffner, O.; Leskovský, R. Multimedia Application for Object-oriented Programming Education Developed by Unity Engine. In Proceedings of the 2020 Cybernetics & Informatics (K & I), Velke Karlovice, Czech Republic, 29 January–1 February 2020; IEEE: New York, NY, USA, 2020; pp. 1–8.
68. Kim, H.S.; Hong, N.; Kim, M.; Yoon, S.G.; Yu, H.W.; Kong, H.J.; Kim, S.J.; Chai, Y.J.; Choi, H.J.; Choi, J.Y.; et al. Application of a perception neuron®system in simulation-based surgical training. *J. Clin. Med.* **2019**, *8*, 124. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Gleicher, M. Retargetting motion to new characters. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, USA, 19–24 July 1998; ACM: New York, NY, USA, 1998; pp. 33–42.