IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 November 2024, accepted 30 December 2024, date of publication 3 January 2025, date of current version 9 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3525630

APPLIED RESEARCH

Designing and Analyzing Virtual Avatar Based on Rigid-Body Tracking in Immersive Virtual Environments

MYEONGSEOK PARK[1], **JIWON LEE**[2], **HYEONYONG YANG**[1,3], **AND JINMO KIM**[1,2,3]
[1]Department of Computer Engineering, Graduate School, Hansung University, Seoul 02876, South Korea
[2]Department of Information and Computer Engineering, Graduate School, Hansung University, Seoul 02876, South Korea
[3]Division of Computer Engineering, Hansung University, Seoul 02876, South Korea

Corresponding author: Jinmo Kim (jinmo.kim@hansung.ac.kr)

This research was financially supported by Hansung University.

ABSTRACT This study proposes a novel method for virtual avatar applications based on rigid-body tracking in immersive virtual environments (IVEs). This method aims to design a method to efficiently and accurately estimate motions defined with few conditions. For this purpose, we design a virtual avatar motion process that proceeds from rigid-body tracking to motion estimation using Final IK (inverse kinematics). This study uses optical motion capture equipment based on markers for rigid-body tracking and calculating the movement of a defined virtual avatar based on the transformation information of the precisely tracked rigid-body object. To analyze the effectiveness of the proposed method, a comparative analysis of equipment and environment with existing research on animation fidelity in self-avatars is performed. In addition, two upper body-centered actions (simultaneous two-hand motion and separate two-hand motion) are presented to analyze latent factors of virtual avatar embodiment. A comparative survey experiment using the virtual embodiment questionnaire (VEQ) is conducted based on full-body tracking and the proposed rigid-body tracking-based method. Furthermore, to compare and analyze users' realism in an environment where they coexist with virtual avatars in IVE, a survey experiment using a presence questionnaire is conducted. Therefore, we confirmed that similar avatar motion generation is possible compared to full-body tracking even with fewer tracking targets (e.g. type of tracking device or number of markers), inputs, and processing steps. Additionally, it is possible to design a virtual avatar that provides satisfactory realism in IVE.

INDEX TERMS Virtual avatar, immersive virtual environment, rigid-body tracking, motion estimation, motion capture, virtual embodiment.

I. INTRODUCTION

Virtual reality (VR) is a technology that provides an experience similar to that in the real world in a virtual environment. Recently, various VR head-mounted displays (HMDs) have been developed to provide three-dimensional visual information, hand and motion tracking for highly immersive interaction, and various senses such as hearing and touch to break down the boundaries between virtual reality and provide higher immersion and new experiences [1]. The development of these technologies has ignited interest in the metaverse, an extended virtual world interconnected to the physical world now widely applied across the economy, society, and culture [2]. Additionally, the importance of extended reality (XR) is increasing as a technology that can create experiences in immersive virtual environments (IVEs) encompassing VR, augmented reality (AR), and mixed reality (MR) and build a metaverse close to reality [3], [4]. Moreover, applied research to increase user immersion in virtual environments is being conducted through various technological approaches or experience analyses. Research has been conducted to increase immersion through spatial

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

sense in hearing along with stereoscopic visual information using HMDs such as MetaQuest 3 and HTC Vive Pro [5], [6], and to perform actions close to reality through more direct interaction with the virtual environment using the body such as gaze and hands [7], [8]. Furthermore, studies on applying haptic systems that accurately feedback physical reactions occurring during interaction processes [9] and on redirected walking for free walking expression in limited spaces and motion platforms using walking in place [10], [11] have been conducted from various perspectives. Additionally, experimental studies are being conducted on various topics and fields to analyze the effects of experiential environments or interactions provided in IVEs on users' sense of presence, experience, and satisfaction [12], [13].

Virtual avatars or agents are used as elements of a user's social presence in a virtual environment and can provide a highly immersive experience by imitating the human shape or similarly expressing actions and facial expressions [14]. A self-avatar is a virtual representation of oneself that is distinguished from other people's avatars because it has different requirements. How one recognizes one's avatar is more important than the others in terms of latency and pose accuracy. In this regard, the impact of user avatar animation fidelity on various tasks was studied, and the effect of animation quality on the sense of embodiment (SoE) was analyzed [15]. In particular, the sense of agency (SoA) refers to the feeling that one is the self that causes or creates an action, and it also means the feeling of becoming an agent that performs the actions of an avatar in VR. This can be utilized as an important element in user immersion and experience by synchronizing real and virtual body movements [16].

Based on previous studies showing that animation fidelity affects SoE and user performance, we propose a rigid-body tracking-based motion estimation method using Final IK [17] for accurate avatar motion calculation while reducing equipment dependency. Additionally, we conducted a comparative analysis with a self-avatar based on a highly accurate optical system for motion capture and performed user evaluations of the realism and experience of the experiential environment provided through a self-avatar in an IVE constructed through the proposed method. Our results were mainly aimed at demonstrating that fast and accurate motion estimation, from the perspective of animation fidelity related to SoAs, can be applied to efficient avatars in IVEs without requiring expensive equipment. The contributions of this study are as follows.

1. We propose a novel virtual avatar design method using rigid-body tracking-based motion estimation for self-avatar generation that minimizes equipment use while maintaining animation accuracy.
2. We systematically studied the technical quantitative evaluation of accuracy with avatars using optical motion capture for various applications in IVEs, and user evaluations of user realism in IVE and the virtual embodiment of virtual avatars.

II. RELATED WORK

The IVE is evolving into a metaverse that blurs the boundaries between the virtual and real by passing through the economy, society, and culture. For this purpose, technology is moving toward applying and utilizing XR, including VR, AR, and MR. Furthermore, research has been conducted to increase user immersion by allowing the body, such as the hands and eyes, to interact more directly with the virtual environment or objects. In addition, research is being conducted on ways to increase immersion by considering actions and thoughts, such as detecting and imitating perceptions or understanding intentions during the interaction process. A study on a gaze-controlled multimedia user interface that controls the interface through eye tracking was conducted [18], and a new interaction technique was proposed in which the user's hand becomes a virtual object by imitating the object itself, unlike in the existing work in which the user's hand grabs a virtual object [19]. Tatzgern et al. [20] confirmed the possibility of creating immersive scenarios by developing an AirRes mask that controls breathing in real-time. Moreover, a new deep learning-based interface that understands the user's behavioral intention and intuitively performs the desired action was proposed [21], and a method using visual information and tactile sensation together through a visuo-haptic interface was suggested [22]. A pseudo-haptic interface was also proposed to increase immersion using electromyography (EMG) sensors to determine the action of grasping a virtual object with the hand and to visually deliver physical feedback to induce an illusion-like phenomenon [9]. Research has been conducted on the methods, equipment, and interfaces through which users participate in IVEs as well as on interactions with other entities within the virtual environment, such as avatars and agents. Kammerlander et al. [23] proposed a method to generate virtual characters of various proportions in a shared collaborative virtual environment using a combination of VR and motion capture and analyzed the impact of virtual avatars on the user. Furthermore, increasing the level of facial and body expressions of virtual avatars can improve the realism and quality of social interaction in VR [24]. VR, which expresses conversation partners as realistic moving avatars, has a positive effect on social interaction with other people [25].

A self-avatar is a virtual representation of one's body from a first-person perspective in a virtual environment. The presence or absence of such an avatar can significantly impact the users' cognitive load. Pan and Steed [26] analyzed the effect of the avatar type on cognitive task performance in a VR environment. They also implemented shared VR for consumer VR systems that can be represented by a self-avatar. This confirms that using a self-avatar has a significant effect on the presence and interactions. Ogawa et al. [27] set up a specific situation in which a person could not pass through a virtual boundary (a wall) and found that using an approach other than sensory feedback was most effective in preventing a realistic full-body self-avatar from passing through the wall. Aseeri and Interrante [28] divided avatar types into No_Avatar (HMD and controllers only), Scanned_Avatar (wearing an HMD), and Real_Avatar (video-see-through) and investigated the influence of visual and nonverbal cues provided by avatar expressions. They reported that the Real_Avatar condition had a higher level of trust and was preferable. Animation accuracy is an important factor when creating self-avatars. Gao et al. [29] analyzed the effect of full-body tracking rather than avatar visibility on accurate avatar behavioral responses. In addition, delays inevitably affect motor performance, agency, body ownership, and simultaneity perception during the process of improving animation accuracy [30]. When creating virtual characters, a library was developed that dramatically reduced the production time along with high-level functions that perform accurate movements [31]. For animation accuracy, IK is utilized to precisely calculate the virtual avatar's motions. Particularly in IVEs, methods are being studied to track the position and direction information of joints through the tracker provided with the VR HMD and to estimate the motion of the self-avatar based on this. Zeng et al. [32] proposed PE-DLS, a novel method for performing full-body motion reconstruction in two steps: pose estimation (PE) and damped least-squares (DLS) optimization based on the IK solver. In addition, a new human upper-body IK algorithm was proposed to track only the head and hand sensors, and directions for applying an animated virtual arm with an IK system were suggested [33]. Using motion capture equipment is a common method for expressing accurate avatar motions. There are various methods of motion capture, such as inertial tracking, which attaches inertial sensors to major joint parts for tracking; optical tracking, which estimates near-real-time animation motions by photographing sensors (markers) with two or more cameras; or methods that estimate motions using computer vision or ultrasound. However, because of technical and environmental issues, such as cost and space, research has been conducted to estimate motion in real-time using an RGB camera [34] or to increase accuracy by utilizing a motion capture system consisting of a set of low-cost consumer-grade sensors [35]. Additionally, research is being conducted on the application of motion capture technology to digitalize and preserve dance, an intangible cultural heritage site [36].

One of the important elements of virtual avatars, particularly self-avatars, is the SoA. This refers to the feeling that 'I' am the one causing the action. The SoA also indicated that the user's actual motions were accurately synchronized with those of the virtual avatar in the IVEs. From this perspective, Kilteni et al. [37] defined SoE as consisting of three subcomponents: sense of self-position, sense of subjectivity, and body ownership. Therefore, research is still being conducted to analyze whether a high level of animation is necessary under various environmental conditions to provide higher immersion to users from the perspective of SoA and user performance [15].

This study proposes a novel method for designing virtual avatars in IVEs that minimizes the use of additional equipment such as trackers by approaching the animation accuracy affecting the SoA using a rigid-body tracking-based motion estimation method. Although optical markers have been utilized for rigid-body tracking, the goal is to focus on a method of estimating motion based on a tracked rigid body rather than rigid-body tracking, to ensure that it can be sufficiently applied to tracking objects using Kinect. In addition, we analyzed the potential of using the virtual avatar expressed through estimated motions, and its impact on the user's realism in an IVE.

III. VIRTUAL AVATAR BASED ON RIGID-BODY TRACKING

Virtual avatars or agents, which are expressed by tracking and calculating a user's motions in an IVE, increase user immersion and help deliver and understand the experiences provided by the virtual environment more effectively and diversely. Animation accuracy may be important when utilizing virtual avatars as elements to increase user immersion. However, there are limitations in that increasing accuracy may require a lot of calculations and additional equipment. We proposed a new motion estimation method based on rigid-body tracking for efficient virtual avatar applications in IVEs. This was inspired by the existing method of generating the motion of a self-avatar by animating through the IK Solver based on the position and direction information tracked from the trackers connected to the joints. Here, rather than directly calculating the movement through the joints, a rigid body was tracked and set as the target. The motion is then inversely estimated through the correspondence between the end effector and the target. Figure 1 shows the overall motion estimation flow for the proposed rigid-body tracking-based virtual avatar application. The aim is to minimize the tracking target and simultaneously present a virtual avatar application method that can provide users with immersive and meaningful experiences in an IVE by generating relatively accurate animation motions.

A. RIGID-BODY TRACKING

A prerequisite for the proposed virtual avatar motion estimation process is the assumption that the movement (position and orientation) of a rigid-body object is accurately tracked. Therefore, accuracy is given the highest priority in rigid-body tracking. In this study, 16 OptiTrack Prime 17 W cameras were used to configure an environment for rigid-body tracking, and the optical markers attached to real objects were tracked to calculate their positions and directions in real-time. The tracking results can be evaluated using Motive software.

Markers were attached to the real objects to be tracked and the markers were tracked using optical equipment. If there are multiple objects according to the placement information, each can be identified using Motive software and the motion can be recognized using degree of freedom (6DoF) data. Figure 2 shows the rigid-body tracking experimental environment. The number of markers attached to real objects is not fixed, and the markers are adjusted such that they can be recognized

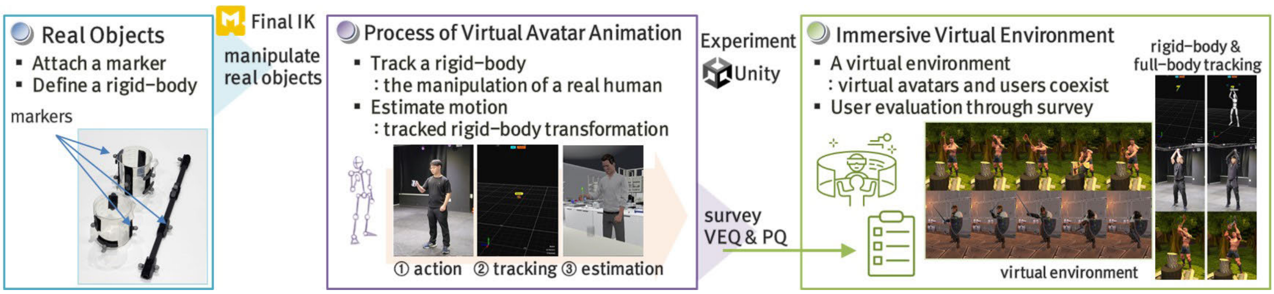

FIGURE 1. Workflow of the proposed rigid-body tracking based virtual avatar's motion estimation.

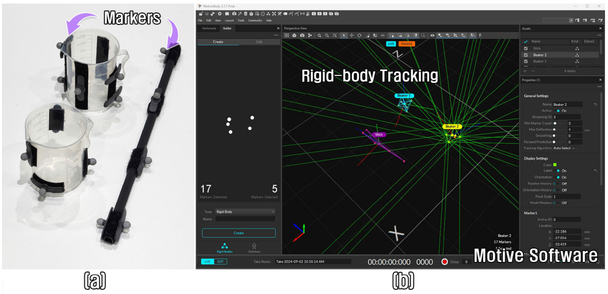

FIGURE 2. Development and experimental environment for rigid-body tracking: (a) The real-world objects with markers attached for rigid-body tracking, (b) The example of rigid-body tracking results in the motive software.

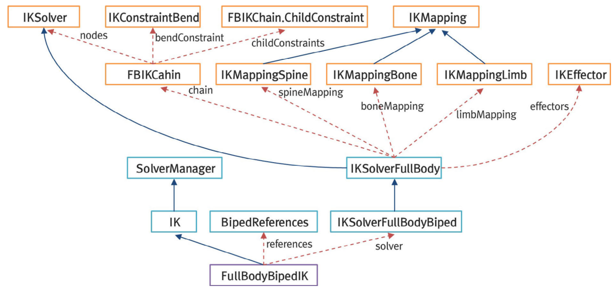

FIGURE 3. Class structure of the FullBodyBipedIK of the Final IK [39].

by the camera within a range that does not interfere with the user directly holding or grabbing the object and acting.

The important point in the proposed method is not to accurately track the rigid-body object, but to verify the motion of the avatar estimated from the movement of the rigid-body. Therefore, although the experimental environment was considered optical, it was concluded that low-cost and lightweight rigid-body tracking methods could be considered together after analyzing the accuracy and realism of the virtual avatar based on the rigid-body tracking proposed through the experiment.

B. MOTION ESTIMATION

This study aims to design a method to inversely estimate the motion of a virtual avatar based on the movement of real-world objects calculated through rigid-body tracking. Therefore, we defined a development environment for virtual avatar motion estimation based on rigid-body tracking that considers user participation in IVE. To estimate the motion of a virtual avatar from the movement of a rigid-body object, the position and direction of the rigid body were set as targets. Subsequently, IK is applied to inversely calculate the movement using the body joints that control the object as end-effectors. For the IK of the virtual avatar, we used the Final IK [17] based on the Unity engine. The Final IK is an IK system supported by the Unity engine that provides functions used in various animation implementations, such as generating natural movements of characters and objects, as well as building an interaction system with objects. It also provides various features such as automatically finding the appropriate motion for the terrain, limiting the range of movement of joints, and creating and managing generated motions through the Final IK as animation files. Therefore, motion estimation is processed by analyzing the user poses of multiple end effectors for each limb from the same input, which produces higher-quality results than the Unity IK system [38].

This study utilizes the Full Body Biped IK and the Look At IK among the functions of the Final IK to estimate virtual avatar motion. The Full Body Biped IK automatically recognizes the joints of the avatar and manages them separately by dividing them into Body, Arm, Leg, etc. This allows for the specification of a target to track each body part. Thus, the most natural motion for controlling the tracked rigid body is estimated by specifying the tracked rigid body as the target and calculating the movement of the joints in reverse. The FullBodyBipedIK's overall structure is divided into the roles of Effector, Chain, and Mapping, and it performs motions based on constraints defined based on the relationship between the spine, bone, and limb (Figure 3). Specifically, the Effector controls the effect that affects generating movement for each body part target, and the Chain divides the influence between joints into detailed steps and controls it. The Mapping is used to adjust the amount of computation and expression to generate the movement. At this time, Pull, Reach, Bend Goal Weight, etc. are adjusted with detailed properties, which are defined as properties of

FIGURE 4. Example of virtual avatar joint movement according to the Full body biped IK component in the final IK and adjusting parameters.

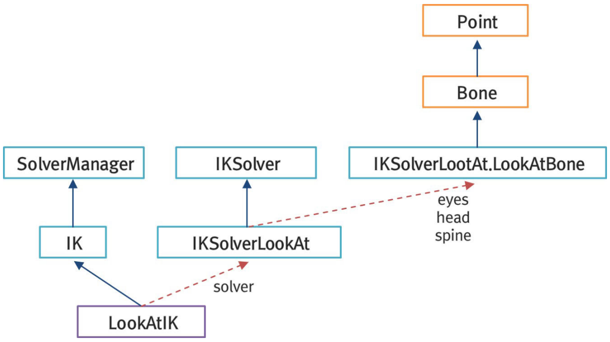

FIGURE 5. Class structure of the LookAtIK of the final IK [39].

the FBIKChain class. Figure 4 shows the properties of the Full Body Biped IK based on a random avatar.

Next, is the Look At IK, a component used to generate natural gaze handling for virtual avatars. There is no rule for gaze, but naturally adjusting the avatar's gaze according to the situation may influence natural movements rather than continuously looking in one direction. Figure 5 shows the class structure of LookAtIK.

The specific implementation process provided by the Final IK is as follows: First, we specify the joint information of the avatar involved in the gaze. The central head was automatically defined, the upper-body joints moved according to the movement of the head, and the eyes could be specified. The extent to which the upper body moves according to the movement of the head can be specified by adjusting the property values. Furthermore, if the avatar is created with eyes separately, a more realistic gaze can be generated by specifying the eyes. In this component, the amount of registered information reflected in the generated movement can be controlled by adjusting the Head, Body, and the Eyes Weight values. It provides precise control of the Spine, Weight Curve and Clamp Weight. And each parameter is defined as a property of the IKSolverLootAk class. Finally, a natural gaze can be generated according to a set of options by specifying the target to be looked at. Figure 6 shows this and the results of natural gaze processing.

The Solver is a general class structure that calculates transformation information for joints in inverse kinematics. In the FullBodyBipedIK class of the Final IK's, it calculates the motion of the whole body using information such as spine, bone, and limb. And in the LootAtIK class, it calculates the motion for overall gaze processing using the movement of the head, eyes, etc.

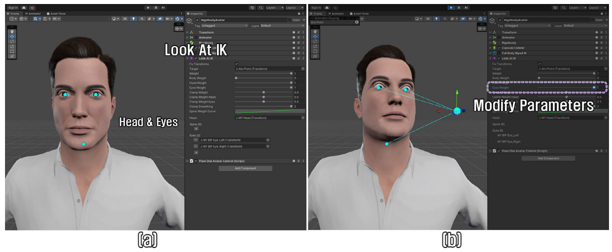

FIGURE 6. The look at IK component in the final IK component and the process of setting the head and eyes of the virtual avatar.

Figure 7 shows an example of the results obtained using the proposed rigid-body tracking-based virtual avatar motion estimation.

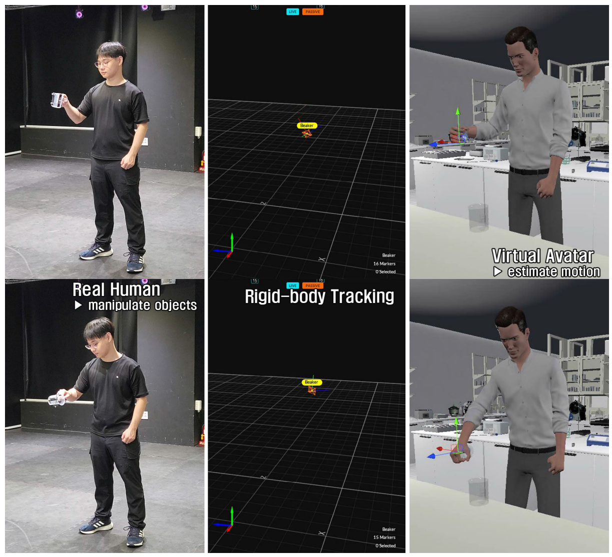

FIGURE 7. Process of generating animation of a virtual avatar through motion estimation based on rigid-body tracking.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The development and experimental environment for the application of virtual avatars based on rigid-body tracking in IVEs were created using the Unity 3D engine (Unity 2022.3.6f1) [40], the Final IK [17], and a plugin for a motion capture studio linkage (OptiTrack Unity Plugin 1.4.0) [41]. In addition, experimental content was created to confirm the user's sense of immersion in the IVE based on the behavior of the virtual avatar and to qualitatively evaluate the virtual embodiment of the virtual avatar's behavior. Rigid-body tracking was performed using Motive software in a studio space with 16 OptiTrack Prime 17 W cameras, and full-body tracking for comparison experiments was performed in the same environment. The PC for the integrated development environment and experiments was equipped with an Intel Core i7-11800H, 16GM RAM, and GeForce RTX 3050Ti Laptop GPU.

A. EXPERIMENT CONFIGURATION

The experimental actions of the virtual avatar were divided into two types: simultaneous two-hand motion and separate two-hand motion, as illustrated in Figure 8. According to previous research on user performance and SoA as specified by the animation accuracy of a self-avatar [15], focusing on the upper body when animating an avatar is recommended. This is because the movements that humans can perform are wider in the upper body than in the lower body. It was also experimentally confirmed that angular errors occurred more frequently in the arms than in the legs under all conditions. Therefore, this study designed two actions based on previous research that high-quality animation of the upper body (especially the arms) is necessary. In addition, we appropriately defined the situation for each action and provided meaning to the motions. Through this, we aim to consider the user performing actions in an IVE and the immersion and experience of other users interacting with the avatar.

The first action is a simultaneous two-hand motion. In this action, the character holds an axe with both hands and chops wood. The second action is a separate two-hand motion, holding a shield in one hand and a sword in the other, where the background acts like a knight character in a game. The reason for configuring actions and situations in this manner is to establish an environment in which users can consider their subjective senses as self-avatars from various angles in various cases. In full-body tracking, the motion of the actor wearing the suit was tracked; however, the movement of the object was not tracked simultaneously. The actor holding the object is only used for natural action. Therefore, a situation in which the object (axe, sword, or shield) is dislocated from the hand is considered for full-body tracking-based virtual avatar motion.

B. TECHNICAL COMPARISON EXPERIMENT

To accurately estimate the motion of a virtual avatar based on rigid-body tracking, this study generated an animation that moves according to the movement of a rigid body using the Final IK. Both the basic one-hand motion in Figure 7 and the two actions proposed for the experiment were generated based on the Full Body Biped IK and the Look At IK in the Final IK. However, the component property values for each action must be defined differently, depending on the operating range and situation of the action. Table 1 summarizes the values defined for each action of the key parameters of the Full Body Biped IK that are necessary to express an accurate animation. The parameter Pull is the weight (involvement) that pulls the parent chain (shoulder, elbow), and Reach is the value that pulls the first bone of the limb (e.g., shoulder) closer to the last bone (e.g., hand). The parameter Bend Goal Weight is the value that bends the limb in the Bend Goal direction. The two actions proposed in this study had a high involvement of the elbow. Hence, they were used as variables to control the elbow involvement

TABLE 1. Key parameters of the Final IK set for virtual avatar motion generation.

		One-hand motion	Simultaneous two-hand motion	Separate two-hand motion
Full Body Biped IK	Pull (L/R)	1	1	1
	Reach (R)	0.2	0.1	0.1
	Reach (L)	0.1	0.1	0.1
	Bend Goal Weight (R)	0	0.65	0.5
	Bend Goal Weight (L)	0	0.65	0.5
Look At IK	Head Weight	1	1	1
	Eyes Weight	1	1	1

according to the movement of the rigid body. The parameters were adjusted according to the action characteristics to enable natural motion estimation. In addition, by setting the Head and Eyes Weights among the Look At IK parameters, the head and eyes could follow naturally according to the movement of the rigid body.

The following describes an action comparison experiment with an avatar based on full-body tracking using optical equipment. The motions that a user can perform may vary when performing a specific action in each case. In other words, it was difficult to obtain a single correct motion. This study aimed to intuitively verify the difference by comparing the action-focused key frames rather than comparing them with the ground truth between animation expressed through full-body tracking, which can be expressed relatively naturally, and the animation expressed through estimated motion by the proposed rigid-body tracking in the same situation. Figure 9 illustrates this by showing frames at the beginning, end, and key situations of each of the two defined actions. First, gaze is not an important subject for comparison because it is expressed randomly, without any set rules. No awkward or inaccurate parts (e.g., the arm digging into the skin) were detected in the motions estimated using the proposed method. Additionally, the animation connection process proceeded smoothly without any unexpected behavior. However, because the proposed rigid-body tracking method was set to match the avatar's gaze with the movement of the rigid body, the results showed differences, particularly for the head.

Finally, the number of controllers, trackers, sensors, and markers used for accurate animation were compared. We compared and analyzed the conditions for motion estimation with an existing study [15] that compared and analyzed the user experience, from the perspective of the animation accuracy of the self-avatar (Table 2). First, the method of generating animation through the Unity IK and the Final IK using a consumer VR tracker in existing studies uses five pieces of equipment, including an HMD that tracks the head, controllers for both hands, the pelvis, and trackers

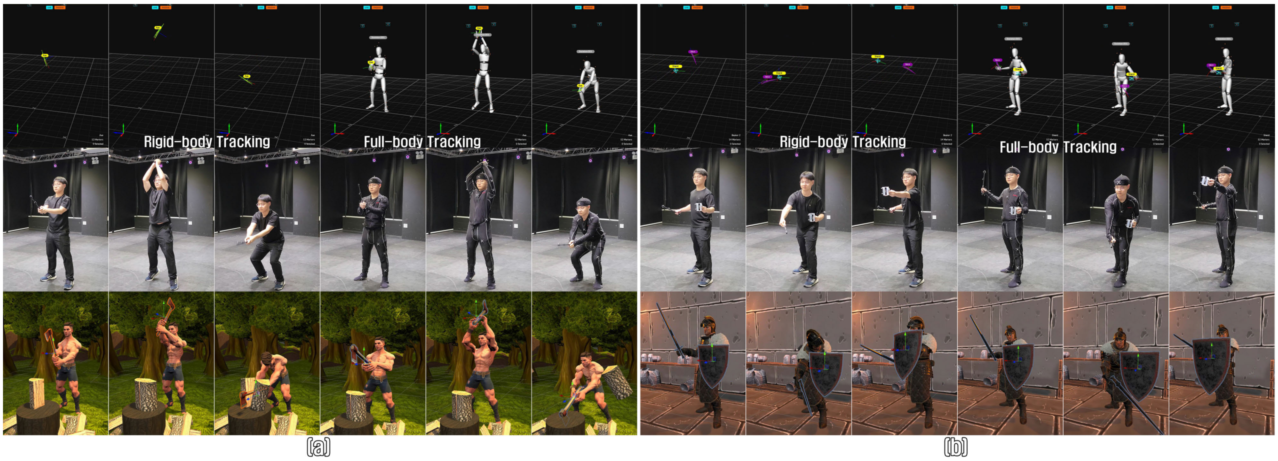

FIGURE 8. Motion process of virtual avatar estimated through full-body tracking and rigid-body tracking: (a) simultaneous two-hand motion (axe), (b) separate two-hand motion (knight: sword and shield).

FIGURE 9. Example of key frame comparison for virtual avatar motion animation analysis.

attached to each foot. When using the inertial motion capture equipment, joint movements were precisely calculated using 17 sensors, including those on the head. For the full-body tracking designed for comparison, animation was estimated by wearing a suit with 37 attached markers. However, the proposed rigid-body tracking method can be used to estimate animation motion using 4-11 markers. By default, three markers per tracking object were set as one set. We increased the number of markers to improve the tracking accuracy, but the most ideal results were achieved with four to five markers. In other words, it was important to find the optimal arrangement and number of markers that could specify the object movement rather than numerous markers. When an experiment was conducted by attaching 3 markers, which is the minimum constraint for the number of markers, it caused inaccurate tracking results as in Figure 10(a). This is because some markers are missed for reasons such as the markers going to the hand holding the object, resulting in inaccurate positions being calculated and problems with broken frames in continuous motions. On the other hand, when a large number of markers (e.g. 8) are attached as in Figure 10(b), this also frequently results in a phenomenon where the number of markers actually tracked decreases during the process of holding and acting on the object. As a result, it was confirmed

TABLE 2. Comparison of experimental conditions with previous studies.

	Type	Number of tracking devices
Previous [15]	Consumer-level VR device	5 + 1 (head: VR HMD)
	Xsens IMU sensors	16 + 1 (head sensor)
Our	Full-body	37 (include head marker, controlling 14 joints)
	Rigid-body	4∼11 (basis: 3 pieces 1set, One-hand motion: 4 pieces, Simultaneous two-hand motion: 5 pieces, Separate two-hand motion: 11 pieces (sword: 5 pieces, shield: 6 pieces)

that unnecessarily increasing the number of markers only causes inconvenience in object manipulation actions and does not help in accurate object tracking.

Another important point is that optical equipment is not required for rigid-body tracking. If the experimental environment used for rigid-body tracking can accurately extract stereoscopic data through equipment such as Kinect,

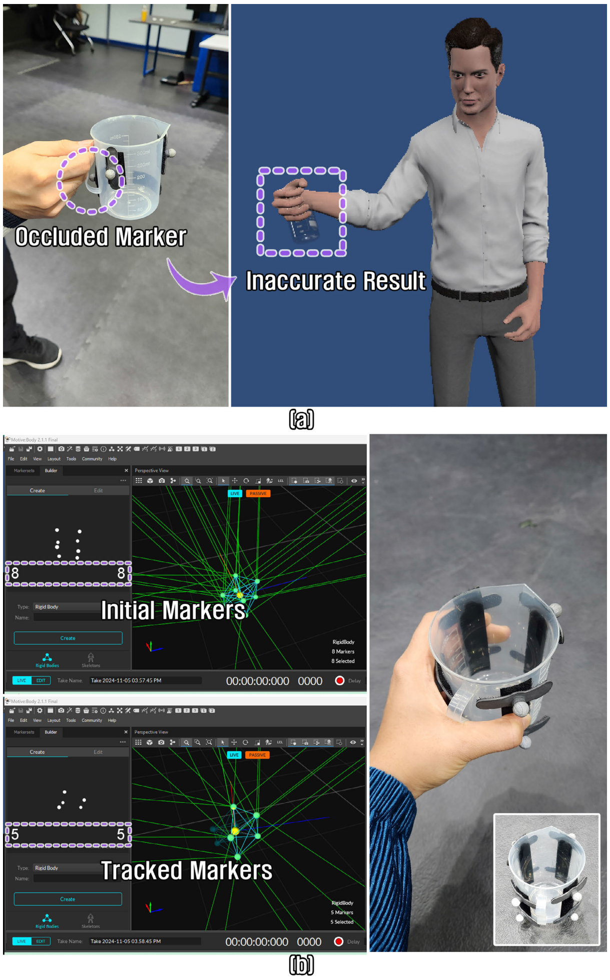

FIGURE 10. Problems in the rigid-body tracking process according to the number of markers: (a) setting the minimum number of markers, (b) results of attaching a large number of markers.

it can be expected that efficient animation construction will be possible without markers.

C. SURVEY EVALUATION

Finally, we qualitatively analyzed the user experience in IVEs using a virtual avatar based on rigid-body tracking. Table 3 shows the agency item that was modified and organized for the questionnaire of this study among the three factors of agency of embodiment used in the study of agency of agent [15] in the virtual embodiment questionnaire (VEQ) [42]. This study aims to consider applications from the perspective of a third-party user who interacts and communicates with a virtual avatar acting in a virtual environment rather than a self-avatar hosted by the user. Therefore, the survey was conducted from the perspective of the person facing the avatar, rather than recognizing the avatar as oneself.

The experiment consisted of 12 people (male: 9, female: 3) between the ages of 23 and 32 years, and the participants experienced the proposed avatar based on rigid-body tracking and the avatar expressed through full-body tracking. The participants experienced the completed motions for each of the two proposed actions and recorded their values according to the items in Table 3. The results of the participant's responses are presented in Table 4.

The survey results showed that satisfaction was relatively high for full-body tracking, in which the motions of the virtual avatar were presented realistically and naturally to the user. Full-body tracking tracks the movements of all the joints of the human body and reflects them in the avatar's motion. In contrast, rigid-body tracking calculates motion based on the movement of the rigid body; therefore, it has the limitation of slightly lowering the smoothness of the overall movement. In addition, the estimated motions that

TABLE 3. Questions for comparative analysis of actions of virtual avatar embodiment.

Agency	Question
1	The movements of the virtual body felt like real movements.
2	I felt the virtual arms moved as real arms.
3	I felt the virtual elbows were in the natural position of elbows.
4	I felt the virtual hands were in the natural position of hands.
5	I felt the virtual legs moved like natural legs.
6	I found it natural to control the virtual body pose to complete the exercises.

TABLE 4. Comparative analysis of actions of virtual avatar embodiment based on full-body tracking and rigid-body tracking (simultaneous two-hand motion: axe, separate two-hand motion: knight).

Two Hand	Simultaneous motion Mean (SD)		Separate motion	
	Rigid-body	Full-body	Rigid-body	Full-body
1	4.923(1.639)	5.461(1.278)	4.769(1.576)	5.923(1.269)
2	4.846(1.610)	5.769(1.250)	5.077(1.774)	5.846(1.167)
3	4.615(1.595)	5.846(1.167)	5.462(1.737)	5.923(1.542)
4	5.384(1.689)	5.462(1.447)	5.539(1.692)	5.384(1.273)
5	5.616(1.389)	5.538(1.278)	5.077(1.542)	5.384(1.689)
6	4.923(1.542)	5.615(0.923)	4.615(1.643)	5.769(1.367)
Total	5.051(1.448)	5.615(1.087)	5.090(1.514)	5.705(1.188)
Pairwise Comparison (Rigid-body vs Full-body)				
1	$F(1,22)= 0.805$, $p=0.378$		$F(1,22)= 3.902$, $p=0.060$	
2	$F(1,22)= 2.461$, $p=0.130$		$F(1,22)= 1.575$, $p=0.222$	
3	$F(1,22)= 4.655$, $p<0.05$*		$F(1,22)= 0.474$, $p=0.498$	
4	$F(1,22)= 0.014$, $p=0.906$		$F(1,22)= 0.927$, $p=0.345$	
5	$F(1,22)= 0.020$, $p=0.889$		$F(1,22)= 0.217$, $p=0.645$	
6	$F(1,22)= 1.780$, $p=0.195$		$F(1,22)= 3.497$, $p=0.074$	
Total	$F(1,22)= 1.179$, $p=0.288$		$F(1,22)= 1.227$, $p=0.279$	

reflect the physical characteristics of the virtual character (such as broad shoulders or long arms) are less flexible than those of full-body tracking. For this reason, the character used in the simultaneous two-hand motion (axe) had broad shoulders, resulting in the elbows turning inward during the chopping of firewood. This negatively affected the survey results. In contrast, full-body tracking focuses on the hands and arms while chopping firewood; therefore, the movement of the lower body is expressed relatively unnaturally, whereas rigid-body tracking expresses movement using joint transformation information calculated from the upper body to the lower body. Therefore, there were differences in the overall consistency of the expressions for all joints. This is because the actors do not chop wood but only act, and because the actors who participated in full-body tracking are ordinary people who do not specialize in acting. This phenomenon has also been observed in other studies. Because the rigid-body tracking-based method estimates motions centered on an object, all motions are estimated based on the movement of the object. This has led to the view that the weight of the axe in simultaneous two-hand motions is conveyed more realistically than in full-body tracking. Finally, statistical significance was confirmed using a one-way ANOVA. One-way ANOVA is a type of statistical test method that compares the variance of the means of groups. It considers one independent variable or factor. This study considers the hypothesis that rigid-body and full-body tracking have no difference in the satisfaction of virtual embodiment as an independent factor. In other words, the null hypothesis is that there is no difference between the rigid-body and full-body tracking survey groups. Therefore, the alternative hypothesis means that there is a significant difference. At this time, the smaller the p-value, the higher the possibility of rejecting the null hypothesis. If the significance level is 5%, that is, if the p-value is less than 0.05, the null hypothesis of this study is rejected. However, in this study, both actions showed no significant differences in the comprehensive virtual embodiment analysis ($p>0.05$). Based on this survey experiment process, in an environment where the action of a virtual avatar is not generated by a professional actor, it is possible to utilize natural and efficient motion estimation by considering varying the Final IK parameters to supplement the main joints.

Next, we compared and analyzed the realism of users in a space where they participated with a virtual avatar that acted through motions estimated in the IVE. The questionnaire was created by modifying the realism section of the presence questionnaire (PQ) [43] to suit the purpose of this study. The questionnaire recorded the values on a 7-point scale. Table 5 contains questions used to analyze the realism experienced by users through virtual avatars in IVE. The survey participants were configured similarly as in the virtual embodiment survey experiment.

Table 6 presents a comparative analysis of the survey results. The survey results showed that realism had a narrower gap than the naturalness of the avatar's motions.

TABLE 5. Questions for analyzing the realism of users in IVE with virtual avatars.

Realism	Question
1	How natural did the interactions with the environment through the avatar's actions seem?
2	How much did the visual aspects of the avatar's actions involve you?
3	How compelling was the sense of movement to manipulate objects through avatar in virtual environments?
4	How much did avatar's behavioral experience in the virtual environment seem consistent with your real world experiences?
5	How compelling was the sense of moving around through avatar inside the virtual environment?
6	How involved were you in the actions of the virtual avatar?

TABLE 6. Comparative analysis of realism through virtual avatars based on full-body tracking and rigid-body tracking (simultaneous two-hand motion: axe, separate two-hand motion: knight).

Two Hand	Simultaneous motion Mean (SD)		Separate motion	
	Rigid-body	Full-body	Rigid-body	Full-body
1	5.077(1.639)	5.231(1.476)	5.077(1.730)	5.462(1.216)
2	5.000(1.617)	5.308(1.323)	5.231(1.804)	5.385(1.595)
3	5.231(1.576)	5.846(0.948)	5.154(1.791)	5.615(1.389)
4	4.846(1.460)	5.385(1.003)	4.846(1.747)	5.538(1.278)
5	5.308(1.896)	5.462(1.646)	5.154(1.657)	5.538(1.646)
6	5.231(1.476)	5.308(1.202)	5.385(1.546)	5.692(1.588)
Total	5.115(1.545)	5.423(1.087)	5.141(1.630)	5.538(1.237)
Pairwise Comparison (Rigid-body vs Full-body)				
1	$F(1,22)= 0.058$, $p=0.811$		$F(1,22)= 0.397$, $p=0.535$	
2	$F(1,22)= 0.260$, $p=0.615$		$F(1,22)= 0.049$, $p=0.827$	
3	$F(1,22)= 1.343$, $p=0.258$		$F(1,22)= 0.498$, $p=0.487$	
4	$F(1,22)= 1.109$, $p=0.303$		$F(1,22)= 1.227$, $p=0.279$	
5	$F(1,22)= 0.045$, $p=0.834$		$F(1,22)= 0.325$, $p=0.574$	
6	$F(1,22)= 0.020$, $p=0.890$		$F(1,22)= 0.231$, $p=0.635$	
Total	$F(1,22)= 0.318$, $p=0.578$		$F(1,22)= 0.453$, $p=0.508$	

Alternatively, although the avatar's motions were relatively unnatural, they had a slight effect on the user's immersion in the virtual environment. We verified that the proposed rigid-body tracking-based virtual avatar did not interfere with user realism. It was also confirmed through a one-way ANOVA that there was no significant difference in any aspect of realism for either action ($p>0.05$). At this time, the null hypothesis is that there is no difference in realism between the two groups of rigid-body tracking and full-body tracking.

V. LIMITATION AND DISCUSSION

The proposed motion estimation method for a virtual avatar is based on rigid-body tracking. In this process, optical motion capture equipment was used. However, while optical motion capture allows accurate tracking, it has clear limitations, such as high cost and large space requirements. This method has the same constraints, even though it does not involve wearing a full-body suit. However, the proposed method aims to analyze the advantages and features of calculating movement inversely based on a tracked rigid body. Although we utilized optical motion capture equipment for the comparison experiments with full-body tracking using a full-body suit along with accurate rigid-body tracking, there is no prerequisite for the proposed method to use optical equipment. Accordingly, if the proposed method is applied using low-cost camera-based equipment, such as Kinect, it will be possible to increase the efficiency and overcome current limitations.

A quantitative analysis of the animation accuracy of a virtual avatar that moves based on movement calculated based on rigid-body tracking has not yet been conducted. As it is difficult to design an objective index to judge the accuracy of actions, this study evaluates the feasibility of the proposed method through a qualitative comparative analysis of each action with an avatar moving through full-body tracking. Another limitation is that we conducted only a qualitative experiment comparing realism and virtual embodiment through a questionnaire from the perspective of a user experiencing a virtual avatar acting with estimated motions. In this regard, we plan to design a criterion that can apply the quantitative experimental process of existing studies related to animation fidelity to the method proposed in this study and conduct an experiment to enable quantitative analysis.

There are clear differences and problems when expressing motion using objects in a virtual avatar using rigid-body and full-body tracking. For instance, the proposed rigid-body tracking-based method calculates the rigid body and moves the avatar when movement occurs in the object. For this reason, there is a problem in which an unnatural situation occurs where the joint suddenly moves because there is no preliminary motion before the movement begins. However, virtual avatars using full-body tracking only estimate human joints; therefore, additional work is required to match the joints with the moving object.

One thing to consider in terms of motion is that the actors who conducted the full-body tracking experiments were ordinary people who did not specialize in acting. It cannot be ruled out that the experimental results would have changed if the motions had been performed by professional actors with acting experience or experience in using motion capture to produce animations. Experiments that consider these factors should be conducted to verify the usability of the proposed rigid-body tracking-based virtual avatar. In addition, the current movement directly applies to the joint transformation information calculated using the Final IK. Because the user's additional correction work is not considered, unnatural postures may appear depending on the situation compared with full-body tracking (Figure 11). In the case of a simultaneous two-hand motion, the elbows may be strongly drawn inward because they focus on grasping the object. Separate two-hand motions can sometimes show excessive bending due to the head setting moving along the object and have the limitation that there is no change in movement without additional settings for the lower body. Therefore, it will be necessary to diversify the parameters that can subdivide and control joints, as well as generalize the initial settings that consider the characteristics of the character.

FIGURE 11. Example of movement requiring compensation in a virtual avatar based on rigid-body tracking.

The proposed rigid-body tracking based virtual avatar aims to enable highly immersive interaction with users in IVEs by expressing accurate motions with simple inputs that minimize the use of equipment. This can be utilized in entertainment fields such as virtual YouTubers who utilize characters that express human actions in place of others for broadcasting. In the field of sports for all such as Pilates, which uses equipment to exercise, it can be helpful as an exercise assistance function by estimating and judging postures based on the movement of the equipment. In addition, it can be applied as a safety accident prevention technology in industrial fields such as construction and manufacturing. For example, by estimating the motion of virtual human movements based on construction materials, assembly parts, etc., and predicting physical safety accidents that may occur when actual workers work on site, it is possible to plan safe and efficient routes and work methods. In this application process, if an environment is presented where only the movements of necessary objects can be tracked without investing in large physical spaces or expensive equipment, efficient motion estimation is expected to be possible based on this. However, elaborate motions such as hand gestures may be required in this process. Therefore, we aim to overcome the limitation of currently being implemented only as a fixed hand gesture that grabs an object according to the size of the virtual object corresponding to the rigid-body, and to implement it so that the motion of the hand gesture can also be estimated naturally according to the object.

VI. CONCLUSION

This study proposes a novel method for generating animations of virtual avatars coexisting with users in IVE with fewer conditions, costs, and inputs. This method estimates the corresponding motion according to the movement of an object based on rigid-body tracking. Thus far, this study tracks the movement of a rigid body for object-centered virtual avatar motion estimation and designs a process to calculate the movement using the transformation information of the tracked object as a target. In this process, the Final IK is used to calculate the IK for the virtual avatar animation. Yun et al. [15] demonstrated that accurate positioning is possible even with the Final IK-based solution that uses a few sensors. Consequently, this study also developed a motion estimation function for a virtual avatar using the Final IK. The properties and parameters of the Final IK used to generate the proposed virtual avatar motions were analyzed and organized. In addition, the technical conditions and conditions for animation generation were analyzed and compared with those of the existing study. Based on this, we analyzed the behavioral elements and reliability of virtual avatar embodiments and conducted a survey experiment to determine their impact on the realism of virtual avatars coexisting with users in IVE. For the experiment, two virtual avatar actions were defined: simultaneous and separate two-hand motions. We then conducted a comparative survey experiment between a virtual avatar designed elaborately using full-body tracking and a virtual avatar based on the proposed rigid-body tracking. The virtual avatar embodiment experiment utilized a purpose-oriented questionnaire based on the VEQ. In addition, realism was assessed using survey questions modified to suit the purpose of this study based on the PQ. The results of the comparative survey showed that the proposed rigid-body tracking-based avatar had a lower overall potential for virtual avatar motion estimation than full-body tracking; however, there was no significant difference. Additionally, the usability of the rigid-body tracking-based avatar was confirmed to be sufficient. Furthermore, it was found that the realism that virtual avatars provide users in IVE can sufficiently coexist at a level that does not interfere with the user's immersion.

In order to verify the accuracy and usability of the proposed rigid-body based motion estimation, this study conducted a frame-by-frame motion comparison with the full-body tracking method and a user evaluation through survey experiments. This is based on the motion estimation results through full-body tracking due to the limitation of difficulty in establishing a clear standard (ground truth) regarding accuracy. However, this is also not a comparative experiment based on quantitative data with full-body tracking. Therefore, we plan to objectively verify the performance of the proposed method through quantitative experimental results in the future. To this end, we need a standard that can analyze the accuracy through comparison of joint change values in the estimated motion or transform values between joints in the hierarchical structure. Therefore, we will define a standard that can verify the accuracy of the estimated motion and study a comparative experiment. In addition, we currently conducted a motion estimation experiment with one-hand motion, simultaneous two-hand motion and separate two-hand motion. We plan to suggest specific application directions by considering various motions in terms of usability [44]. Finally, we plan to study whether it is possible to build a real low-cost system by applying the proposed method through rigid-body information tracked through computer vision using Kinect, cameras, etc. instead of the motion capture equipment, which is the experimental environment of this study. Based on these processes, we plan to develop a technology that can be applied to various metaverse environments by comprehensively considering the voice and facial motion of virtual avatars.

REFERENCES

[1] Y. Cho, M. Park, and J. Kim, "XAVE: Cross-platform based asymmetric virtual environment for immersive content," *IEEE Access*, vol. 11, pp. 71890–71904, 2023.
[2] S. Tim, "Foundational principles & technologies for the metaverse," in *Proc. ACM SIGGRAPH Talks*, 2019, pp. 1–10.
[3] N. Xi, J. Chen, F. Gama, M. Riar, and J. Hamari, "The challenges of entering the metaverse: An experiment on the effect of extended reality on workload," *Inf. Syst. Frontiers*, vol. 25, no. 2, pp. 659–680, Feb. 2022.
[4] L. Oppermann, F. Buchholz, and Y. Uzun, "Industrial metaverse: Supporting remote maintenance with avatars and digital twins in collaborative XR environments," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, Apr. 2023, pp. 1–12.
[5] W. Fitzpatrick, M. Wickert, and S. Semwal, "3D sound imaging with head tracking," in *Proc. IEEE Digit. Signal Process. Signal Process. Educ. Meeting (DSP/SPE)*, Aug. 2013, pp. 216–221.
[6] D. Hong, T.-H. Lee, Y. Joo, and W.-C. Park, "Real-time sound propagation hardware accelerator for immersive virtual reality 3D audio," in *Proc. 21st ACM SIGGRAPH Symp. Interact. 3D Graph. Games*, Feb. 2017, pp. 1–2.
[7] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves, "Gaze-supported 3D object manipulation in virtual reality," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, p. 21.
[8] Y. Cho, J. Kang, J. Jeon, J. Park, M. Kim, and J. Kim, "X-person asymmetric interaction in virtual and augmented realities," *Comput. Animation Virtual Worlds*, vol. 32, no. 5, Sep. 2021, Art. no. e1985.
[9] M. Kim, J. Kim, K. Jeong, and C. Kim, "Grasping VR: Presence of pseudo-haptic interface based portable hand grip system in immersive virtual reality," *Int. J. Hum.-Comput. Interact.*, vol. 36, no. 7, pp. 685–698, Apr. 2020.
[10] N. L. Williams, A. Bera, and D. Manocha, "Redirected walking in static and dynamic scenes using visibility polygons," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 11, pp. 4267–4277, Nov. 2021.
[11] J. Lee, K. Jeong, and J. Kim, "MAVE: Maze-based immersive virtual environment for new presence and experience," *Comput. Animation Virtual Worlds*, vol. 28, nos. 3–4, May 2017, Art. no. e1756.
[12] J. Lee, M. Kim, and J. Kim, "RoleVR: Multi-experience in immersive virtual reality between co-located HMD and non-HMD users," *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 979–1005, Jan. 2020.
[13] D. Paes, J. Irizarry, M. Billinghurst, and D. Pujoni, "Investigating the relationship between three-dimensional perception and presence in virtual reality-reconstructed architecture," *Appl. Ergonom.*, vol. 109, May 2023, Art. no. 103953.
[14] N. Pelechano, C. Stocker, J. Allbeck, and N. Badler, "Being a part of the crowd: Towards validating VR crowds using presence," in *Proc. 7th Int. Joint Conf. Autonomous*, vol. 1, 2008, pp. 136–142.

[15] H. Yun, J. L. Ponton, C. Andujar, and N. Pelechano, "Animation fidelity in self-avatars: Impact on user performance and sense of agency," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Los Alamitos, CA, USA, Mar. 2023, pp. 286–296.

[16] N. David, A. Newen, and K. Vogeley, "The 'sense of agency' and its underlying cognitive and neural mechanisms," *Consciousness Cognition*, vol. 17, no. 2, pp. 523–534, 2008.

[17] RootMotion. (2023). *Final Ik*. [Online]. Available: https://assetstore.unity.com/packages/tools/animation/final-ik-14290

[18] N. Sidorakis, G. A. Koulieris, and K. Mania, "Binocular eye-tracking for the control of a 3D immersive multimedia user interface," in *Proc. IEEE 1st Workshop Everyday Virtual Reality (WEVR)*, Mar. 2015, pp. 15–18.

[19] S. Pei, A. Chen, J. Lee, and Y. Zhang, "Hand interfaces: Using hands to imitate objects in AR/VR for expressive interactions," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–16.

[20] M. Tatzgern, M. Domhardt, M. Wolf, M. Cenger, G. Emsenhuber, R. Dinic, N. Gerner, and A. Hartl, "AirRes mask: A precise and robust virtual reality breathing interface utilizing breathing resistance as output modality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–14.

[21] T. Kang, M. Chae, E. Seo, M. Kim, and J. Kim, "DeepHandsVR: Hand interface using deep learning in immersive virtual reality," *Electronics*, vol. 9, no. 11, p. 1863, Nov. 2020.

[22] S. Yun, S. Park, B. Park, S. Ryu, S. M. Jeong, and K.-U. Kyung, "A soft and transparent visuo-haptic interface pursuing wearable devices," *IEEE Trans. Ind. Electron.*, vol. 67, no. 1, pp. 717–724, Jan. 2020.

[23] R. K. Kammerlander, A. Pereira, and S. Alexanderson, "Using virtual reality to support acting in motion capture with differently scaled characters," in *Proc. IEEE Virtual Reality 3D User Interface (VR)*, Mar. 2021, pp. 402–410.

[24] A. D. Fraser, I. Branson, R. C. Hollett, C. P. Speelman, and S. L. Rogers, "Expressiveness of real-time motion captured avatars influences perceived animation realism and perceived quality of social interaction in virtual reality," *Frontiers Virtual Reality*, vol. 3, pp. 1–20, Dec. 2022.

[25] S. L. Rogers, R. Broadbent, J. Brown, A. Fraser, and C. P. Speelman, "Realistic motion avatars are the future for social interaction in virtual reality," *Frontiers Virtual Reality*, vol. 2, pp. 1–21, Jan. 2022.

[26] Y. Pan and A. Steed, "Avatar type affects performance of cognitive tasks in virtual reality," in *Proc. 25th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2019, pp. 1–4.

[27] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose, "Do you feel like passing through walls?: Effect of self-avatar appearance on facilitating realistic behavior in virtual environments," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–14.

[28] S. Aseeri and V. Interrante, "The influence of avatar representation on interpersonal communication in virtual social environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 5, pp. 2608–2617, May 2021.

[29] B. Gao, J. Lee, H. Tu, W. Seong, and H. Kim, "The effects of avatar visibility on behavioral response with or without mirror-visual feedback in virtual environments," in *Proc. IEEE Conf. Virtual Reality 3D User Interface Abstr. Workshops (VRW)*, Mar. 2020, pp. 780–781.

[30] T. Waltemate, I. Senna, F. Hulsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch, "The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, New York, NY, USA, Nov. 2016, pp. 27–35.

[31] R. Oliva, A. Beacco, X. Navarro, and M. Slater, "QuickVR: A standard library for virtual embodiment in unity," *Frontiers Virtual Reality*, vol. 3, pp. 1–27, Sep. 2022.

[32] Q. Zeng, G. Zheng, and Q. Liu, "PE-DLS: A novel method for performing real-time full-body motion reconstruction in VR based on vive trackers," *Virtual Reality*, vol. 26, no. 4, pp. 1391–1407, Dec. 2022.

[33] M. Parger, J. H. Mueller, D. Schmalstieg, and M. Steinberger, "Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality," in *Proc. 24th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2018, pp. 1–10.

[34] D. Mehta, O. Sotnychenko, F. Mueller, W. Xu, M. Elgharib, P. Fua, H.-P. Seidel, H. Rhodin, G. Pons-Moll, and C. Theobalt, "XNect: Real-time multi-person 3D motion capture with a single RGB camera," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 1–23, Aug. 2020.

[35] A. Chatzitofis, G. Albanis, N. Zioulis, and S. Thermos, "A low-cost & realtime motion capture system," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21421–21426.

[36] L. Zhang, "Digital protection of dance of intangible cultural heritage by motion capture technology," in *Proc. Int. Conf. Cogn. Based Inf. Process. Appl.*, 2022, pp. 429–436.

[37] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence*, vol. 21, no. 4, pp. 373–387, Nov. 2012.

[38] L. Wagnerberger, D. Runde, M. T. Lafci, D. Przewozny, S. Bosse, and P. Chojecki, "Inverse kinematics for full-body self representation in VR-based cognitive rehabilitation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Nov. 2021, pp. 123–129.

[39] (2024). *Rootmotion Class*. [Online]. Available: http://www.root-motion.com/finalikdox/html/annotated.html

[40] (2024). *Unity Engine*. [Online]. Available: https://unity3d.com/

[41] (2022). *Optitrack*. [Online]. Available: https://optitrack.com/

[42] D. Roth and M. E. Latoschik, "Construction of the virtual embodiment questionnaire (VEQ)," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 12, pp. 3546–3556, Dec. 2020.

[43] B. G. Witmer, C. J. Jerome, and M. J. Singer, "The factor structure of the presence questionnaire," *Presence*, vol. 14, no. 3, pp. 298–312, Jun. 2005.

[44] S. Liu, Y. Li, and G. Hua, "Human pose estimation in video via structured space learning and halfway temporal evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 7, pp. 2029–2038, Jul. 2019.

MYEONGSEOK PARK received the B.E. degree from the Division of Computer Engineering, Hansung University, South Korea, in 2022, and the M.S. degree from the Department of Computer Engineering, Hansung University, in 2024. His research interests include virtual reality, augmented reality, HCI, and computer graphics applications.

JIWON LEE received the B.E. degree from the Department of Software, Catholic University of Pusan, South Korea, in 2018, and the M.S. degree from the Graduate School of Culture Technology, KAIST, South Korea, in 2021. She is currently pursuing the Ph.D. degree with the Department of Information and Computer Engineering, Hansung University, South Korea. Her research interests include virtual reality, augmented reality, HCI, and computer graphics application.

HYEONYONG YANG is currently a bachelor's and master's program student of Division of Computer Engineering, Hansung University, South Korea. His research interests include virtual reality, augmented reality, HCI, and computer graphics application.

JINMO KIM received the B.E. degree in multimedia engineering and the M.E. and Ph.D. degrees in multimedia from Dongguk University, in 2006, 2008, and 2012, respectively. He was an Assistant Professor with the Department of Software, Catholic University of Pusan, South Korea, from March 2014 to February 2019. He is currently an Associate Professor with the Division of Computer Engineering, Hansung University, South Korea. His research interests include virtual reality, augmented reality, computer graphics, metaverse, game engineering, and related applications.