https://doi.org/10.7840/kics.2025.50.1.116

A Human-Centric Interactive Avatar Creation Framework Using Motion Capture and Face Expression Detection in Metaverse

Ahmad Zainudin◆, Esmot Ara Tuli*, Dong-Seong Kim**, Jae-Min Lee°

ABSTRACT

The metaverse is a next-generation internet that merges the physical and virtual worlds, offering users immersive interaction experiences. Avatars are essential components in the metaverse, as they are human digital representations that interact and communicate with virtual objects. However, current avatar development has limited movement capabilities (standing, walking, running, and jumping), and the features of facial expressions are limited. This paper presents an interactive avatar generation framework by leveraging motion capture and facial expression to enable human-like interactive avatars in metaverse ecosystems. The proposed framework utilizes the camera to capture the human skeleton data and provide more detailed human movements with facial expressions. This framework achieved high accuracy and low latency for detailed avatar movement synchronization in physical and virtual worlds.

Key Words : Interactive avatar, motion capture, face expression detection, metaverse

Ⅰ. Introduction

The metaverse is a next-generation internet that is empowered by emerging technologies such as the Internet of Things (IoT), digital twins (DT), artificial intelligence (AI), blockchain, Web 3.0, and augmented/ virtual reality (VR/ AR) to facilitate a dynamic interaction between the physical and virtual worlds[1]. The metaverse presents a virtual environment that mimics real-world activities in the physical world and enables various interaction capabilities and diverse engagements[2]. Avatars are essential components of the metaverse, providing users with a sense of identity and a way to interact and communicate with other objects in virtual environments. Human-driven avatar development depends on the activity context to create interactive avatars[3]. Moreover, integrating AI and IoT in avatar creation can provide intelligent, human-like digital meta-humans and enhance metaverse services.

Techniques for generating 3D avatars are widely deployed using images to digitalize 3D faces and bodies of personalized avatars[4,5]. Nowadays, the reconstruction of 3D human shapes has become more accessible by implementing generative AI for digital assets in virtual environments, reducing time-consuming processes. Integrating text and image-driven techniques makes 3D avatar creation more reliable, facilitating high resolution and customizable features by understanding different perspectives of the model[6]. Utilizing a generative adversarial network (GAN) enables the reconstruction, encoding, and gen-

※ This research was supported by Kumoh National Institute of Technology (2022).

◆ First Author : Kumoh National Institute of Technology, Department of Electronic Engineering, and with the Politeknik Elektronika Negeri Surabaya, Department of Electrical Engineering, zai@kumoh.ac.kr, 학생회원

° Corresponding Author : Kumoh National Institute of Technology, Department of IT Convergence Engineering, ljmpaul@kumoh.ac.kr, 종신회원

* Kumoh National Institute of Technology, ICT Convergence Research Center, esmot@kumoh.ac.kr

** Kumoh National Institute of Technology, Department of IT Convergence Engineering, dskim@kumoh.ac.kr, 종신회원

논문번호 : 202407-127-D-RN, Received June 29, 2024; Revised September 17, 2024; Accepted October 4, 2024

eration of 3D vision of facial avatars. Furthermore, the Contrastive Language-Image Pre-training (CLIP) facilitates textand image-driven 3D manipulation of avatars with user-provided text synchronization[7]. However, existing avatar reconstruction approaches[4,5,7] only focus on 3D model construction and lack movement recognition capabilities to synchronize human objects in physical and avatars in virtual environments.

Most of the recent virtual environment services used joysticks, keyboards, touchscreens, and mouses to guide avatars' movement and their location. Nevertheless, these techniques have limited movement capabilities (standing, walking, running, and jumping) and are not supported with face expression features to make more realistic metaverse applications. Some approaches integrate an AI-based human activity recognition (HAR) method to produce interactive avatars by utilizing WiFi channel state information (CSI)[8], inertial measurement unit (IMU) sensors[9], and image processing techniques[4,5]. However, these methods[8,9] only facilitate restricted activities based on the known activities. The HAR model cannot recognize detailed unknown movements. Furthermore, these techniques[4,5] generated a 3D human shape body and face of personalized avatars from image files that are not directly deployed in metaverse applications.

Applying machine learning (ML) and deep learning (DL) techniques in metaverse services can enhance the immersive experience and improve metaverse application quality[10]. The DL is powerful for representing 3D human avatars[6]. A framework for real-time recognition of body movements and head gestures was developed using DL techniques[11]. This method utilized a 1D CNN model to recognize body actions when performing walkingin-place (WIP) and head gestures for head-mounted display (HMD) in VR applications. Furthermore, a web-based avatar controller uses the hand gestures technique to manage the movement of virtual puppets[12]. In [13], the synthesized 3D human avatars were developed using Hybrid VolumetricTextural Rendering (HVTR++) method. This technique encodes the driving signals of the human body posture and maintains the integrity of a parametric model based on a skeleton's structure. A long shortterm memory (LSTM) model called GeoPose was deployed to learn human joint points for full-body avatar construction using a geometry-incorporated method[14]. Based on captured human poses from two controllers and HMD, the GeoPose transforms the data to the avatar frame and distinguishes the global motion using a nine-connected LSTM model.

The 3D facial avatar digitization is important to enhance the user's immersive experience in a metaverse environment. The facial expression capabilities of the avatars can facilitate trust-building between users in an immersive virtual scene[15]. The creation of facial animation for the basic emotion of the 3D avatar was developed[16]. This method utilized the generation of animation, face capture, and emotion estimation techniques with lip movement synchronization. In [17], the expression speech-driven facial animation framework was implemented to address the limitation of flexibility in emotion control. In this system, a speech-to-face facial animation generated by DL can display various facial expressions with adjustable emotion type and intensity. An avatar expression reenactment was developed using open-flow prediction[18]. This approach eliminated artifacts via illumination consistency and solved the identity issue by estimating local and overall optical flows using the Overall-Local Feature Warping Fusion Model (O-LFWFM) to estimate expression reenactment.

Considering the improvement of users' immersive experience in metaverse services, this study develops an interactive human-centric avatar generation framework by utilizing motion capture to enable unlimited detail human movement capabilities and face expression detection to provide human facial expression for intelligent metaverse services. This system applied an open-source DL-enabled platform, Mediapipe framework, to facilitate real-time human poses and movement detection. Moreover, this framework provides accurate facial expression detection for meta-human avatars in the metaverse. At first, to capture human movements in the physical environment, the framework uses the skeleton-based human detection technique. The decisions are synchronized with avatar animation to generate human-like digital meta-humans in the virtual environment. Considering the unlimited human movement capabilities with face recognition features, this study proposes potential contributions as follows:

1. **Integrated movement detection and face expression recognition 3D avatar reconstruction framework capabilities**: We developed a human-centric interactive avatar reconstruction framework using motion capture and face expression detection to improve users' immersive experience in metaverse services. Motion capture facilitates unlimited human movement synchronization between physical and virtual environments based on captured skeleton data. Additionally, face expression detection enables real-time correspondence of facial expression recognition, contributing to the development of human-like avatars.
2. **Metaverse platform integrated development**: We utilized the Mediapipe framework and integrated it with the Unreal Engine (UE) based metaverse platform to provide avatar development in a practical approach. Human movement and facial expressions are captured using a camera in the physical environment and represent the mapped features through an interactive metahuman avatar in the virtual environment platform.
3. **Intensively analyzed interactive avatar generation framework reliability**: We analyzed the proposed system performance in realtime integrated framework scenarios regarding movement and face expression detection accuracy. Moreover, the delay process is analyzed to validate the reliability of the proposed framework against real-time requirements in the metaverse services.

The rest of this study is organized as follows: Section II presented the related works of interactive avatar reconstruction in the metaverse environments. In Section III, the proposed system is described. Section IV presents the implementation of the proposed system including the human movement avatar detection framework development. The simulation results and discussion are presented in Section V Finally, the conclusion and future works are presented in Section VI.

Ⅱ. Related Works

Creating human-like digital characters or virtual avatars for immersive media was widely developed. 3D face generation using a single image file is a popular method for constructing 3D human avatars. StyleGAN[4], InferGAN[5], PiCA[19], and Avatar-Poser[20] techniques facilitate 3D avatar reconstruction using GAN model capabilities. The GAN technique enables the automatic 3D body and face digitization with enhanced experience to address traditional methods' likeness and facial details issues when using linear models[11,21]. Furthermore, GAN-based 3D model reconstruction provides realistic 3D full-body and face avatar design with a rapid process. In [22], the avatar generation method was implemented using the image's shape, pose, and semantic information to reconstruct realistic 3D avatars. This method employed AvatarSDK and the SMPL model to rebuild the character to a high-quality T-pose avatar. The web browser-based realistic digital human creation tool was utilized to meet the high demands of metahuman virtual production for next-generation platforms[23]. This software facilitates customized photorealistic digital humans, allowing users to change hair and clothing based on their wants. Integrating images and text descriptions as the multi-modal input provides expressive and high-quality 3D avatar models. Avatar-Verse[24] and Guide3D[25] methods enable 3D pose guiding of avatars through 2D images and complex text prompts. However, these approaches[4,5,19-25] concern for developing of 3D human avatar models and lack for avatar movement capabilities.

Leveraging the Internet of Things (IoT) and DL in a metaverse environment can facilitate intelligent metaverse services. In [8], the authors proposed a WiFi-based smart home system for human pose estimation in metaverse avatar simulation. This approach utilized WiFi CSI to classify human pose and generate avatars based on the data obtained from WiFi sensing data. The collection of CSI sensing data suffers from phase shift problems, which negatively impact the extraction of information, leading to a decline in the performance of avatar construction. To address this issue, the authors[26] corrected the CSI phase shifts using phase compensation and sliding window. Instead of depending on a single device, the authors[27] proposed semantic sensing information transmission for the metaverse. Combine multiple sensing data, for example, mobile phones, WiFi, and others, and encode using semantic encoding before sending it to the metaverse to reduce information loss. Furthermore, the avatar creation framework utilized IMU sensors on the jacket to recognize human movement for synchronizing avatar animation in a metaverse platform[9]. To estimate full-body poses, Avatar-Poser[20] technique used 6-DoF sensors with motion input data from the user' s hands and head. This method comprised four modules: stabilizer, transformer encoder, forward-kinematics, and backwardkinematics for enabling holistic avatars. However, the sensor-based metahuman creation processes[8,9,20,26,27] has some major limitations. For instance, facial detail expressions are difficult to create in the metaverse, and it is challenging to identify gender using sensor data.

The effectiveness of avatars is essential in conveying emotions that play a vital role in social presence and user experience, encompassing both verbal and non-verbal facial and body signals. Facial expression capabilities facilitate human-like and realistic avatar creation in metaverse environments. The authors[28] utilized Facial+eye Tracking (FT) and Lip-Sync approximation (LS) techniques for comparing the effectiveness capabilities to provide six basic emotions (surprise, fear, happiness, disgust, sadness, and anger). The Meta-EmoVis[29] method was implemented to enhance the realism of facial expressions and human emotional features in real-time virtual avatars. This technique used the Emoj tool for detecting faces and classifying emotions through a webcam device. An avatar re-enactment method was developed in [18] for a metaverse-based exhibition interaction system. This approach used Optical-Flow field prediction to provide a realistic recreation of face features. In [30], the authors proposed real-time virtual humans in the metaverse, generating six facial expressions by setting parameters. This process is not dynamic, and avatars only use predefined expressions. In [31], adopt Octree-based real-time metahuman simulation in the metaverse environment. However, the performance of synchronization and delay needs to improve. In addition, these approaches[18,28-31] only facilitate facial expression features and do not consider combination with human movement detection features that enable metaverse platform integrated development in real-time scenarios.

Furthermore, in [34], the authors apply the avatar creation strategy to the metaverse platform and deploy it with a more user-friendly interface. This framework enables users to personalize their avatars by customizing facial traits and clothing styles. The authors claim that the customized avatars can increase user participation by providing enhanced immersion and stimulating social connections. An avatar design using the cognitive walkthrough (CW) technique was developed for virtual education[35]. This system utilizes the VRoid platform for avatar creation, offering reliable customization features and a user-friendly interface. By leveraging a user-centric avatar design framework, user interaction can be enhanced, leading to a more intuitive metaverse environment. An avatar ceation framework was implemented to serve as the digital representation of the user in the metaverse based on the event expectation[36]. In this system, users design formal or informal avatars based on the expected hedonic or utilitarian value of the event. The designed avatars represent the users' identities and self-expression, enabling high interactivity. However, these approaches[34-36] lack AI integration to enable interactive avatar creation based on user activities in the physical world for real-time scenarios.

The comparison between existing avatar creation techniques with the proposed system in metaverse environments is presented in Table 1. This table extensively compares several capabilities of the proposed system, such as movement detection (MD), face expression detection (FED), unlimited activity recognition capabilities (UARC), integrated movement, and face expression detection (iM-FED), and metaverse platform integrated development (MPID). This com-

Table 1. Comparison between Existing Avatar Creation Techniques with Proposed System in Metaverse Environments

Ref	Year	Method	Description	MD	FED	UARC	iM-FED	MPID
[4]	2021	StyleGAN	Digitizing 3D faces of personalized avatars from unconstrained images using StyleGAN2 technique. The FaceNet model was utilized for pre-trained and provide 3D face recognition.	✗	✓	✗	✗	✗
[5]	2021	InferGAN	Using a silhouette-based dense correspondence to construct a 3D human avatar. The final shape of the avatar body was created from the image of the human of body, segmented part, and wrapped of the initial geometry of the body.	✓	✓	✗	✗	✗
[19]	2021	PiCA	Develop a 3D human faces using deep generative model that facilitate efficient computation and adaptive rendering.	✗	✓	✗	✗	✗
[23]	2021	MetaHuman Creator	A web browser-based application to create realistic digital human by customize the hair and clothing in fast away.	✓	✗	✗	✗	✓
[9]	2022	IMU sensor-based HAR	Utilizing IMU-based sensor jacket to recognize human movement for synchronizing avatar animation in metaverse environment. This system used six IMU sensors in the jacket pocket as the reference for the human movements.	✓	✗	✗	✗	✗
[20]	2022	AvatarPoser	Full-body poses estimation using 6-DoF sensors to enable holistic avatar in the metaverse platform. This approach was composed by four modules including stabilizer, transformer encoder, forward-kinematics, and backward-kinematics.	✓	✗	✗	✗	✓
[22]	2022	AvatarSDK and SMPL	Avatar generation method utilize image's shape, pose, and semantic information to reconstruct a realistic 3D avatar. This method employs AvatarSDK and the Skinned Multi-Person Linear (SMPL) model to rebuild the character to T-pose avatar with high quality.	✓	✗	✗	✗	✗
[32]	2022	Unity-PUN2	Design of an avatar for multi-player application that developed in Unity3D using Photon Unity Networking 2 (PUN2) plug-in.	✓	✗	✗	✗	✓
[6]	2023	AvatarCraft	Using difusion model for creating human avatars with learning of geometry and texture. An avatar animation was generated and controlled by the shape and pose parameters.	✓	✗	✗	✗	✗
[8]	2023	MetaFi++	Utilizing integrated CSI WiFi signal and camera for avatar simulation. A convolution module was used to recognize human poses.	✓	✗	✗	✗	✗
[33]	2023	Self-avatar configuration	A technique for crafting human-like avatars through the direct configuration of life-sized avatar in virtual environment facilitates a seamless transition.	✓	✗	✗	✗	✗
[24]	2023	AvatarVerse	This method facilitates an expressive high-quality 3D avatar creation based on customized text description and pose guidance estimation using pre-trained DensePose model.	✓	✗	✗	✗	✗
[34]	2024	3D Avatar	apply the avatar creation strategy with user-friendly interface by customizing facial traits and clothing styles.	✗	✗	✗	✗	✓
[35]	2024	VRoid	An avatar design using the cognitive walkthrough (CW) technique was developed for virtual education.	✗	✗	✗	✗	✓
Ours	2024	MD-FED	Integrated motion detection and face expression detection to provide interactive avatar in metaverse. This framework provides unlimited activity recognition capabilities.	✓	✓	✓	✓	✓

* MD : Movement Detection, * FED : Face Expression Detection, * UARC : Unlimited Activity Recognition Capabilities, * iM-FED : Integrated, Movement and Face Expression Detection, * MPID : Metaverse Platform Integrated Development, ✓ : Considered, ✗ : Not Considered

parison indicates that an interactive avatar generation framework that provides movement detection and face expression capabilities in an integrated system for enhancing metaverse services is an unsolved problem. Therefore, this study implements a humancentric interactive avatar reconstruction framework that enables movement detection for unlimited activity recognition capabilities and face expression detection features in a metaverse environment. The proposed system was developed and validated in a realtime integrated metaverse platform. Fig. 1 presents the metaverse components supported by avatar development for interactive avatar-enabled metaverse services.

Ⅲ. Proposed Human-Centric Interactive Avatar Generation Framework

3.1 Problem Formulation

The recent 3D avatar reconstruction techniques[4,5,19-25] utilized image files and text descriptions to provide digital characters or virtual avatars for immersive media. These methods used the GAN model to create 3D body and face digitation with an enhanced experience and rapid process. However, these methods[4,5,19-25] only facilitate for deploying 3D human avatar models and not consider avatar movement capabilities. Some virtual services utilized joysticks, keyboards, touchscreens, and mouses to guide avatars' movement. Nevertheless, these methods have limited movement capabilities (standing, walking, running, and jumping). Implementing an IoT-based human activity recognition (HAR) approach to simulate avatars in the metaverse can provide synchronous movement detection in physical and virtual environments. Several HAR-based avatar creation frameworks were deployed, such as IMU sensor-based[9], WiFi CSI-based[8], image processing-based[4,5], and 6-DoF sensor-based[20] avatar simulation techniques. However, these approaches[4,5,8,9,20] can recognize only the known human movements.

Utilizing ML/DL can improve the immersive experience of metaverse services. The ML/DL technique is powerful enough to recognize full-body actions and head gestures in real-time 3D human avatar reconstruction[6,10]. Several methods[11,12] were deployed to enable recognizing body actions in a physical environment and synchronously construct avatar movement in virtual environments. The hand gestures method was implemented in a webbased avatar controller to guide the actions of virtual puppets. Furthermore, the HVTR++[13] technique can synthesize 3D human avatars using encoded human body pose signals and integrated with the skeleton's structure. However, these approaches[6,10,13] do not consider the avatar's facial expression capabilities. Facial avatar digitization can build trust between users and make it a human-like avatar in an immersive virtual scene. The facial expressiondriven avatar generation platforms[15,16,18] utilized generation animation, face capture, and emotion estimation techniques to synchronize lip movement. Moreover, using expression speech, the speech-toface facial animation framework was generated by a DL model to display various facial expressions[17]. However, these methods[15-18] only focus on providing facial expression features for 3D avatar reconstruction. Therefore, this study proposes integrated motion capture and face expression detection for a human-centric interactive avatar reconstruction framework in a metaverse service.

3.2 Proposed System

This study utilizes Mediapipe to generate real-time human poses, movements, and facial expressions on the metaverse platform. To enhance interactive avatar development, the system integrates motion and expression capture functions to enable dynamic avatar deployment. The proposed human-centric interactive avatar generation framework is shown in Fig. 1. The system consists of physical and virtual environments with full synchronization capability. In the physical environment, a camera captures the users' movements and facial expressions in the edge server. Subsequently, the captured data are processed in the cloud-based metaverse server through B5G networks.

The system comprises body movement recognition (BMR) and facial expression recognition (FER) units. The BMR unit maps the user's skeleton data and skeleton model to the avatar's gesture for providing detailed movements. On the other hand, the FER unit enables an avatar's verbal expression and eye gaze

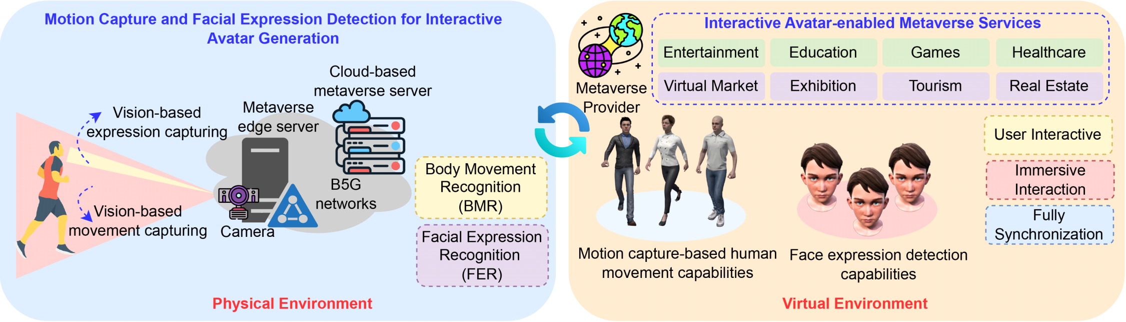

Motion Capture and Facial Expression Detection for Interactive Avatar Generation
Cloud-based metaverse server
Vision-based expression capturing
Metaverse edge server
B5G networks
Vision-based movement capturing
Camera
Body Movement Recognition (BMR)
Facial Expression Recognition (FER)
Physical Environment
Metaverse Provider
Interactive Avatar-enabled Metaverse Services
Entertainment
Education
Games
Healthcare
Virtual Market
Exhibition
Tourism
Real Estate
User Interactive
Immersive Interaction
Fully Synchronization
Motion capture-based human movement capabilities
Face expression detection capabilities
Virtual Environment

Fig. 1. Metaverse Components Supported by Avatar Development for Interactive Avatar-enabled Metaverse Services

using facial motion capture. This system utilized the MediaPipe4U plug-in and Unreal Engine (UE)based metaverse platform to facilitate integrated motion detection and facial expression recognition in interactive avatar real-time development. The MediaPipe4U plugin facilitates development or collaboration efforts to bring MediaPipe functionalities into the UE environment.

3.2.1 MediaPipe4U Plug-in Overview

MediaPipe4U is an open-source UE plug-in for 3D meta-human development that uses MediaPipe to enable complex technologies in UE environments1). MediaPipe is a multifunctional platform developed by Google2). This plug-in consists of six main functions such as motion capture, expression capture, multiple capture sources, large language model (LLM), text-to-speech (TTS), and speech recognition (ASR). In addition, this plugin supports driving any 3D character, including Mannequin, Virtual Reality Model (VRM), Metahuman, ReadyPlayerMe, etc. The MediaPipe4U plugin provides some functions to enable dynamic and reliable capabilities, including full-body/half-body motion capture, facial expression capture, character movement, finger snap, joint distortion correction, smooth animation functions and etc[37].

3.2.2 Body Movement Recognition (BMR) for Avatar Reconstruction

The workflow proposed BMR unit is presented in Fig. 2. The proposed BMR unit consists of three main components: the physical environment, the MediaPipe4U plug-in, and the virtual environment. In the physical environment, the input video frames of the user are captured using a camera as the input of the MediaPipe4U plug-in. The MediaPipe4U plugin performs some processes, including transforming the image to a tensor, utilizing a model for the inference process, transforming the tensor to landmarks, acquiring the required landmarks, and drawing landmarks and connections on the input frame. The virtual environment utilizes the Unreal Engine metaverse platform to generate an avatar. To create an interactive avatar, several steps are required, including avatar construction, full-body animation rigging, inverse kinematics, and avatar rotation.

This system employed the MediaPipe holistic model to combine pose and face landmarks. If we observe this Fig. 2, the pose detection model only covers a few key face points, such as the nose, eyes, and lips. On the contrary, the face detection model precisely recognizes a range of facial expressions based on the numerous identification points. Employing the Holistic model enables the creation of an avatar that faithfully replicates genuine human expressions in real-time. The Holistic landmark leverages a machine model to enable the ongoing generation of gestures, poses, and actions, incorporating 33 pose landmarks, 468 face landmarks, and 21 hand landmarks for each hand.

3.2.3 Facial Expression Recognition (FER) for Avatar Reconstruction

Along with body movement recognition (BMR), avatars also need to exhibit expressions like those of

1) https://opensource.labijie.com/Mediapipe4u-plugin/
2) https://developers.google.com/mediapipe

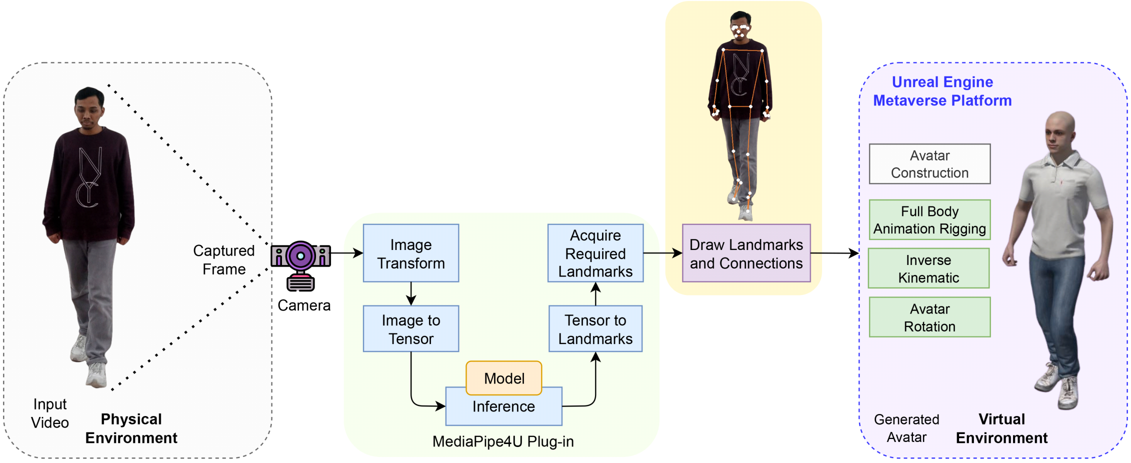


Fig. 2. Proposed MediaPipe-assisted Body Movement Recognition (BMR) Unit for Interactive Avatar Reconstruction

humans in the physical world to make metaverse social interactions more realistic. Therefore, our proposed method also integrates a facial expression recognition (FER) module. Fig. 3 shows the functional diagram of the FER module. As can be seen in the figure, first, the camera is utilized to capture real-time video from the physical world. These images are then fed into the MediaPipe4U plug-in. With the help of MediaPipe, the face area is first detected, followed by face embedding and facial feature extraction. Later, facial features are transformed into facial landmarks. MediaPipe can detect 478 facial landmarks. The next step involves transform the landmark information into the 3D metahuman model (avatar) in the metaverse to reconstruct the same facial expression on the avatar, matching it with the real-world human. The proposed MediaPipe-assisted FER can recognize and reconstruct a total of 51 facial expressions.

3.3 System Configuration

This framework utilizes the open-source MediaPipe4U plug-in to generate interactive avatars based on human poses in the real-world environment. The captured video then passes through the Mediapipe plugin, which estimates landmarks. A camera captures the human poses and maps to skeleton data or pose landmarks. Furthermore, the captured human face is transformed to face landmarks. After that, the metahuman avatar in the metaverse will be connected with the Mediapipe plugin using an animation blueprint. The avatar movement and facial expression will reflect the Mediapipe plug-in's realtime input. The Mediapipe functions used are as follows: MediapipePoseSolver, MediapipeHandSolver, and MediapipeLocationSolver, and lastly, everything connected with the Output Pose. MediaPipePoseSolver involves compensating for body movement and determining the rotation of body bones. MediaPipePoseSolver calculates finger bone rotation and motion compensation. The calculation of character displacement is performed by the MediaPipeLocationSolver module. The detailed facial expression is captured by the MediaPipeLiveLinkComponent, which functions as the interface responsible for streaming and constructing facial data from the external world to the metaverse avatar.

Ⅳ. System Implementation

We utilize the Unreal Engine (UE) 5.1.1 platform to create the metaverse environment. However, it is worth noting that Mediapipe can also be seamlessly connected with other platforms, such as Unity 3D. The metaverse platform working on a Windows 10 operating system with a Core i5-8500 processor and

16GB RAM. To support the development of an avatar in the UE environment, we employ Visual C++, Visual Studio 2022, and Windows SDK software tools. The streaming video of the users is captured using a Logitech BRIO 500 webcam with 1280 ╥ 720

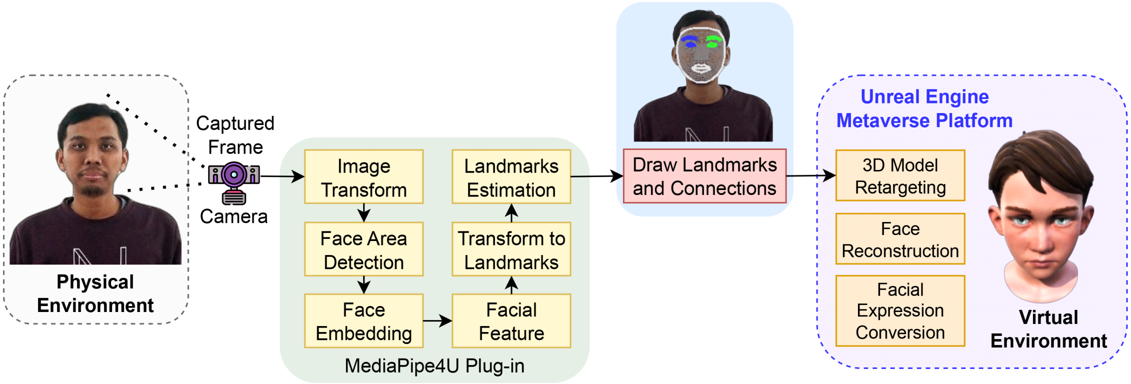


Fig. 3. Proposed MediaPipe-assisted Facial Expression Recognition (FER) Unit for Interactive Avatar Reconstruction

pixels

60 fps. To utilize Mediapipe within the UE, activating certain plugins, including GStreamer, MediaPipe4U, MediaPipe4UBVH, MediaPipe4UGStreamer, MediaPipe4ULiveLink, and MediaPipe4UNvAR, is essential. This system conducts several steps in developing an interactive avatar framework, such as installing the MediaPipe4U plugin, GStreamer, and human movement avatar implementation.

4.1 Plugin Installing and Dependencies

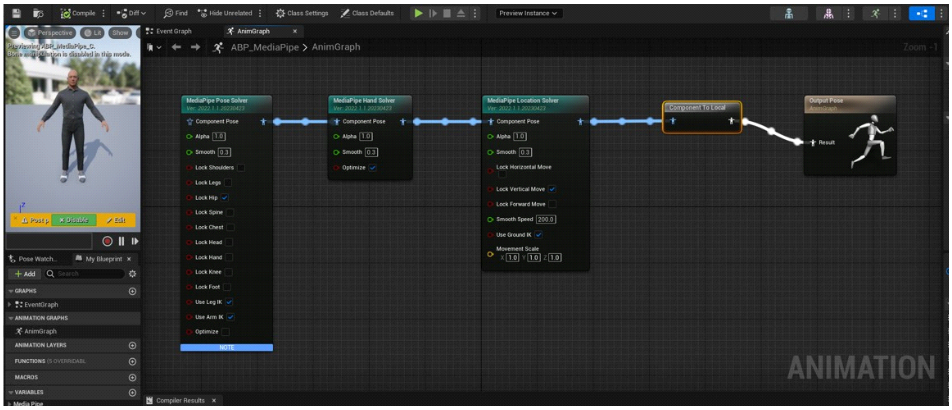
To install and integrate the plugin with the UE environment, it is necessary to enable specific plugins, including MediaPipe4ULiveLink, MediaPipe4UGStreamer, GStreamer, and MediaPipe4U. GStreamer is a multimedia framework designed as a pipeline to process a wide variety of multimedia components. GStreamer is utilized in Unreal Engine for real-time streaming, multimedia processing, synchronization, and integration with other multimedia formats. The MediaPlayer in Unreal Engine experiences challenges in decoding certain H264 encoded video files effectively. As a result, MediaPipe4U utilizes GStreamer as an optimal alternative to MediaPlayer for processing video streams, particularly for motion capture from video. Live Link plugin provides a unified interface for streaming and consuming animation data from external sources. However, the MediaPipe4ULiveLink plugin is a modified version of the Live Link plugin, and the MediaPipe4UGStreamer is a modified version of GStreamer. Both are used to support the MediaPipe4U plugin.

After completing the necessary plugin installations, the next step involves configuring GStreamer. The default runtime installer functions only as a basic decoder, necessitating the installation and adjustment of custom plugins. During installation, it is essential to ensure that "libav" is installed and to exclude "QT" to reduce the plugin size. Finally, the plugin path needs to be added and configured in the *environment variables*.

4.2 Human Movement Avatar Detection

We prepare to animate the character by creating an animation blueprint in this step. This animation blueprint selects the MediaPipeAnimInstance as the base class. Subsequently, the three nodes MediaPipePoseSolver, MediaPipeHandSolver, and MediaPipeLocationSolver are put in cascade configuration as in Fig 4. The MediaPipePoseSolver node is responsible for calculating the body bone rotation and body movement compensation calculation. The MediaPipeHandSolver is a finger motion compensation solution node. It is responsible for calculating the finger bone rotation, and the parameters are the same as those of MediaPipePoseSolver. Additionally, the MediaPipeLocationSolver node is the dynamic compensation displacement calculation node responsible for calculating the character displacement.

A webcam captures human pose, movement, and facial expression. The captured data is then processed using machine learning techniques, specifically Mediapipe. The processed information is then trans-



MediaPipe Pose Solver
Ver. 2022.1.1.20230423
Component Pose
Alpha 1.0
Smooth 0.3
Lock Shoulders □
Lock Legs □
Lock Hip ☑
Lock Spine □
Lock Chest □
Lock Head □
Lock Hand □
Lock Knee □
Lock Foot □
Use Leg IK ☑
Use Arm IK ☑
Optimize □
NOTE

MediaPipe Hand Solver
Ver. 2022.1.1.20230423
Component Pose
Alpha 1.0
Smooth 0.3
Optimize ☑

MediaPipe Location Solver
Ver. 2022.1.1.20230423
Component Pose
Alpha 1.0
Smooth 0.3
Lock Horizontal Move □
Lock Vertical Move ☑
Lock Forward Move □
Smooth Speed 200.0
Use Ground IK ☑
Movement Scale X 1.0 Y 1.0 Z 1.0

Component To Local

Output Pose
AnimGraph
Result

Fig. 4. Creating Animation Blueprint with Cascade Configuration of MediaPipePoseSolver, MediaPipeHandSolver, and MediaPipeLocationSolver nodes

mitted to the avatar, causing its skeleton bones to move in accordance with the data received from Mediapipe. In our study, we employ the Unreal Metahuman character. In the case of Maximo 3D character, it is necessary to connect the skeleton bones according to the Mediapipe bone hierocracy.

4.3 Performance Optimization

In our work, we utilized the MediaPipe framework for pose, facial expression, and activity detection. The vision information feed from the camera generates landmarks using MediaPipe. This landmark information is subsequently transferred to the metaverse avatar to replicate movements, expressions, and activities. However, there is a delay in the transition from real-world image/video capture to avatar generation within the metaverse, which hinders the performance of real-time avatar formation. To address this issue, we used optimization techniques needed to enhance the performance. There are several potential solutions to improve performance, including enhancing the end device hardware and network capabilities. Additionally, reducing the frame rate can decrease computational costs, and model quantization can be applied to reduce latency. Furthermore, integrating a Kalman filter could be effective for generating smooth movements, particularly in facial expression detection.

Ⅴ. Simulation Results and Discussion

In this section, the experimental results are evaluated to analyze the robustness of the proposed system. We investigate the reliability of body movement recognition (BMR) and facial expression recognition (FER) frameworks. First, the delayed response of movement in both physical and virtual environments is calculated. This measurement is conducted to ensure the real-time capability of the proposed body movement and facial expression detection systems. The measurement results of delayed response BMR and FER framework are resented in Table 2 and Table 3, respectively. In this evaluation, we use activities including raising the right hand, raising the left hand, raising both hands, moving a hand forward, moving a hand backward, lifting the right leg, lifting the left leg, and facing backward for evaluating the BMR framework. We compare the user activities in a physical and virtual environment by calculating the delay response and precision. Based on these measurement results, the average delay for raising the right hand requires a delay of 2.06 seconds until finishing the raising hand. A similar value of 1.96 seconds is achieved for raising the left hand. Raising both hands requires more time around 2.15 seconds to complete this activity. The reason is raising both hands does a more complex activity than raising one hand, therefore the system requires more complicated to identify the skeleton data. For these three activities perform the precision of 98.82%, 98.85%, and 97.95%, respectively. The precision of raising both hands is low because the skeleton data of this activity is slightly detected.

Table 2. Delay and Precision Performances for BMR Framework

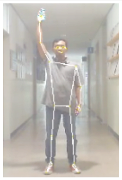
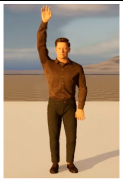
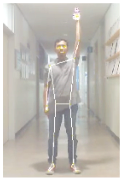
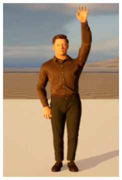
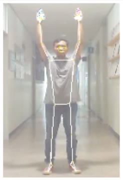
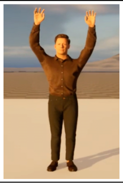
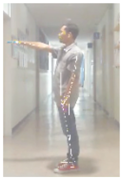
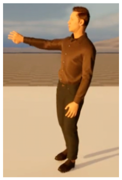
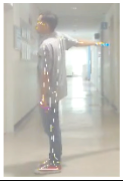
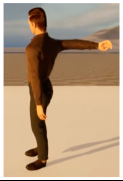
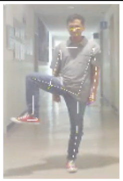
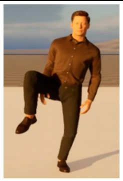
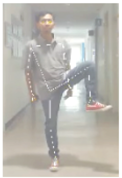
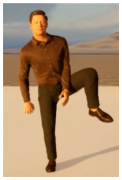
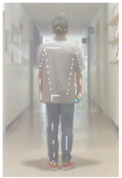
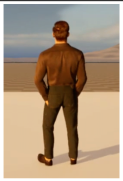
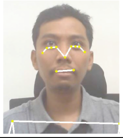
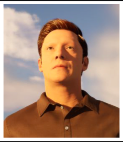
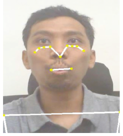
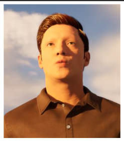
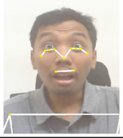
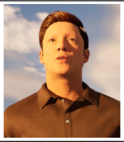
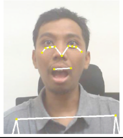
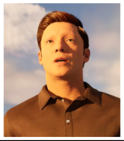
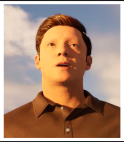
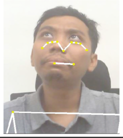
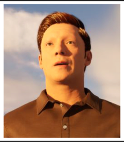
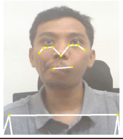
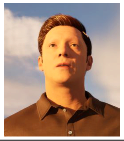
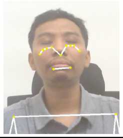
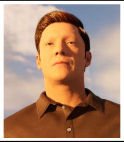
No	Activity	User Activity in Physical Environment	Captured Avatar in Virtual Environment	Delay (s)	Precision (%)
1	Raising the right hand			2.06	98.82%
2	Raising the left hand			1.96	98.85%
3	Raising both hands			2.15	97.95%
4	Moving a hand forward			1.84	98.56%
5	Moving a hand backward			1.72	98.04%
6	Lifting the right leg			1.83	98.95%
7	Lifting the left leg			1.96	98.92%
8	Facing backward			2.18	97.53%

Table 3. Delay and Precision Performances for FER Framework

No	Expression	User Expression in Physical Environment	Avatar Expression in Virtual Environment	Delay (s)	Precision (%)
1	Natural			1.24	97.85%
2	Accepting with a pouty mouth			1.46	97.56%
3	Surprise with mouth open			1.55	94.27%
4	Just knowing with mouth open			1.35	95.38%
5	pouting with lips tilted to the left			1.42	93.15%
6	Thinking with eyes looking upwards			1.46	93.26%
7	pouting with lips tilted to the right			1.38	93.38%
8	Concentrating with eyes closed			1.64	93.52%

The other activities such as moving a hand forward, moving a hand backward, lifting the right leg, lifting the left leg, and facing backward require a delay of 1.84 seconds, 1.72 seconds, 1.83 seconds, 1.96 seconds, and 2.18 seconds, respectively. The activity of moving a hand backward achieves the lowest time. On the other hand, the facing backward activity requires the highest time delay. Furthermore, the precision achieves 98.56% for moving a hand forward, 98.04% for moving a hand backward, 98.95% for lifting the right leg, 98.92% for lifting the leg, and 97.53% for facing backward. The delay and precision performance for the FER framework are presented in Table 3. This measurement investigates the delay and precision of several expressions including natural, accepting with a pouty mouth, surprise with mouth open, just knowing with mouth open, pouting with lips titled to the left, thinking with eyes looking upwards, pouting with lips titled to the left, and concentrating with eyes closed. Based on the investigation results, these facial expressions require an avatar expression generation delay of 1.24 seconds, 1.46 seconds, 1.55 seconds, 1.35 seconds, 1.42 seconds, 1.46 seconds, 1.38 seconds, and 1.64 seconds, respectively. Furthermore, the natural expression has the highest precision of 97.85%. The lowest precision is achieved by the 'pouting with lips titled to the left' expression with 93.15% of precision. The FER framework performs poorly to generate the expressions at the mouth and eyes parts. Therefore, this framework needs to improve the capability for particular expressions such as concentrating with eyes closed, pouting with lips titled to the right, thinking with eyes looking upwards, and pouting with lips titled to the left.

Ⅵ. Conclusion

This study integrate human body movement using a skeleton and facial expression to generate an interactive avatar on a metaverse platform. We use Mediapipe4U plugin for real-time avatar creation in the metaverse. In this framework, we can utilize a webcam to capture the users' movement and facial expression and generate the representative avatar in the vitual environment. This avatar generation framework suitable for meetings, conferences, and classrooms inside the metaverse, with accurate human movement and facial expression detection capabilities. Based on the measurement results, the proposed BMR framework performs the highest precision of

98.95% for lifting the right led with delay of 1.83 seconds. Furthermore, the proposed FER framework achieve the highest precision of 97.85% for natural expression with delay process of 1.24 seconds.

In future work, there is an opportunity to utilize additional MediaPipe4U features such as audio classification, text embedding, and language detection for the metaverse. As well, there is the possibility to adjust real-time scheduling to minimize delay, Kalman filter-based error correction, and synchronization for avatar performance improvement. Moreover, the open research issues for avatar creation in the metaverse environment include: (a) AIbased avatar creation scalability, (b) privacy and security in interactive avatar generation framework, and (c) reliable and real-time avatar development techniques for metaverse platform. To address the challenges of these research issues, exploring generative AI and large language model (LLM) presents a potential solution for an AI-empowered interactive avatar creation platform. Integration decentralized learning and blockchain can enhance privacy and security by implementing decentralized authentication for avatar customization. Lastly, adopting edge computing with a task off-loading mechanism can provide reliable and real-time avatar development techniques.

References

[1] T. Huynh-The, Q.-V. Pham, X.-Q. Pham, T. T. Nguyen, Z. Han, and D.-S. Kim, "Artificial intelligence for the metaverse: A survey," *Eng. Appl. Artificial Intell.*, vol. 117, p. 105581, 2023. (https://doi.org/10.1016/j.engappai.2022.105581)

[2] K. Li, Y. Cui, W. Li, et al., "When internet of things meets metaverse: Convergence of physical and cyber worlds," *IEEE Internet of Things J.*, vol. 10, no. 5, pp. 4148-4173, 2022. (https://doi.org/10.1109/JIOT.2022.3232845)

[3] D. Zimmermann, A. Wehler, and K. Kaspar, "Self-representation through avatars in digital environments," *Current Psychol.*, vol. 42, no. 25, pp. 21775-21789, 2023. (https://doi.org/10.1007/s12144-022-03232-6)

[4] H. Luo, K. Nagano, H.-W. Kung, et al., "Normalized avatar synthesis using stylegan and perceptual refinement," in *Proc. IEEE/CVF Conf. CVPR*, pp. 11662-11672, Nashville, TN, USA, 2021. (https://doi.org/10.1109/cvpr46437.2021.01149)

[5] Z. Li, L. Chen, C. Liu, et al., "Animated 3d human avatars from a single image with ganbased texture inference," *Computers & Graphics*, vol. 95, pp. 81-91, 2021. (https://doi.org/10.1016/j.cag.2021.01.002)

[6] R. Jiang, C. Wang, J. Zhang, et al., "Avatarcraft: Transforming text into neural human avatars with parameterized shape and

pose control," *ICCV*, pp. 14371-14382, 2023. (https://doi.org/10.1109/iccv51070.2023.01322)

[7] Z. Canfes, M. F. Atasoy, A. Dirik, and P. Yanardag, "Text and image guided 3D avatar generation and manipulation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vision,* pp. 4421-4431, Waikoloa, HI, USA, 2023. (https://doi.org/10.1109/wacv56688.2023.00440)

[8] Y. Zhou, H. Huang, S. Yuan, H. Zou, L. Xie, and J. Yang, "Metafi++: Wifi-enabled transformer-based human pose estimation for metaverse avatar simulation," *IEEE Internet of Things J.*, vol. 10, no. 16, pp. 14128-14136, 2023. (https://doi.org/10.1109/jiot.2023.3262940)

[9] D. Kim, J. Jeong, and Y. Chai, "A study on basic human activity recognition based on sensor jacket for interaction in metaverse environment," *Moving Image & Technol.* (*MINT*), vol. 2, no. 3, pp. 6-10, 2022. (https://doi.org/10.15323/mint.2022.8.2.3.6)

[10] Z. Lv, "Generative artificial intelligence in the metaverse era," *Cognitive Robotics*, vol. 3, pp. 208-217, 2023. (https://doi.org/10.1016/j.cogr.2023.06.001)

[11] J. Zhao, M. Shao, Y. Wang, and R. Xu, "Real-time recognition of in-place body actions and head gestures using only a head-mounted display," *2023 IEEE Conf. Virtual Reality and 3D User Interfaces* (*VR*), pp. 105-114, Shanghai, China, 2023. (https://doi.org/10.1109/vr55154.2023.00026)

[12] H. Luo, Y.-B. Lin, C.-C. Liao, and Y.-H. Huang, "An IoT-based microservice platform for virtual puppetry performance," *IEEE Access*, vol. 11, pp. 103014-103032, 2023. (https://doi.org/10.1109/access.2023.3315656)

[13] T. Hu, H. Xu, L. Luo, et al., "Hvtr++: Image and pose driven human avatars using hybrid volumetric-textural rendering," *IEEE Trans. Visualization and Computer Graphics*, vol. 30, no. 8, pp. 5478-5492, 2024. (https://doi.org/10.1109/tvcg.2023.3297721)

[14] T. Anvari and K. Park, "Geometry-incorporated posing of a full-body avatar from sparse trackers," *IEEE Access*, vol. 11, pp. 78858-78866, 2023. (https://doi.org/10.1109/access.2023.3299323)

[15] L. Luo, D. Weng, N. Ding, J. Hao, and Z. Tu, "The effect of avatar facial expressions on trust building in social virtual reality," *The Visual Computer*, vol. 39, pp. 5869-5882, 2023. (https://doi.org/10.1007/s00371-022-02700-1)

[16] D. Mukashev, M. Kairgaliyev, U. Alibekov, N. Oralbayeva, and A. Sandygulova, "Facial expression generation of 3D avatar based on semantic analysis," *2021 30th IEEE Int. Conf. Robot & Human Interactive Commun.* (*RO-MAN*), pp. 89-94, Vancouver, BC, Canada, 2021. (https://doi.org/10.1109/ro-man50785.2021.9515463).

[17] Y. Chen, J. Zhao, and W.-Q. Zhang, "Expressive speech-driven facial animation with controllable emotions," *2023 IEEE ICMEW*, pp. 387-392, Brisbane, Australia, 2023. (https://doi.org/10.1109/icmew59549.2023.00073)

[18] S. He, H. Zhao, and L. Yu, "The avatar facial expression reenactment method in the metaverse based on overall-local optical-flow estimation and illumination difference," *2023 26th Int. Conf. CSCWD*, pp. 1312-1317, Rio de Janeiro, Brazil, 2023. (https://doi.org/10.1109/cscwd57460.2023.10152763)

[19] S. Ma, T. Simon, J. Saragih, et al., "Pixel codec avatars," in *Proc. IEEE/CVF Conf. CVPR*, pp. 64-73, Nashville, TN, USA, 2021. (https://doi.org/10.1109/cvpr46437.2021.00013)

[20] J. Jiang, P. Streli, H. Qiu, et al., "AvatarPoser: Articulated full-body pose tracking from sparse motion sensing," *Computer Vision ECCV 2022,* pp. 443-460, Springer Nature Switzerland, 2021. (https://doi.org/10.1007/978-3-031-20065-6_26)

[21] M. T. Nguyen, T. V. Dang, M. K. T. Thi, and P. T. Bao, "Generating point cloud from measurements and shapes based on convolutional neural network: An application for building 3d human model," *Comput. Intell. and Neuroscience*, vol. 2019, no. 1, pp. 1-15, 2019. (https://doi.org/10.1155/2019/1353601)

[22] A. Beacco, J. Gallego, and M. Slater, "Automatic 3d avatar generation from a single RBG frontal image," *2022 IEEE Conf. Virtual Reality and 3D User Interfaces Abstracts and Workshops* (*VRW*), pp. 764-765, Christchurch, New Zealand, 2022. (https://doi.org/10.1109/vrw55335.2022.00233)

[23] Z. Fang, L. Cai, and G. Wang, "MetaHuman creator: The starting point of the metaverse," *2021 Int. Symposium on Comput. Technol. and Inf. Sci.* (*ISCTIS*), pp. 154-157, Guilin, China, 2021. (https://doi.org/10.1109/isctis51085.2021.00040)

[24] H. Zhang, B. Chen, H. Yang, et al., "Avatarverse: High-quality & stable 3d avatar creation from text and pose," in *Proc. AAAI Conf. Artificial Intell.,* pp. 7124-7132, Vancouver, Canada, 2024. (https://doi.org/10.1609/aaai.v38i7.28540)

[25] Y. Cao, Y.-P. Cao, K. Han, Y. Shan, and K.-Y. K. Wong, "Guide3d: Create 3d avatars from text and image guidance," *arXiv preprint arXiv:2308.09705*, 2023.

[26] J. Wang, H. Du, X. Yang, D. Niyato, J. Kang, and S. Mao, "Wireless sensing data collection and processing for metaverse avatar construction," *arXiv preprint arXiv:2211.12720*, 2022.

[27] J. Wang, H. Du, Z. Tian, D. Niyato, J. Kang, and X. Shen, "Semantic-aware sensing information transmission for metaverse: A contest theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5214-5228, 2023. (https://doi.org/10.1109/TWC.2022.3232565)

[28] A. Visconti, D. Calandra, and F. Lamberti, "Comparing technologies for conveying emotions through realistic avatars in virtual reality-based metaverse experiences," *Computer Animation and Virtual Worlds*, vol. 34, no. 3-4, pp. 1-11, 2023. (https://doi.org/10.1002/cav.2188)

[29] E. Spadoni, M. Carulli, M. Mengoni, M. Luciani, and M. Bordegoni, "Empowering virtual humans' emotional expression in the metaverse," *Int. Conf. Human-Computer Interaction,* pp. 133-143, Copenhagen, Denmark, 2023. (https://doi.org/10.1007/978-3-031-35897-5_10)

[30] M. Zhang, Y. Wang, J. Zhou, and Z. Pan, "SimuMan: A simultaneous real-time method for representing motions and emotions of virtual human in metaverse," *Internet of Things-ICIOT 2021: 6th Int. Conf., Held as Part of the Services Conf. Federation, SCF 2021,* pp. 77-89, Virtual Event, Dec. 2021. (https://doi.org/10.1007/978-3-030-96068-1_6)

[31] K. Y. Lam, L. Yang, A. Alhilal, L.-H. Lee, G. Tyson, and P. Hui, "Human-avatar interaction in metaverse: Framework for full-body interaction," in *Proc. 4th ACM Int. Conf. Multimedia in Asia,* pp. 1-7, Tokyo, Japan, 2022. (https://doi.org/10.1145/3551626.3564936)

[32] R. Jakob, V. Schmücker, T. J. Eiler, F. Grensing, and R. Brück, "The design of an avatar in a multiplayer serious game," *Current Directions in Biomedical Eng.*, pp. 153-156, 2022. (https://doi.org/10.1515/cdbme-2022-1040)

[33] J. Ichino and K. Naruse, "Virtual avatar creation support system for novices with gesture-based direct manipulation and perspective switching," in *Proc. 18th Int. Joint Conf. Comput. Vision, Imaging and Comput. Graphics Theory and Applications,* vol. *2,* VISIGRAPP, pp. 143-151, 2023. (https://doi.org/10.5220/0011630800003417)

[34] S. Noisri, P. Wisessing, K. Srisupakwong, H. B. Santoso, and L. Wittisittikulkij, "Designing avatar system and integrate to the metaverse,"

2024 ITC-CSCC, pp. 1-6, 2024. (https://doi.org/10.1109/itc-cscc62988.2024.10628399)

[35] H. Tinmaz and P. K. S. Dhillon, "User-centric avatar design: A cognitive walkthrough approach for metaverse in virtual education," *Data Sci. and Manag.*, 2024. (https://doi.org/10.1016/j.dsm.2024.05.001)

[36] S. Barta, S. Ibáñez-Sánchez, C. Orús, and C. Flavián, "Avatar creation in the metaverse: A focus on event expectations," *Computers in Human Behavior*, vol. 156, pp. 1-16, 2024. (https://doi.org/10.1016/j.chb.2024.108192)

[37] E. A. Tuli, A. Zainudin, M. J. A. Shanto, J. M. Lee, and D.-S. Kim, "MediaPipe-based real-time interactive avatar generation for metaverse," in *Proc. Korea Commun. Soc. Conf.* pp. 1370-1371, 2023.

Ahmad Zainudin

He received a Ph.D. degree in electronic engineering from Kumoh National Institute of Technology (KIT), Gumi, South Korea, in 2024. He pursued his B.Eng. and M.Eng. degrees in telecommunication engineering and electrical engineering from the Electronic Engineering Polytechnic Institute of Surabaya (EEPIS), Indonesia, in 2011 and 2014, respectively. Currently, he is a Lecturer in the Department of Electrical Engineering, EEPIS, and a member of the Mobile Communication and Security Research Group. Dr. Zainudin received the Excellent Academic Achievement Award from the NIIED, Ministry of Education, Republic of Korea, in 2023. He received the Grand Prize of Haedong Excellent Paper Award from the Korean Society of Communications, the Korean Institute of Communications and Information Sciences (KICS), in 2024. He received a Special Recognition Award from the IEEE R10 Asia-Pacific Graduate Student Research Paper Contest 2024. He was also a recipient of the Best Graduate Award from the Electronic Engineering Department at KIT. His research interests include intrusion detection systems, industrial IoT vulnerabilities, federated learning, blockchain, digital twins and metaverse applications.

[ORCID:0000-0001-7941-9733]

Esmot Ara Tuli

She received her Ph.D. degree in IT Convergence Engineering from Kumoh National Institute of Technology (KIT), South Korea, in 2024. She pursued her B. Sc. Eng. degree in Computer Science and

Engineering from Jatiya Kabi Kazi Nazrul Islam University, Bangladesh in 2012. Currently, she is working as a Post-doctoral Research Fellow in ICT Convergence Research Center, KIT, South Korea. She received the Outstanding Academic Paper Award in Doctoral Degree Program from the IT Convergence Engineering Department at KIT. Her major research interests include metaverse, digital twins, signal processing, and quantum computing.

[ORCID:0000-0002-7099-398X]

Dong-Seong Kim

He received his Ph.D. degree in Electrical and Computer Engineering from Seoul National University, Seoul, Korea, in 2003. From 1994 to 2003, he worked as a full-time researcher at ERC-ACI at Seoul National University. From March 2003 to February 2005, he served as a postdoctoral researcher at the Wireless Network Laboratory in the School of Electrical and Computer Engineering at Cornell University, NY. Between 2007 and 2009, he was a visiting professor in the Department of Computer Science at the University of California, Davis, CA. He served as Dean of IACF from 2019 to 2022. He is currently a Professor with the Department of IT Convergence Engineering, School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, South Korea. He is also the director of the KIT Convergence Research Institute and the ICT Convergence Research Center (ITRC and NRF advanced research center program), supported by the Korean government at Kumoh National Institute of Technology, and the director of NSLab Co. Ltd. He is a senior member of IEEE and ACM. His primary research interests include realtime IoT and smart platforms, industrial wireless control networks, networked embedded systems, Fieldbus, metaverse, and blockchain.

[ORCID:0000-0002-2977-5964]

Jae Min Lee

He received the Ph. D degree in electrical and computer engineering form the Seoul National University, Seoul, Korea, in 2005. From 2005 to 2014, he was a Senior Engineer with Samsung Electronics, Suwon, Korea. From 2015 to 2016, he was a Principle Engineer in Samsung Electronics, Suwon, Korea. Since 2017, he has been an Associate professor with School of Electronic Engineering and Department of IT-Convergence Engineering, Kumoh National Institute of Technology, Gyeongbuk, Korea. Since 2024, he has been the Director of Smart Defense Logistics Innovation Convergence Research Center (ITRC Program). He is a member of IEEE. He is the Executive Director of the Korean Institute of Communications and Information Sciences (KICS). His current main research interests are smart IoT convergence application, industrial wireless control network, UAV, Metaverse and Blockchain.
[ORCID:0000-0001-6885-5185]