

Article

DAVE: Deep Learning-Based Asymmetric Virtual Environment for Immersive Experiential Metaverse Content

Yunsik Cho ¹, Seunghyun Hong ², Mingyu Kim ³ and Jinmo Kim ^{1,2,*}¹ Department of Computer Engineering, Graduate School, Hansung University, Seoul 02876, Korea² Division of Computer Engineering, Hansung University, Seoul 02876, Korea³ Teneleven Inc., Seoul 03925, Korea; kmg2917@naver.com

* Correspondence: jinmo.kim@hansung.ac.kr; Tel.: +82-2-760-4046

Abstract: In this study, we design an interface optimized for the platform by adopting deep learning in an asymmetric virtual environment where virtual reality (VR) and augmented reality (AR) users participate together. We also propose a novel experience environment called deep learning-based asymmetric virtual environment (DAVE) for immersive experiential metaverse content. First, VR users use their real hands to intuitively interact with the virtual environment and objects. A gesture interface is designed based on deep learning to directly link gestures to actions. AR users interact with virtual scenes, objects, and VR users via a touch-based input method in a mobile platform environment. A text interface is designed using deep learning to directly link handwritten text to actions. This study aims to propose a novel asymmetric virtual environment via an intuitive, easy, and fast interactive interface design as well as to create metaverse content for an experience environment and a survey experiment. This survey experiment is conducted with users to statistically analyze and investigate user interface satisfaction, user experience, and user presence in the experience environment.



Citation: Cho, Y.; Hong, S.; Kim, M.; Kim, J. DAVE: Deep Learning-Based Asymmetric Virtual Environment for Immersive Experiential Metaverse Content. *Electronics* **2022**, *11*, 2604. <https://doi.org/10.3390/electronics11162604>

Academic Editors: Kangsoo Kim, Myungho Lee, Dongsik Jo, Gun Lee and Hanhoon Park

Received: 21 July 2022

Accepted: 16 August 2022

Published: 19 August 2022

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



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

Keywords: asymmetric virtual environment; gesture interface; text interface; immersive interaction; virtual reality; augmented reality; deep learning; immersive experiential metaverse content

1. Introduction

Virtual reality (VR) is a technology that maximizes the sense of reality by providing an immersive environment for users wearing head-mounted displays (HMD) such as the Oculus Quest and HTC Vive. Users experience a virtual environment via visual information that is completely cut off from the real world, as well as from senses such as hearing and touch. Augmented reality (AR) is a technology that synthesizes digitally produced content with the real world. Here, users experience interactions with elements that have been augmented in the real world. AR is based on reality, and it provides virtual environments and objects in the real world. Therefore, compared with VR, users can feel familiar with AR and experience it without significant inconvenience. Users can experience AR easily using smart devices such as mobile devices, and they can experience high immersion with wearable devices such as smart glasses or Microsoft HoloLens [1,2]. Expanded reality is an umbrella term, and the related technology encompasses VR, AR, and mixed reality (MR); presently, the sense of reality is expanding into a virtual environment that can provide new experiences and opportunities beyond time and space [3]. In addition to technological advancements, studies on core technologies that enable users to directly and realistically interact with a virtual environment or objects, such as user interface, haptic and motion platforms, and applied research related to the production of applications, are being conducted from various perspectives [4–8].

Relevant studies address topics such as calculating the movement of joints in the human body and reflecting it in a virtual environment, or providing haptic feedback as

physical responses from the perspective of a single user's presence in immersive VR. In addition, studies have been conducted on redirected walking, which allows a user to walk freely in a limited space, as well as motion platform technology [9–11]. These studies are being applied and developed as multi-user experience environments. Moreover, research is being conducted on collaborative interactions that can efficiently process collaborative work among users via an experience environment and system design that allows multiple HMD users to participate immersively. Based on these studies, research can be conducted on an asymmetric virtual environment in which non-immersive users who use PCs or mobile devices and AR users can participate together with VR users [12–15]. With the growing interest in the Metaverse, which is an expanded virtual world where reality and virtual reality coexist, and as interoperability and openness as key components evolve, there is a need for an experience environment that considers various platforms and a user-centered intuitive interface technology in virtual space [16]. In this study, we design a deep learning-based interface that considers the characteristics of the platform and experience environments in an asymmetric virtual environment where VR users wearing HMDs with hand tracking function and mobile platform-based AR users participate together. This approach seeks to minimize the use of additional equipment in an expanded virtual space where many users using multiple platforms participate together; to this end, we propose a highly immersive and intuitive interface via the direct expression of thoughts using body movements. The key objective is a deep learning-based asymmetric virtual environment (DAVE), a novel asymmetric experience environment which can provide all users with interesting experiences and satisfaction. The core structure of the proposed interface is designed by dividing the users' participating environment.

- For VR users, a convolutional neural network (CNN) to design a gesture interface that links real hand gestures to actions.
- For AR users, extending the modified national institute of standards and technology database (EMNIST) dataset and using a CNN to design an interface that links text to actions.

All user interfaces in the proposed DAVE infer the user's intentions and actions based on data trained with a deep learning model as a shortcut key. This approach enables users to quickly and intuitively interact with other users, the virtual environment, and objects. Because additional equipment and unnecessary graphical user interface (GUI) elements have been minimized, it is easy and convenient to use these interfaces. In order to check whether VR and AR users of DAVE are satisfied with the interface, interactive experience, and presence, we performed a comparative analysis through a survey experiment.

2. Previous Work

Virtual reality provides users with an experience similar to the real world in a computer-generated virtual environment. However, it is developing into a technology that blurs the boundary between virtual reality and reality based on various senses such as sight, hearing, and touch. This is an expanded concept known as the Metaverse, which is a compound word made up of "meta" and "verse." This concept is evolving toward presenting virtual environments in which both reality and virtual unreality are connected in society, economy, and culture [17].

Collaborative virtual environments (CVEs) are used in collaboration and interaction between several users dispersed over a long distance, and various application programs are being developed for CVEs based on shared virtual environments, including distributed simulation, 3D multi-player games, and collaborative engineering software. In addition, various studies are being conducted, ranging from studies that define the interaction between the user and environment or between users as an interactive structure, as in a distributed interactive virtual environment [18], to studies that provide an immersive environment in distributed collaboration [19]. Studies focused on the interaction between several users in an immersive virtual environment have evolved into asymmetric virtual environment research, where non-immersive (non-HMD) users using mobile devices or PCs, AR users,

and various VR users wearing HMDs can participate together. Studies that propose asymmetric interactions between HMD and PC users and between HMD users and users using the tabletop interface present a novel environment in which users experience different roles based on the user's platform and experience environment [20,21]. Gugenheimer et al. [12] proposed the ShareVR experience environment, which assigns more active actions to non-HMD users in an asymmetric virtual environment. In a role VR study, Lee et al. [14] verified that both HMD and non-HMD users could be provided with a satisfying experience and presence by presenting roles and interactions that are applicable to the user's experience environment. Furthermore, Cho et al. [15] conducted a study to design interactions and roles that are optimized for the platform in an asymmetric virtual environment where VR and AR users participate together. In addition to these studies, various studies are currently being conducted on the asymmetry of sensory feedback with respect to the user's experience environment by comparing the difference to the user between symmetrical and asymmetrical sensory feedback in a VR system that conveys multiple senses, such as visual, auditory, tactile, and olfactory feedback [22]. To implement virtual city tours based on these studies, asymmetric collaborative environments have been presented in which VR and AR users participate together [23]. In addition, asymmetric multiplayer VR games have been implemented in which HMD and non-HMD users participate together, and an analysis performed on the user's roles and perceptions in an asymmetric environment [24]. Hence, studies are required on asymmetric virtual environments that consider various experience environments and platforms to present new experiences and a satisfying presence in a virtual space via user participation. Accordingly, immersive interactions that are suitable for the user's characteristics need to be supported.

Studies are being conducted from various technical perspectives to enhance presence in virtual environments by providing highly immersive interactions to users utilizing various senses, such as sight, hearing, and touch. HMDs such as the Oculus Quest convey stereoscopic visual information in a virtual environment. In addition to HMDs, research has been conducted on sound processing, which maximizes the auditory sense of space based on a head-related transfer function (HRTF) [25–27]. Furthermore, studies have been conducted on interactions that directly utilize bodily movements using the hands, legs, etc., as well as the gaze of the eyes [15,28,29]. Based on these studies, research is being conducted on haptic systems that provide physical feedback, such as wearable equipment like 3-RSR and electric turntable haptic platforms, to implement a realistic virtual environment by reducing the gap between virtual reality and reality via high immersion [30,31]. Furthermore, as an alternative that improves the cost burden and complexity of haptic systems, studies are being conducted on pseudo-haptics that induce the illusion of receiving physical feedback via visual effects from the cognitive response according to the user's experience [6]. In addition, studies have been conducted on portable haptic systems [32], methods for directly controlling virtual objects by using a gaze pointer [28], and a hand interface that maps the controller and hand gestures [33], for the purpose of providing high accessibility and immersion in the interaction. In addition, applied research on motion platforms, such as redirected walking for walking naturally and walking in place, has been actively conducted [10,29,34]. Recently, a few studies have applied deep learning technology to VR. For example, previous research adopted deep imitation learning in a complex and elaborate process of operating a robot remotely, using VR HMD and hand tracking hardware [35]. Furthermore, another study was conducted on an interface that allows an easy and convenient way of performing the user's actions and decision-making in an intuitive structure via deep learning [33,36]. Hence, interest in convergence research is growing. However, there has been insufficient applied research on interfaces that consider various experience environments and platforms to allow users to easily and conveniently participate in asymmetric virtual environments with high immersion. Therefore, this study adopts deep learning to propose a novel user-centered interface in a DAVE experience environment.

3. DAVE: Deep Learning Based Asymmetric Virtual Environment

This study proposes an asymmetric virtual environment, DAVE, which VR and AR users can participate and experience together. The purpose of this study is to provide a novel experience environment that provides a satisfying experience and presence via an interface design which utilizes deep learning and considers the user's characteristics, such as their experience environment and platform. Figure 1 presents the structure of the proposed DAVE and the common virtual space comprising each interface that considers the participating user's characteristics (platform, device, etcetera). VR users wear an HMD with a built-in hand tracking function; hence, they design interactions using their hands and gestures. Conversely, AR users design interactions using touch screen input based on their mobile platform (shown in Supplementary Materials).

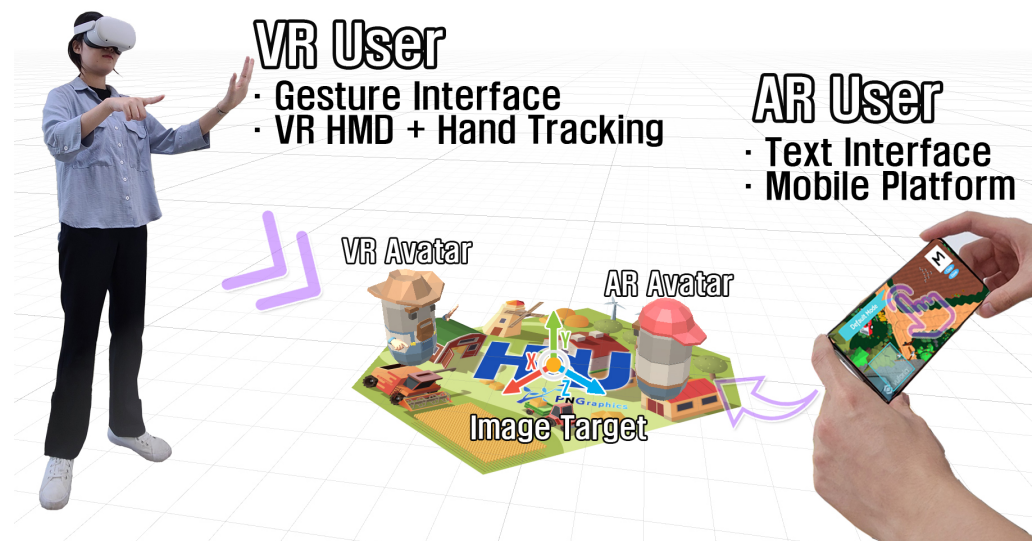


Figure 1. DAVE Configuration where VR and AR users participate together.

The goal of the proposed user interface is to utilize deep learning to interact with the virtual environment, objects, and users in an intuitive structure. In addition, the interface is designed such that users can learn and utilize the interface easily and conveniently, similar to operating shortcut keys. The main objective of this study is to provide a transition between various modes (e.g., movement, settings, communication, implement, picture, etc.) using this interface for broad participation, action, and communication in the virtual space. First, VR users use gestures, while AR users use text as the interface that corresponds to the mode based on a uniform input method using alphabets to maintain a consistent experience environment. The interface is designed to enhance learning, ease of use, and accessibility by mapping gestures to sign language and text to handwriting. Via this process, we present an experience environment that allows users to quickly and conveniently select and control desired actions, similar to shortcut keys (keys or a sets of keys) that perform a predefined function.

To implement the novel asymmetric virtual environment (DAVE) presented in this study, VR users use Oculus Quest 2 HMD and hand tracking functions [37], while AR users use a mobile device in the Vuforia development environment [38]. In addition, the integrated development environment is built in the Unity 3D engine [39], and Photon Unity Networking (PUN) is used to synchronize communication between the users and their actions.

3.1. Virtual Reality

In immersive VR, hands are often used to express the user's intentions or actions during their interactions with the virtual environment, objects, or other users. Therefore, a gesture interface is designed to allow VR users to use their hands for interactions in

the proposed asymmetric virtual environment. The proposed gesture interface comprises three stages.

1. HMD with a built-in hand tracking function (or that includes equipment for hand tracking function) is used to realistically match the real hand with the virtual hand without using an additional input equipment.
2. Gestures that directly link the virtual hand (not the traditional GUI or controller key input) to the mode transition action are defined.
3. The input gesture is recognized and inferred based on the trained gesture data, and the defined interaction is performed.

VR applications interact with virtual environments and objects by placing a GUI in a three-dimensional space and selecting a menu or by using a dedicated controller or keyboard input. However, this study proposes a gesture interface that can perform the user's intentions and actions easily and quickly without a separate GUI or additional equipment such as a controller. Considering scalability and applicability to a variety of VR systems, sign language, which conveys meaning by adopting hand gestures, is used as the gesture. Here, the gestures are defined using manual alphabets rather than fingerspelling to define specific words. This approach allows users to customize gestures easily and to make gestures easily and quickly. Figure 2 illustrates the process of making gestures with the virtual hand using the real hand based on manual alphabets [40].

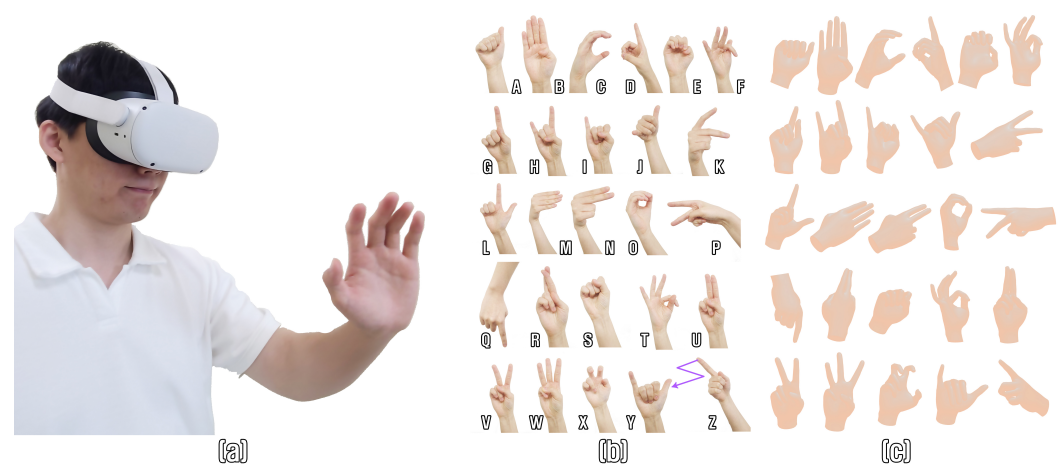


Figure 2. Gesture configuration of the VR user interface: (a) gesture is input using an HMD with a built-in hand tracking function, (b) real hand, and (c) virtual hand.

An important goal for the interaction of the proposed gesture interface is to intuitively reflect the user's desired action in the virtual environment easily and quickly. When linking the gesture to the mode transition action, this study utilizes a CNN, a deep learning model adopted by Kang et al. [33] in the DeepHandsVR interface. The CNN model used in the VR user's gesture interface has the structure and parameters of the Inception v3 neural network model, which has an efficient structure for image recognition (inference). In addition, the CNN model performs transfer learning, which involves retraining the manual alphabet gesture dataset that was photographed and classified beforehand [41]. The Inception v3 is a model that has applied BN(Batch Normalization)-auxiliary + RMSProp + Label Smoothing + Factorized 7×7 to the Inception v2 structure, which comprises 42 layers.

Figure 3 illustrates the process by which the gesture interface processes interactions via deep learning. Seven gestures and mode transition actions were first defined as the prototype based on the structure of the Inception v3 model. In addition, a dataset comprising 4,541 images was built by capturing each gesture image and augmenting them (shift, rotation, scale, etcetera). Then, training was performed for gesture classification. The trained data file was converted and generated in the input (jpeg image format) and output nodes in a format suitable for recognizing and inferring the VR user's gestures in

the Unity 3D engine development environment. Finally, the gesture made by the virtual hand corresponding to the VR user's real hand was stored as an image, and the single action was inferred by mapping the stored image to the labeled mode transition action.

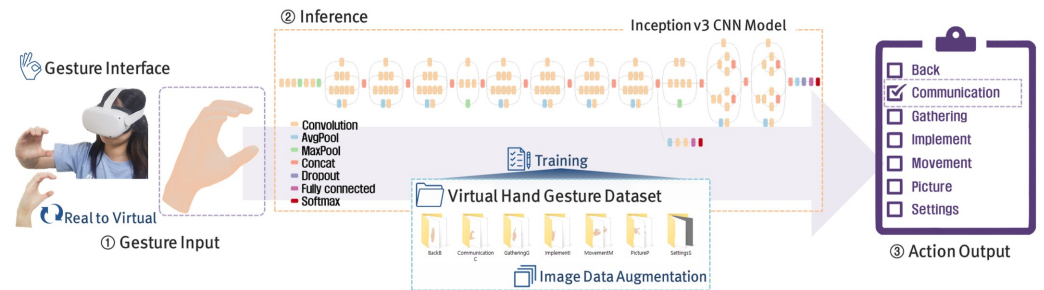


Figure 3. Deep learning model structure of the gesture interface and process of inferring mode transition action.

The CNN model used in the gesture interface adopts the softmax function as the activation function. Softmax is a function used in multi-class classification; the probability values are calculated by performing transfer learning based on the seven gesture classes. The output ($\varphi(v_i)$) of the i -th output node was calculated using Equation (1). Here, v_i represents the weighted sum of the i -th output node, while m denotes the number of output nodes.

$$\varphi(v_i) = \frac{e^{v_i}}{\sum_{j=1}^m e^{v_j}} \quad (1)$$

Each element of the softmax output is a real number greater than or equal to 0.0 and less than or equal to 1.0. In addition, the sum of the output of all nodes is always equal to 1. (Equation (2)).

$$\varphi(v_1) + \varphi(v_2) + \varphi(v_3) + \dots + \varphi(v_m) = 1 \quad (2)$$

3.2. Augmented Reality

The AR experience environment is classified into a method that involves wearing a dedicated HMD or glasses for AR/MR and a method that uses a mobile device. This study proposes a mobile platform-based experience environment and interface in order to reduce the user's equipment cost and configure an experience environment that is easy to access. Compared with the experience environment of VR users, the experience environment of AR users provides relatively low immersion. However, it has no spatial restrictions, and AR users have the advantage of being able to freely control and explore virtual scenes. Therefore, this study provides interactions with various actions and viewpoints by adopting the characteristics of the AR experience environment. The AR user interface comprises the following three stages.

1. Handwriting is input by interacting with the screen in a mobile platform environment.
2. Alphabet characters that directly link the input text to the mode transition action are defined based on the EMNIST dataset.
3. The defined interactions are performed by recognizing and inferring the handwritten input text using a CNN learning model.

The mobile platform interface is primarily based on touch screen input and interacts through input methods such as touch, drag, and swipe or through GUI elements such as buttons and menus. This study designs a text interface that interacts with virtual environments, objects, and VR users. The proposed interface performs various mode transition actions in a general structure while minimizing the existing touch input and GUI. This interface enhances the ease of learning and use by drawing the user's handwriting; in addition, it provides an experience environment and interactions consistent with use of

the alphabet input by VR users. The handwriting is drawn by connecting the coordinates of the fingertips in order with a line. Here, the line properties (width, color, etc.) are adjusted by considering the font properties of the EMNIST dataset used for recognizing handwriting [42].

The proposed text interface inputs alphabet text by drawing the user's handwriting. The text interface utilizes deep learning to recognize and infer the handwriting, then links it to various mode transition actions easily and quickly using a minimum set of GUI elements. Accordingly, this study utilizes the extended MNIST (EMNIST) to design the recognition of handwritten text input from the text interface. The EMNIST extended the MNIST (Modified National Institute of Standards and Technology database) dataset, which consists of numbers, to consider alphabets as well. The EMNIST dataset images, comprising a collection of handwritten alphanumeric characters, were converted to a 28×28 pixel image format. The training set consists of 697,932 images and the test set of 116,323 uppercase and lowercase alphabets and numbers 0–9 that map to their corresponding classes. Each image represents the intensity value of 784 pixels (28×28 grayscale pixels), ranging from 0 to 255. The test and training images have a white foreground and a black background, respectively, and consist of a combination of 62 characters: 10 numbers from 0 to 9 and 26 alphabet letters in both upper and lower cases [43,44].

In addition, the AR user interface adopts a CNN for fast and accurate interactions that link the handwritten text to mode transition actions. The user's handwriting is recognized accurately by training the CNN with the EMNIST data and via the inference made by the CNN. However, because AR users participate in the asymmetric virtual environment with their mobile devices, the CNN structure of Mor et al. [45] was adopted because it can be effectively applied to a mobile platform. This CNN structure is based on a CNN consisting of two convolutional layers and one dense layer to develop an Android handwritten text classification app for the EMNIST classification. Figure 4 illustrates the process of recognizing and inferring the handwriting input from the AR user text interface based on the information learned by training the CNN model with the EMNIST dataset. The diagram of the CNN model applied to the text interface consists of eight layers, beginning with the reshape layer that converts the handwriting image stored at 256×256 pixel resolution into a 28×28 pixel image using two convolutional layers, a Maxpooling layer, a flatten layer, dense and dropout layers, and the output layer.

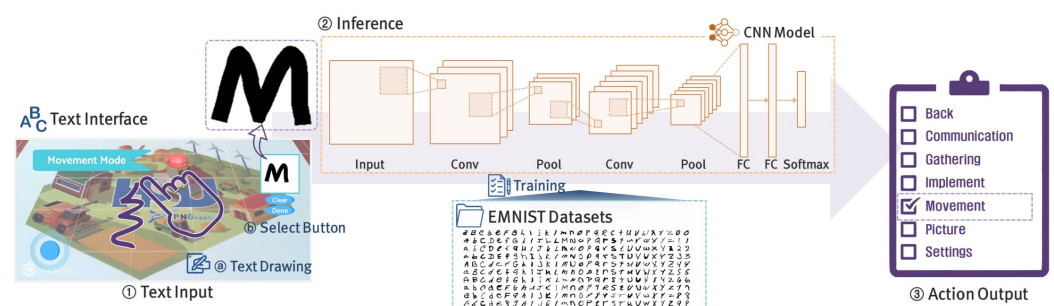


Figure 4. Text interface input method and the inference process using the deep learning model.

Similar to the gesture interface, the text interface adopts the softmax function as the activation function (Equation (1) and (2)). However, the text interface comprises 62 neurons (classes) corresponding to numbers 0–9 and letters A–Z in the upper and lower cases, as it utilizes the EMNIST database. Here, processing is required in order to infer the actions that correspond to n alphabet letters. Algorithm 1 summarizes the process of selecting and determining one of the n actions (alphabet) defined by us from the 62 classes. This determines whether the proposed action can be performed by selecting an alphabet letter with a high probability among the inferred probability values.

Algorithm 1 Process for calculating the probability of inferring the AR user's action.

1: **procedure** RECALCULATION OF PROBABILITIES FOR THE DEFINED ACTIONS

2: p_j , (for $j = 1, \dots, 3$) \leftarrow Top 3 alphabet class probabilities among $\varphi(v_i)$, (for $i = 1, \dots, 62$).

3: α_k , (for $k = 1, \dots, n$) \leftarrow Alphabet letters corresponding to n actions (uppercase and lowercase are distinguished. if there are 7 actions, $n = 2 \times 7 = 14$).

4: $p'_j = p_j * w$ \leftarrow Weight is applied if the alphabet letter same as α_k is in the j -th class.

5: $p'_x = \max(p'_x + p'_y, 1.0)$ \leftarrow Accumulate the two probabilities if the x -th alphabet letter and the y -th alphabet letter are the same, except the letter cases are different (the probability should not exceed 1.0).

6: Update the calculated p'_j value in $\varphi(v_i)$.

7: **end procedure**

Similar to VR interaction, the AR interaction defines mode transition actions that correspond to seven texts. If the user inputs one of the defined alphabet text via handwriting, the CNN infers the handwriting. Then, one of the mode transition actions corresponding to the input text is directly performed without interacting with a separate GUI.

3.3. Metaverse Experience Environment

An asymmetric virtual environment in which all users participate together is built based on the VR user's gesture interface and the AR user's text interface. All users have independent roles in the same virtual space. They can perform actions such as collaboration and exchange to achieve common goals. To present an experience environment where both VR and AR users can perform various roles and actions using an input method with a minimum set of GUI in the asymmetric virtual environment of this study, we design a structure in which the proposed interface controls the actions that effectively changes the mode (e.g., movement, settings, communication, gathering, implement, picture, etcetera). The key is to maintain a consistent interface with a uniform input pattern using the alphabet, regardless of the platform used by the users participating in the asymmetric virtual environment. This concept is similar to that of shortcut keys used for quick commands in many software applications. Although users need to learn the shortcut keys, they are convenient and easy to use when the users become familiar with them. This characteristic of shortcut keys was applied to the proposed interface.

This study adopts a virtual farm as the background and classifies VR users and AR users by considering the possible activities and roles on the farm as an example. In addition, the same seven types of mode transition actions are defined for all users. Figure 5 presents the process by which each user performs the seven mode transition actions using the proposed interface. A gesture made by a VR user and a text input by an AR user are recognized and inferred via deep learning based on the seven defined gestures for VR users and the seven defined texts for AR users. Then, the mode transition action corresponding to the recognition and inference result is performed. The defined mode transition actions are communication, settings, back, movement, gathering, implement, and picture. The alphabet letters *C*, *S*, *B*, *M*, *G*, *I*, and *P* are adopted as the input for these actions. The seven mode transition actions are described as follows.

1. **Communication:** A chat mode for users. Users can communicate with each other by exchanging messages preset using a macro function.
2. **Settings:** In this mode, users can change settings for the virtual environment.
3. **Back:** Cancel or return mode. Users can cancel the current input or action or return to the previous state.

4. **Movement:** In movement mode, VR users move through portals, and AR users create 3D avatars and move them with a virtual joystick.
5. **Gathering:** In gathering mode, users can collect objects in the virtual environment. VR users use their hands to grab objects, and AR users select objects by touching the screen.
6. **Implement:** A tool operation mode for manipulating and controlling virtual objects. Again, VR users interact using their hands and AR users interact by touching the screen.
7. **Picture:** A sharing mode in which users can share the scene they see with other users.

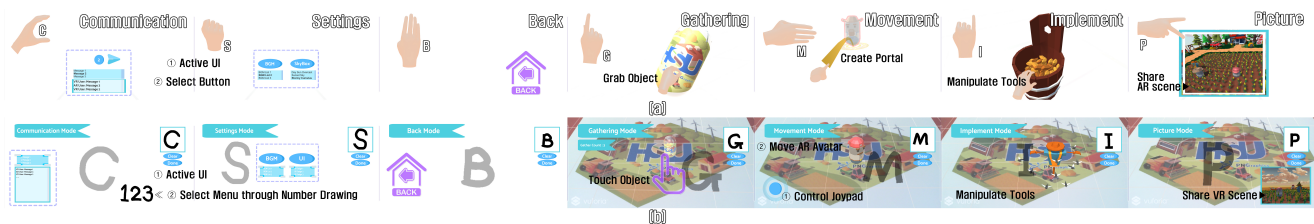


Figure 5. Process by which VR and AR users perform the mode transition action based on the proposed interface via deep learning: (a) VR user and (b) AR user.

The inferring of the seven types of actions and the alphabet letters corresponding to them is determined according to the highest probability value calculated using Equation (1). Here, the number of output nodes (m) refers to the seven types of action. However, for AR users the probability values are updated using Algorithm 1, based on the seven actions and the alphabet texts corresponding to these actions from the 62 classes. Through this process, a single action is selected. If the calculated probability value is less than 80%, the input gesture or text is considered inaccurate, and the mode transition action is not performed.

Figure 6 presents the result of creating the proposed DAVE experience environment via the process illustrated in Figure 5. In this study, commercial graphic resources were used for the 2D GUI and 3D objects required for content creation of the metaverse experience environment [46–48]. This example illustrates the process of experiencing the virtual farm via the adoption of various mode transition actions. The important point is that this process is not limited to the seven types of actions and alphabet letters presented in Figure 5, including the corresponding gestures (VR users) and texts (AR users). However, it is scalable, such that it can be modified or customized in various ways according to the application field and application methods.



Figure 6. Results of creating an asymmetric experience environment with a virtual farm background: (a) Results of creating the virtual environment and the experience process, (b) VR user, (c) AR user.

4. Experimental Results and Analysis

The deep learning model applied to the proposed interface was implemented using Anaconda 3, conda 4.6.12, and TensorFlow 3.6.8. The experiment for the training model was implemented in the Unity 3D engine using the TensorFlowSharp 1.15.1 plug-in [49]. In addition, the DAVE experience environment, in which VR and AR users participate together, as well as the content, were developed using Unity 2019.2.3f1(64bit), Oculus SDK, and Vuforia 9.7. VR users used Oculus Quest 2 HMDs to interact, and AR users used Samsung Galaxy S20+ mobile devices to interact. Finally, the PC used for the integrated development environment and experiment was equipped with an Intel Core i7-6700, 16GB RAM, and a GeForce 1080 GPU. Figure 7 presents the process in which VR and AR users participate in the DAVE experience environment. The environment was configured with sufficient space ($1.5\text{ m} \times 1.5\text{ m}$) for a VR user to experience DAVE while comfortably sitting on a chair or standing. For AR users, the environment was configured with a roomy space ($2\text{ m} \times 2\text{ m}$) for the user to freely control virtual objects on the image target.

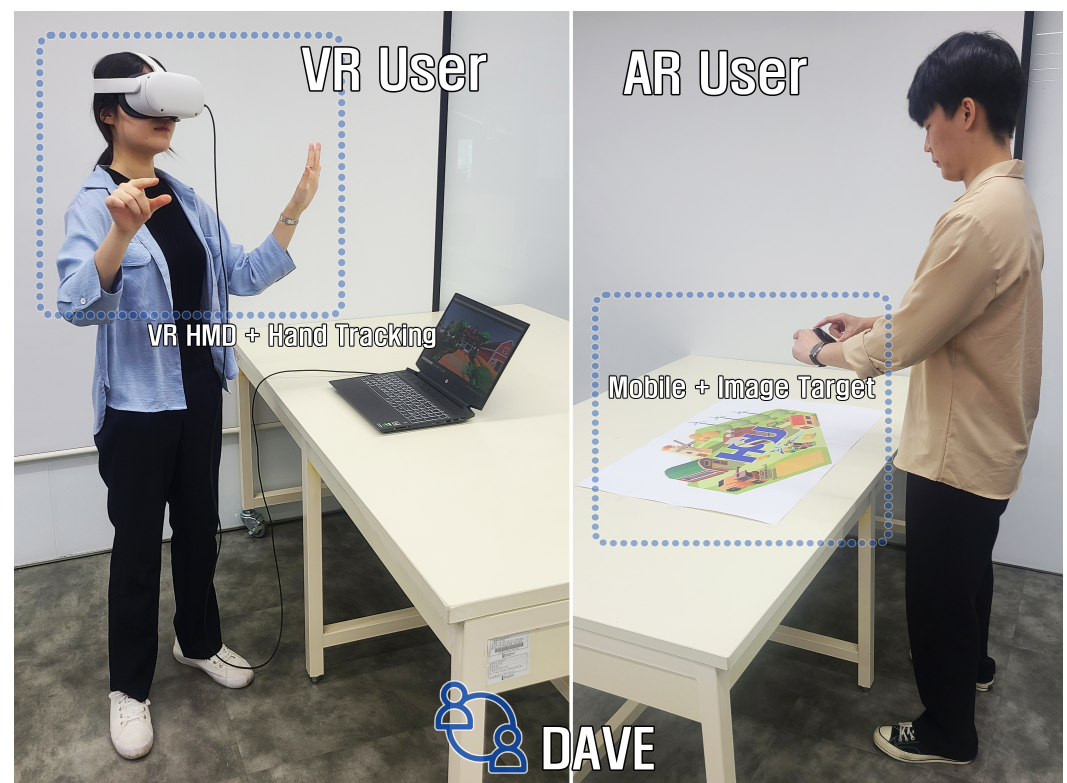


Figure 7. Configuration of the DAVE experience environment where VR and AR users participate.

A survey experiment was conducted with users in the experience environment shown in Figure 7 to comparatively analyze the user experience of the proposed interface and presence experienced by the user. The participants consisted of ordinary people who have had experience with interactive content, including VR and AR, as well as computer science students and researchers with development experience. The participants understand the process of interacting with content using traditional GUI methods, and they had the knowledge and experience to compare their experiences of the proposed interface. The purpose of the survey comprised three main parts. First, we checked whether the new type of interface optimized for the platforms and utilizing deep learning is easy to learn, easy to use, and satisfying to the user. In addition to overall satisfaction, we checked whether it provides a satisfying experience for all users with each action, role, and interaction considering the differences in the experience methods and platforms between VR and AR users. Finally, we compared whether the AR user and VR users were both able to experience enhanced presence in the immersive experience environment. The survey

participants experienced the virtual farm content once in VR and once in AR, and the order of experience was arranged such that ten participants experienced VR first while the other ten participants experienced AR first. This was intended to reduce any difference in the survey results owing to the order of experience.

The first survey experiment compared and analyzed the satisfaction level of the VR and AR users with their respective interfaces. The participants in the first survey comprised twelve people (male: eight, female: four) between the ages of 22 and 39. Interactions via the proposed interface provided users with a convenient experience by allowing them to directly interact with the virtual environment or objects with a minimum GUI set and without additional equipment. Accordingly, values were recorded on a 7-point scale based on the four categories, including 30 questions from A. Lund's USE (Usefulness, Satisfaction, and Ease of use) questionnaire [50]. Table 1 presents statistical data based on the survey results. It was confirmed that the proposed gesture and text interface showed improved results compared to the traditional GUI method in usefulness, ease of use, and overall satisfaction. However, in the case of ease of learning, most of the survey participants took time and time to learn a new structure to perform actions in a new way, such as taking a gesture or entering text, while they familiar with the traditional GUI. In addition to this extra required effort, the proposed interface showed relatively low results. Nevertheless, the results when calculating statistical significance using one-way analysis of variance (ANOVA) confirmed that there were no meaningful differences in any factors or cases. Although the proposed interface using deep learning for each user affected the overall improvement in satisfaction compared to the traditional GUI, it did not have a significant difference. However, it was judged that even a sufficiently significant difference could be induced if it was improved in such a way that it went through a sufficient adaptation process, such as shortcut keys, or was mixed with the traditional GUI.

Table 1. Comparative analysis result for satisfaction between the proposed interface using deep learning and the traditional GUI method.

	VR		AR	
	Traditional GUI	Gesture Interface	Traditional GUI	Text Interface
Mean (SD)				
usefulness	5.813 (1.093)	6.021 (1.128)	5.802 (0.844)	6.156 (0.271)
ease of use	5.705 (1.054)	5.712 (1.342)	6.129 (0.770)	6.235 (0.445)
ease of learning	6.146 (0.681)	5.854 (1.431)	6.417 (0.562)	6.396 (0.710)
satisfaction	5.774 (1.223)	6.274 (1.093)	5.929 (1.095)	6.464 (0.535)
Pairwise Comparison				
usefulness	F (1, 22) = 0.195, $p = 0.664$		F (1, 22) = 1.757, $p = 0.199$	
ease of use	F (1, 22) = 0.0002, $p = 0.988$		F (1, 22) = 0.157, $p = 0.696$	
ease of learning	F (1, 22) = 0.373, $p = 0.548$		F (1, 22) = 0.006, $p = 0.940$	
satisfaction	F (1, 22) = 1.022, $p = 0.323$		F (1, 22) = 2.126, $p = 0.159$	
VR gesture interface vs AR text interface				
usefulness	F (1, 22) = 0.150, $p = 0.702$			
ease of use	F (1, 22) = 1.504, $p = 0.233$			
ease of learning	F (1, 22) = 1.265, $p = 0.273$			
satisfaction	F (1, 22) = 0.296, $p = 0.609$			

The second survey experiment analyzed presence based on immersion. The participants in the second and last survey comprised twenty people (male: eleven, female: nine) between the ages of 20 and 53. By analyzing the characteristics of the user's experience environment and appropriately designing roles, actions, as well as through the communication between users that are suitable for these characteristics, this study assumed that all

users would be satisfied with the presence they experienced. Therefore, a survey experiment was conducted to verify this assumption. The survey participants recorded values for nineteen questions of the presence questionnaire by Witmer et al. [51] on a 7-point scale. Table 2 presents the comparative analysis result of the detailed categories based on the recorded values. Generally, the overall satisfaction level of the VR users was high because they interacted directly in the virtual space. However, differences were found in detailed categories according to the experience environment. In the VR case, realism and the possibility of examination, which is related to manipulation, were high because the VR users interacted with the virtual environment and objects using their hands without additional equipment such as a controller. For AR, users could easily adapt to the interface, and their satisfaction with the interface was high. Owing to these characteristics, it was easy to predict actions or adapt to situations. Consequently, high values were recorded for the possibility of acting, quality of the interface, and self-evaluation of performance. It was confirmed that the presence results, which the AR and VR users were both satisfied with, could be improved if the experience environment was designed based on an interface suitable for the experience environment and platform. In addition, the result obtained when calculating statistical significance using one-way ANOVA confirmed that there were no meaningful differences in presence for all categories.

Table 2. Comparative analysis result of presence experienced by each user in asymmetric experience environment (DAVE).

	VR	AR
Mean (SD)		
total	6.132 (0.390)	6.092 (0.471)
realism	6.143 (0.593)	6.050 (0.545)
possibility to act	5.900 (0.718)	5.988 (0.731)
quality of interface	6.300 (0.526)	6.433 (0.507)
possibility to examine	6.400 (0.663)	6.067 (0.478)
self-evaluation of performance	5.900 (0.718)	5.975 (0.798)
Pairwise Comparison		
total	F (1, 38) = 0.059, $p = 0.809$	
realism	F (1, 38) = 0.253, $p = 0.618$	
possibility to act	F (1, 38) = 0.139, $p = 0.712$	
quality of interface	F (1, 38) = 0.633, $p = 0.431$	
possibility to examine	F (1, 38) = 3.156, $p = 0.084$	
self-evaluation of performance	F (1, 38) = 0.093, $p = 0.762$	

The last survey experiment was the analysis result of social presence. This experiment analyzed users' presence and experience via interactions and social relationships established when VR and AR users participated in the virtual environment together using the DAVE experience environment based on the proposed interface. Three factors (behavioral interaction, mutual assistance, dependent action) of behavioral engagement among the dimensions of the Networked Minds Measure of Social Presence [52] questionnaire were utilized to check the behavioral experience between users. The values of the social presence questionnaire were recorded on a 7-point scale, and statistical analysis was performed based on the result values. Table 3 presents the result of the social presence questionnaire. VR and AR users performed predefined actions using their respective interfaces. During this process, they had social relationships, such as cooperation and sharing. However, differences may exist in the shared information or the results of actions, such as collaboration and dependence, owing to differences in their experience environments. The experience environment proposed in this study is based on collaboration between VR and AR users. Hence, it can be confirmed that they affect each other during the interaction process. In

the VR case, relatively high values were recorded for mutual assistance because VR users performed their actions based on help from AR users, because AR users can observe the overall situation and make decisions based on it. For AR, high values were recorded for dependent action owing to the ability of VR users to change scenes and scenarios by interacting directly with the virtual environment and objects, while AR users performed actions depending on these changes. The results obtained when calculating statistical significance using one-way ANOVA confirmed that there were no significant differences in any factors.

Table 3. Analysis results of social presence via the interactions between users in the asymmetric experiment environment (DAVE).

	VR	AR
Mean (SD)		
behavioral interaction	4.508 (0.777)	4.500 (0.850)
mutual assistance	4.700 (0.897)	4.675 (0.926)
dependent action	3.975 (1.978)	4.050 (2.018)
Pairwise Comparison		
behavioral interaction	F (1, 38) = 0.001, $p = 0.975$	
mutual assistance	F (1, 38) = 0.007, $p = 0.933$	
dependent action	F (1, 38) = 0.013, $p = 0.909$	

The last experiment confirmed the accuracy of gesture or text inference in the process of performing seven actions through an interface using deep learning. For this purpose, we recorded whether accurate actions corresponding to the user's intended input (gesture or text) were inferred at the same time when the user experienced the content in the process of the survey experiment. Table 4 shows the results from calculating the interface accuracy of VR and AR users, respectively, along with the corresponding actions (alphabet B, C, G, I, M, P, and S) and inferred probability values.

Table 4. Accuracy analysis result in the action inference process (unit: %).

	C	S	B	G	M	I	P
VR user's gesture interface							
basic accuracy	100	100	90.7	97.8	97.9	98.6	99.5
threshold (10%)	100	100	97.7	99.4	98.8	99.7	100
AR user's text interface							
basic accuracy	72.7	66.0	92.0	85.3	89.3	66.7	97.1
recalculation (Algorithm 1)	97.3	99.3	98.0	96.7	91.3	94.0	98.7

Regarding the results of the accuracy analysis, first, the basic accuracy was high because the VR users implemented only the seven gestures corresponding to the defined actions through the manual alphabet. In the gesture image inference process of the gesture interface, a threshold value was applied to the difference between the two gestures (alphabetic) with the highest inferred value. This study implemented a filtering process when a similar or ambiguous gesture was input by applying the process of recalculating the inference when the difference between the upper two inference values was less than the threshold. As a result of the experiment, it was confirmed that improved accuracy results were obtained when the threshold was applied at 10%. It can be seen that AR users show lower accuracy compared to VR users, as they infer using all of the alphabet collections (62 characters: 10 numbers from 0 to 9 and 26 alphabet letters both in upper and lower cases) provided by EMNIST. For example, in the case of I, the handwritten text I is recognized as the number 1 or the lowercase text l (L) of the alphabet, which causes a problem in that the incorrect answer rate increases. Therefore, we tried to solve this problem

through the probability recalculation procedure of Algorithm 1. As a result of calculating the accuracy by applying Algorithm 1, it was confirmed that similar or ambiguous hand-written texts showed improved results. If these points are comprehensively considered and learning is conducted by newly combining input (gesture and text) suitable for the content to be produced, we judge that the inconvenience of the interface due to accuracy can be prevented.

5. Limitations and Discussion

A few limitations or inconvenient factors exist in the interactions made by VR and AR users who participate in the asymmetric virtual environment. For example, VR users wearing the Oculus Quest HMD can use their hands to control the virtual environment or objects directly; however, they cannot enter keys because they do not have any input device, such as a controller. They are limited in performing actions and altering various modes as well. For mobile platform AR users, providing various functions complicates the GUI. This could be a factor that causes inconveniences to AR users, who quickly and easily interact in the experience environment where multiple users participate. Accordingly, further research is required on a novel interface that considers characteristics such as the user's experience environment, platform, and equipment in an asymmetric virtual environment. Nevertheless, there are limitations or factors that need to be improved in the proposed interface in terms of its effectiveness. For example, a difficult or unfamiliar gesture can require separate training, as can cases in which each user has different handwriting in the handwriting process. Therefore, in order to address these limitations, we intend to improve the deep learning-based interface technology, which includes combining the gesture and text functions with voice.

The proposed interfaces and interactions can be adopted in various industries, such as games and education for immersive experiential metaverse content. Ultimately, we attempt to present a novel direction through which these interfaces and interactions can be applied as the interface technique in the development of the metaverse technology, as well as in a new experience environment where users of various platforms participate. Therefore, an example of metaverse content was created for a purpose other than the virtual farm experience environment which comprises the seven mode transition actions mentioned above. Figure 8 presents the results obtained from the content we created, consisting of five types of actions: communication (C), description (D), selection (S), movement (M), and back (B). The content allows users to cooperate with each other to achieve educational goals on the theme of eco-friendly life education (saving electricity, recycling, etcetera). The interface operates similarly to the previous experience environment; only the alphabet letters used for gestures and handwritten texts were changed.



Figure 8. Results obtained from creating educational content by applying the DAVE experience environment: (a) content scene, (b) VR user, (c) AR user.

In addition, the primary purpose of the proposed DAVE is to build an asymmetric experience environment for several users, not just one VR user and one AR user, as well as for various platforms such as PCs and mobile devices. Therefore, the breadth of user participation, roles, and communication between users should be expanded to an experience environment for several users. Furthermore, in addition to the content created for the experiment, other content can be created via various plans to verify that DAVE can be

applied to various fields, such as manufacturing and medical service. We intend to conduct comparative analysis studies on the experience environment and user experience using existing studies, such as existing social VR platforms, that enable flexible user participation through VR HMD, mobile/tablet, and PC [53]. In this way, DAVE can be developed to present a new user-centered experience environment.

6. Conclusions

In this study, a novel deep learning-based interface was designed by considering characteristics such as the user's platform and experience environment in an asymmetric virtual environment where VR and AR users participate together. We proposed a novel experience environment called DAVE which can provide an experience and presence that satisfies all users. For VR users wearing an HMD with a hand tracking function, we designed a gesture interface that directly links real hand gestures to various mode transition actions via virtual hand gestures in order to interact with virtual environments, objects, and users. Regarding AR, a virtual environment was superimposed on an actual image captured by a mobile camera. An AR user exploring and experiencing the general virtual environment can interact with the virtual environment and other users as well as perform various mode transition actions via the text interface. The text interface draws the handwritten text based on the touch-based input method of the mobile platform. We adopted deep learning in order to directly link the text input to action. Using this process, all users can perform interactions and roles suitable for their experience environment. Accordingly, they are provided with a satisfying interface, experience, and presence. To analyze these, a survey was conducted to compare the VR and AR experience environments, and the obtained results verified that all users were satisfied. Consequently, if users can directly learn and adapt to the proposed interface as a shortcut key, it is expected that the novel experience environment would be able to provide satisfying experiences and interactions.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/electronics11162604/s1>.

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

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1F1A1063442).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Lhachemi, H.; Malik, A.; Shorten, R. Augmented Reality, Cyber-Physical Systems, and Feedback Control for Additive Manufacturing: A Review. *IEEE Access* **2019**, *7*, 50119–50135.
2. Pellas, N.; Mystakidis, S.; Kazanidis, I. Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature. *Virt. Real.* **2021**, *25*, 835–861.
3. Reiners, D.; Davahli, M.R.; Karwowski, W.; Cruz-Neira, C. The Combination of Artificial Intelligence and Extended Reality: A Systematic Review. *Front. Virt. Real.* **2021**, *2*, 1–13. <https://doi.org/10.3389/frvir.2021.721933>.
4. Sebastian, M.; Maximilian, B.; Lukas, W.; Lung-Pan, C.; Floyd, M.F.; Patrick, B. VirtualSpace—Overloading Physical Space with Multiple Virtual Reality Users. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 241:1–241:10. [[CrossRef](#)]
5. Sharma, V.S.; Mehra, R.; Kaulgud, V.; Podder, S. An extended reality approach for creating immersive software project workspaces. In Proceedings of the 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE '19), Montreal, QC, Canada, 27 May 2019; pp. 27–30.

6. Kim, M.; Kim, J.; Jeong, K.; Kim, C. Grasping VR: Presence of Pseudo-Haptic Interface Based Portable Hand Grip System in Immersive Virtual Reality. *Int. J. Hum. Comput. Interact.* **2020**, *36*, 685–698. <https://doi.org/10.1080/10447318.2019.1680920>.
7. Aoki, H. Study on pseudo-haptics during swimming motion in a virtual reality space. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Utrecht, The Netherlands, 14–18 December 2020; pp. 333–336.
8. Gong, L.; Fast-Berglund, Å.; Johansson, B. A framework for extended reality system development in manufacturing. *IEEE Access* **2021**, *9*, 24796–24813.
9. Choi, I.; Ofek, E.; Benko, H.; Sinclair, M.; Holz, C. *Claw: A Multifunctional Handheld Haptic Controller for Grasping, Touching, and Triggering in Virtual Reality*, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Montreal QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 654:1–654:13. <http://doi.acm.org/10.1145/3173574.3174228>.
10. Thomas, J.; Rosenberg, E.S. Reactive alignment of virtual and physical environments using redirected walking. In Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; pp. 317–323.
11. Kammerlander, R.K.; Pereira, A.; Alexanderson, S. Using Virtual Reality to Support Acting in Motion Capture with Differently Scaled Characters. In Proceedings of 2021 IEEE Virtual Reality and 3D User Interfaces (VR), Lisboa, Portugal, 27 March–1 April 2021; pp. 402–410.
12. Gugenheimer, J.; Stemasov, E.; Frommel, J.; Rukzio, E. *ShareVR: Enabling Co-Located Experiences for Virtual Reality Between HMD and Non-HMD Users*, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), Denver, CO, USA, 6–11 May 2017; ACM: New York, NY, USA, 2017; pp. 4021–4033. <https://doi.org/10.1145/3025453.3025683>.
13. Grandi, J.G.; Debarba, H.G.; Maciel, A. Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 127–135.
14. Lee, J.; Jeong, K.; Kim, J. RoleVR: Multi-experience in immersive virtual reality between co-located HMD and non-HMD users. *Multim. Tools Appl.* **2020**, *79*, 979–1005. <https://doi.org/10.1007/s11042-019-08220-w>.
15. Yunsik, C.; Jiewon, K.; Jaekyung, J.; Jongchan, P.; Mingyu, K.; Jinmo, K. X-person asymmetric interaction in virtual and augmented realities. *Comput. Animat. Virt. Worlds* **2020**, *32*, e1985. <https://doi.org/10.1002/cav.1985>.
16. Mystakidis, S. Metaverse. *Encyclopedia* **2022**, *2*, 486–497. <https://doi.org/10.3390/encyclopedia2010031>.
17. Sweeney, T. *Foundational Principles & Technologies for the Metaverse*, Proceedings of the ACM SIGGRAPH 2019 Talks (SIGGRAPH '19), Los Angeles, CA, USA, 28 July–1 August 2019; ACM: New York, NY, USA, 2019; p. 38:1. <https://doi.org/10.1145/3306307.3339844>.
18. Carlsson, C.; Hagsand, O. Dive a multi-user virtual reality system. In Proceedings of the IEEE Virtual Reality Annual International Symposium, Seattle, WA, USA, 18–22 September 1993; pp. 394–400.
19. Le, K.; Fjeld, M.; Alavi, A.; Kunz, A. *Immersive Environment for Distributed Creative Collaboration*, Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17), Gothenburg, Sweden, 8–10 November 2017; ACM: New York, NY, USA, 2017; pp. 16:1–16:4. <http://doi.acm.org/10.1145/3139131.3139163>.
20. Duval, T.; Fleury, C. *An Asymmetric 2d Pointer/3d RAY for 3d Interaction within Collaborative Virtual Environments*, Proceedings of the 14th International Conference on 3D Web Technology (Web3D '09), Darmstadt, Germany, 16–17 June 2009; ACM: New York, NY, USA, 2009; pp. 33–41. <https://doi.org/10.1145/1559764.1559769>.
21. Ibayashi, H.; Sugiura, Y.; Sakamoto, D.; Miyata, N.; Tada, M.; Okuma, T.; Kurata, T.; Mochimaru, M.; Igarashi, T. *Doll-house VR: A Multi-View, Multi-User Collaborative Design Workspace with VR Technology*, Proceedings of the SIGGRAPH Asia 2015 Emerging Technologies (SA '15), Kobe, Japan, 2–6 November 2015; ACM: New York, NY, USA, 2015; pp. 8:1–8:2. <http://doi.acm.org/10.1145/2818466.2818480>.
22. Jung, S.; Karki, N.; Slutter, M.; Lindeman, R.W. On the use of multi-sensory cues in symmetric and asymmetric shared collaborative virtual spaces. *PROC. ACM Hum. Comput. Interact.* **2021**, *5*, 1–25. <https://doi.org/10.1145/3449146>.
23. Feld, N.; Weyers, B. Mixed reality in asymmetric collaborative environments: A research prototype for virtual city tours. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VR Workshops 2021, Lisbon, Portugal, 27 March–1 April 2021; pp. 250–256. <https://doi.org/10.1109/VRW52623.2021.00053>.
24. Karaosmanoglu, S.; Rogers, K.; Wolf, D.; Rukzio, E.; Steinicke, F.; Nacke, L.E. *Feels Like Team Spirit: Biometric and Strategic Interdependence in Asymmetric Multiplayer VR Games*, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 8–13 May 2021; ACM: New York, NY, USA, 2021; pp. 443:1–443:15. <https://doi.org/10.1145/3411764.3445492>.
25. Fitzpatrick, W.; Wickert, M.; Semwal, S. 3d sound imaging with head tracking. In Proceedings of the 2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), Napa, CA, USA, 11–14 August 2013; pp. 216–221.
26. Carl, S.; Aaron, N.; Ravish, M. Efficient HRTF-based Spatial Audio for Area and Volumetric Sources. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 1356–1366. <https://doi.org/10.1109/TVCG.2016.2518134>.
27. Hong, D.; Lee, T.; Joo, Y.; Park, W. *Real-Time Sound Propagation Hardware Accelerator for Immersive Virtual Reality 3d Audio*, Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '17), San Francisco, CA, USA, 25–27 February 2017; ACM: New York, NY, USA, 2017; pp. 20:1–20:2. <https://doi.org/10.1145/3023368.3036842>.
28. Sidorakis, N.; Koulouris, G.A.; Mania, K. Binocular eye-tracking for the control of a 3D immersive multimedia user interface. In Proceedings of the 2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR), Arles, France, 23 March 2015; pp. 15–18.

29. Lee, J.; Jeong, K.; Kim, J. MAVe: Maze-based immersive virtual environment for new presence and experience. *Comput. Anim. Virt. Worlds* **2017**, *28*, e1756. <https://doi.org/10.1002/cav.1756>.
30. Leonardis, D.; Solazzi, M.; Bortone, I.; Frisoli, A. A 3-RSR Haptic Wearable Device for Rendering Fingertip Contact Forces. *IEEE Trans. Haptics* **2017**, *10*, 305–316. <https://doi.org/10.1109/TOH.2016.2640291>.
31. Huang, H.-Y.; Ning, C.-W.; Wang, P.-Y.; Cheng, J.-H.; Cheng, L.-P. *Haptic-Go-Round: A Surrounding Platform for Encounter-Type Haptics in Virtual Reality Experiences, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA, 25–30 April 2020*; ACM: New York, NY, USA, 2020; pp. 1–10. <https://doi.org/10.1145/3313831.3376476>.
32. Kim, M.; Jeon, C.; Kim, J. A Study on Immersion and Presence of a Portable Hand Haptic System for Immersive Virtual Reality. *Sensors* **2017**, *17*, 1141. <https://doi.org/10.3390/s17051141>.
33. Kang, T.; Chae, M.; Seo, E.; Kim, M.; Kim, J. DeepHandsVR: Hand Interface Using Deep Learning in Immersive Virtual Reality. *Electronics* **2020**, *9*, 1863.
34. Schroeder, A.; Goodnight, E.; Kühner, M.; Gerner, W.; Hebel, M.; Görlich, D. *Infinite Walking in Three Dimensions in Virtual Reality: A Shopping Mall Simulator Game, Proceedings of the Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20), Virtual Event, Canada, 2–4 November 2020*; ACM: New York, NY, USA, 2020; pp. 76–79. <https://doi.org/10.1145/3383668.3419918>.
35. Zhang, T.; McCarthy, Z.; Jow, O.; Lee, D.; Goldberg, K.; Abbeel, P. Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 1–8.
36. Kim, J. VIVR: Presence of Immersive Interaction for Visual Impairment Virtual Reality. *IEEE Access* **2020**, *8*, 196151–196159.
37. Facebook Technologies, LLC. Oculus. 2021. Available online: <https://www.oculus.com/> (accessed on 15 March 2021).
38. PTC, Inc. Vuforia Engine. 2021. Available online: <https://developer.vuforia.com/> (accessed on 3 Feb 2021).
39. Unity Technologies. Unity Engine. 2019. Available online: <https://unity3d.com/> (accessed on 28 August 2019).
40. Nisar, S.; Khan, A.A.; Javed, M.Y. A Statistical Feature Based Decision Tree Approach for Hand Gesture Recognition, *Proceedings of the 7th International Conference on Frontiers of Information Technology (FIT '09), Abbottabad, Pakistan, 16–18 December 2009*; ACM: New York, NY, USA, 2009; pp. 27:1–27:6. <https://doi.org/10.1145/1838002.1838033>.
41. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. <https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.308>.
42. Cho, Y.; Kim, J. Production of mobile english language teaching application based on text interface using deep learning. *Electronics* **2021**, *10*, 1809. <https://doi.org/10.3390/electronics10151809>.
43. Gregory, C.; Saeed, A.; Jonathan, T.; van Schaik, A. EMNIST: Extending MNIST to handwritten letters. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 2921–2926.
44. Alejandro, B.; Yago, S.; Pedro, I. A Survey of Handwritten Character Recognition with MNIST and EMNIST. *Appl. Sci.* **2019**, *9*, 3169.
45. Mor, S.S.; Solanki, S.; Gupta, S.; Dhingra, S.; Jain, M.; Saxena, R. Handwritten text recognition: With deep learning and android. *Int. J. Eng. Adv. Technol.* **2019**, *8*, 819–825.
46. Innovana Games. My Farm-Low Poly Farm Pack. 2020. Available online: <https://assetstore.unity.com/packages/3d/environments/my-farm-low-poly-farm-pack-162147> (accessed on 23 June 2020).
47. A.R.S|T. Low Poly Men. 2020. Available online: <https://assetstore.unity.com/packages/3d/characters/low-poly-men-160038> (accessed on 1 September 2020).
48. Freepik Company. Vector Icons Resources. 2021. Available online: <https://www.flaticon.com/> (accessed on 10 August 2021).
49. Google Brain. TensorFlow. 2021. Available online: <https://www.tensorflow.org/> (accessed on 28 July 2021).
50. Lund, M.A. Measuring Usability with the USE Questionnaire. In *Usability and User Experience Newsletter of the STC Usability SIG*; Society for Technical Communication: Fairfax, VI, USA, 2001; Volume 8, pp. 3–6.
51. Witmer, B.G.; Jerome, C.J.; Singer, M.J. The Factor Structure of the Presence Questionnaire. *Presence Teleoper. Virt. Environ.* **2005**, *14*, 298–312.
52. Biocca, F.; Harms, C.; Gregg, J. The networked minds measure of social presence : Pilot test of the factor structure and concurrent validity. In Proceedings of the 4th Annual International Workshop on Presence, Philadelphia, PA, USA, 21–23 May 2001; pp. 1–9.
53. Eiris, R.; Wen, J.; Gheisari, M. iVisit-Collaborate: Collaborative problem-solving in multiuser 360-degree panoramic site visits. *Comput. Educ.* **2022**, *177*, 104365. <https://doi.org/10.1016/j.compedu.2021.104365>.