electronics

MDPI

Article

DeepHandsVR: Hand Interface Using Deep Learning in Immersive Virtual Reality

Taeseok Kang [1], Minsu Chae [1], Eunbin Seo [1], Mingyu Kim [2] and Jinmo Kim [1,*]

[1] Division of Computer Engineering, Hansung University, Seoul 02876, Korea; goxotjr@naver.com (T.K.); alstn328@naver.com (M.C.); sebbin99@naver.com (E.S.)

[2] Program in Visual Information Processing, Korea University, Seoul 02841, Korea; kmg2917@naver.com

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

Received: 28 September 2020; Accepted: 4 November 2020; Published: 6 November 2020

Abstract: This paper proposes a hand interface through a novel deep learning that provides easy and realistic interactions with hands in immersive virtual reality. The proposed interface is designed to provide a real-to-virtual direct hand interface using a controller to map a real hand gesture to a virtual hand in an easy and simple structure. In addition, a gesture-to-action interface that expresses the process of gesture to action in real-time without the necessity of a graphical user interface (GUI) used in existing interactive applications is proposed. This interface uses the method of applying image classification training process of capturing a 3D virtual hand gesture model as a 2D image using a deep learning model, convolutional neural network (CNN). The key objective of this process is to provide users with intuitive and realistic interactions that feature convenient operation in immersive virtual reality. To achieve this, an application that can compare and analyze the proposed interface and the existing GUI was developed. Next, a survey experiment was conducted to statistically analyze and evaluate the positive effects on the sense of presence through user satisfaction with the interface experience.

Keywords: hand interface; immersive virtual reality; deep learning; interaction; presence

1. Introduction

There are various studies currently being conducted on immersive virtual reality to provide users with an interface and experience environment that enable real-world interaction through virtual environments or objects with high immersion, as well as realistic and diverse experiences [1–4]. Based on these studies, applications are being developed in various fields, such as education, tourism, manufacturing, and entertainment (e.g., gaming) by focusing on the sense of presence that determines how realistic the user feels about the experience of where the user is and what the user is doing. As for the related technologies, developments are being made to provide the users with a more immersive experience environment by combining virtual reality head-mounted displays (HMDs) such as Oculus Rift S, Oculus Quest, and HTC Vive with other systems such as leap motion, treadmills, and actuators.

The important factors in providing enhanced sense of presence in the immersive virtual reality are user interaction with the virtual environment, supporting devices, and input handling methods. Thus, studies have been conducted on haptic systems that utilize physical information (changes in joints and strength measure) to accurately measure and represent user actions and movements and provide feedback on the physical actions that occur over the interaction [5–7]. In addition, algorithms [8] and portable walking simulators [9] that enable the users to walk freely in a wide virtual reality space in a limited real-world experience space are being studied as well. However, as the experience environment, which depends on the physical devices, is expensive and has a limited range of expression with complicated structure, the applications based on this type of environment face difficulties in

becoming widely used. Recently, to solve this issue, various studies such as the redirection study [10], which enhances the sense of presence by using real and virtual objects at the same time to visually control the virtual objects while providing tactile sensations through real objects, and the pseudo haptic method study [11,12], which enhances the realism over the course of experience by using feedback of physical reaction with visual, have widely been conducted. In addition, a study on providing realistic physical experiences in a virtual space using real props and actuators [13] has been conducted as well. However, these solutions do not fully resolve the dependence on the devices or environment. Thus, there is a need for a method of approach that can provide new experiences and a sense of presence while utilizing the existing virtual reality devices to intuitively interact with virtual environments or objects.

The main objective of this study is to design a new interface that can provide an easy and intuitive interaction with the virtual environment without any additional devices except hands, which are the body parts that users mainly use in immersive virtual reality. To achieve this, the following key structure designs are integrated in the proposed interface.

1. Controller-based hand interface design that directly expresses the gestures taken from real hands to virtual hands
2. Real-time interface design using a deep learning model (CNN) that intuitively expresses the process of gesture to action without GUIs

Convolutional neural network (CNN) is a type of multi-layered artificial neural network used to analyze visual images. This extracts the features of data through a preprocessing step with convolution and polling and performs classification by putting the data into a multi-layered perceptron. In deep learning, it is classified as a deep neural network and is mainly applied to visual image analysis such as image and video recognition, recommendation system, and image classification. This study proposes an interface that creates gesture images through a virtual camera for hand gesture recognition of virtual reality users, and intuitively performs actions through image recognition using CNN.

The proposed interface is expected to provide a more satisfactory virtual experience environment than the GUIs used in the existing interactive applications. A survey was conducted to check whether it yields positive impacts on improving the sense of presence.

2. Related Work

In the field of immersive virtual reality, various studies involving different senses such as visual, auditory, and tactile have been conducted to enhance the sense of presence by providing users with realistic interactions using virtual environments or objects. As for the environment that provides stereoscopic visual information using virtual reality HMDs, various studies (on surround sound processing using audio sources, haptic systems using human body features such as hands and legs, and motion platforms for natural walking) are actively being conducted [9,14–16]. To provide users with a sense of presence that is close to reality through high immersion, realistic interactions that can reduce the gap between virtual and reality are required. To achieve this, multiple studies are being conducted to acquire accurate detection in changes of the joints in human body in real space and grasp the intention of actions to realistically reflect the actions in cyberspace. The study of attaching surface or optical markers to the joints for detecting and tracking the movements with a camera to map them to the actions of virtual model [5,6] and the study of capturing facial expressions, body movements, and hand gestures [4] are examples that aim to express the realistic motion in the cyberspace. Furthermore, to provide a more accessible interaction, various methods such as interacting with Oculus Touch and HTC Vive controllers [7,17], controlling objects based on the gyroscope sensors of the smartphone in mobile virtual reality [18], and directly controlling virtual objects using gaze pointers and hands [2,19] have been studied. Recently, an interactive breath interface that can be applied to virtual reality contents was proposed using the user's breath and the acceleration sensor of

a mobile device. As such, studies that suggest intuitive and convenient interfaces from the user's point of view are being perceived as important and are actively being conducted [20].

During the interaction process, it is also important to express the actions that realistically match the user's intentions and purposes and provide a feedback of the physical force and reactions that occur during this process to the user. Jayasiri et al. [21] proposed a haptic system and interface that express the physical interaction based on the force applied by the user. Further, various applied studies featuring 3-RSR [22], 3-DoF wearable haptic device [23], and a portable hand haptic system [19] have been conducted as well. Additionally, for the study of free free-walking in a limited space, flexible space [8] or portable walking simulators [9] have been proposed. However, due to cost burdens and the complexity limitations of haptic systems, studies on pseudo-haptic have also been established to induce the illusion of receiving physical feedback through visual effects from the recognition response of user's experience [24,25]. Nonetheless, this method is limited to only providing feedback of physical response similar to the haptic systems, and studies on systems and interfaces for directly providing feedback of user intentions and actions have not yet been conducted. Therefore, this study used a deep learning model to propose an interface featuring intuitive interaction.

As for the studies applying deep learning technology to virtual reality, there have been various studies such as applying deep imitation learning in the complex and sophisticated processes of remotely controlling robots using virtual reality HMDs and hand tracking hardware [26], using deep learning methods in tracking 3D objects in augmented reality, and estimating the lighting conditions of images [27]. In addition, CNN, which is a popular deep learning model, has often been used for evaluating or enhancing the quality of 360° panoramic images [28]. Recently, a study on a new CNN-based model (DGaze) for gaze prediction in HMD-based applications [29] and a VIVR study using CNN for walking interaction for visual impairment in immersive virtual reality [30] were also performed. However, few studies have been conducted on the use of deep learning technology for the user-oriented intuitive and popular interface in immersive virtual reality. Thus, to design interfaces that provide new experiences and an enhanced sense of presence, this study proposes a hand interface that enables intuitive gesture to action using CNN.

As an example study of analyzing the sense of presence in immersive virtual reality, Slater et al. made various attempts to analyze relationships based on user actions in a virtual environment and various academic approaches such as psychology and neuroscience [31]. Recently, applied studies have been performed as well to analyze the factors that can enhance the sense of presence in terms of immersive interactions [9,19]. Based on this, we intend to conduct the study to compare and analyze the experience and sense of presence of the proposed DeepHandsVR interface with the GUIs used in the existing interactive applications.

3. DeepHandsVR

The proposed DeepHandsVR interface provides a real-time interface using deep learning to enable intuitive interactions with virtual environment or objects along with a hand interface using a virtual reality controller on the premise of immersive and accessible interactions. Its immersive interaction and experience environment use Oculus Rift CV1 HMD and touch controller and implement an integrated development environment in the Unity 3D engine. Figure 1 presents the process of the proposed DeepHandsVR interface and the interaction with immersive virtual reality.

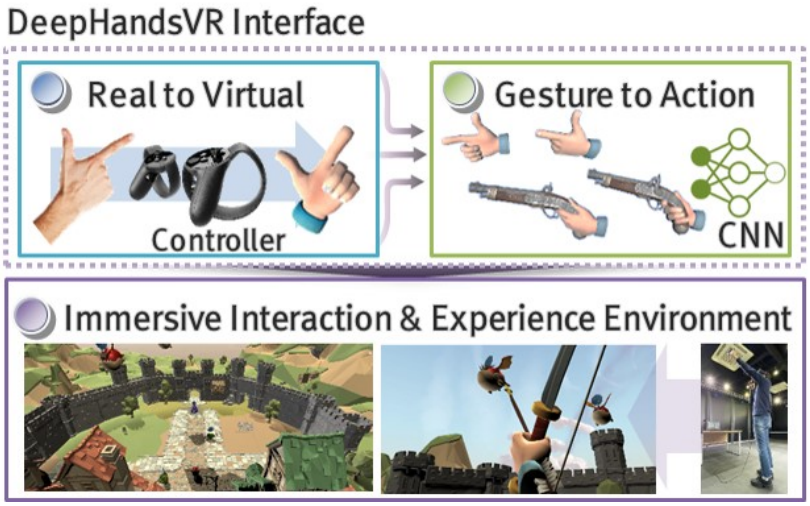

Figure 1. Processes of DeepHandsVR interface and immersive interaction.

3.1. Real to Virtual Direct Hand Interface

Hands are the body parts that the users use the most in expressing their intents or actions; thus, the proposed interface is also designed to allow the users to interact with the virtual environment or objects using hands. Han and Kim [32] found that hand interactions in immersive virtual reality provide a more immersive experience than using gaze pointers, which are used for virtual reality UI and traditional input devices such as keyboards and gamepads. Based on the previous research, this study also proposes a direct hand interface that can minimize the difference in recognition between real hand and virtual hand during hand gestures and actions. Figure 2 shows this interface, where the real hand gestures are defined and keys are mapped to allow natural gestures of the hand holding the controller; thus, the gestures could be seamlessly reflected in the virtual hand. The key part of this method is designing the interface that acts upon the gestures instead of the traditional method of having controllers rely on the keys to interact.

Figure 2. Real-to-virtual direct hand interface structure.

When mapping a controller directly to an action, an action must be defined for every key. For this reason, the number of controllable actions may be limited. Combining GUI and the key to solve this issue is not favorable as the process for the action is lengthened. Therefore, gestures are defined with relatively few key inputs, and various actions are processed intuitively and in a simple structure by performing actions that correspond to the gestures.

Algorithm 1 is the summary of the process of passing the gestures from the real hand to the virtual hand using a controller. In this study, hand gestures are classified into five categories, and the corresponding controller input settings are defined. Further, structures of bleding gestures (e.g., generate grip motion from fist gesture and palm gesture) that can be derived from the five gestures are configured. The structures for the gestures are constructed such that they can be deleted or added as needed.

Algorithm 1 Process of real-to-virtual direct hand interface using controller.

```
keys[] ← key input array of the controller.
0: thumb(top), 1: index trigger, 2: middle trigger
procedure HAND GESTURE CONTROL PROCESS(keys)
    if keys[2] is True then
        if keys[0] is True then
            if keys[1] is True then
                set as fist gesture.
            else
                set as pointing gesture.
            end if
        else if keys[1] is True then
            set as thumb up gesture.
        else
            set as hand gun gesture.
        end if
    else
        set as palm gesture.
    end if
end procedure
gestures[] ← the defined gesture array.
procedure BLENDING GESTURES(gestures)
    i, j ← gestures array index.
    create new gesture by blending i-th gesture (gesture[i]) and j-th gesture (gestures[j]).
    (e.g., grip gesture = blending fist gesture and palm gesture)
end procedure
```

3.2. *Gesture-to-Action Real-Time Interface*

The general interactions using hands in an immersive virtual reality involve the actions of pressing, grabbing, and throwing virtual objects. In the existing interaction systems, the process of selecting tools or changing actions according to the situation requires passing through a GUI. As the virtual reality applications use stereoscopic visual information, GUIs are generally designed in 3D space, rather than 2D windows, to provide menu options in a window (Figure 3). However, the virtual reality user can face inconveniences during the selection process of the GUI method in rapidly changing application environment, and this could eventually hinder the user immersion. Thus, this study proposes an interface that connects gesture to action in real-time without having to go through an additional GUI.

Figure 3. Example of GUI configuration in virtual reality application using images or objects in 3D space [33].

The main goal of the proposed hand interface is to reflect the user-intended actions from the intuitive structure to the virtual environment. Hence, a deep learning method is used in the process of changing gestures to actions.

The DeepHandsVR interface used CNN, which is the most widely used deep learning model in classifying images. In addition, structures and parameters of Google's Inception v3 [34], which is a neural network model that has an efficient structure for image recognition (inference), are applied to the gesture dataset of this study (as part of the retraining process of conducting transfer learning). Figure 4 displays the flow of applying the training process on the proposed gesture-to-action interface. Based on the Inception v3 model structure, 5675 datasets composed of defined gesture and augmented (shift, rotation, size, brightness, etc.) images were collected for gesture to action. Next, the training process was performed for gesture classification. Based on the trained data file, the final training data were generated in the Unity 3D engine development environment by changing the input and output nodes to set the formats according to the input gesture image format and inference results. Lastly, a hand capturing camera was configured in addition to the main camera of the HMD user to save the 3D hand gesture input from the real-to-virtual interface as a 2D image. The hand gesture images taken from the camera are saved in the image buffer of input node format to infer a result from one of the labeled images. To capture the gesturing hand as accurately as possible from the center, the camera position (p_c) and direction ($\vec{v_c}$) are calculated using the positions of the left and right gesturing hand (p_l, p_r) and the vertical vectors of the palm plane of right hand ($\vec{v_r}$) and the palm plane of left hand ($\vec{v_l}$). Here, the direction (θ) and distance (d) of both hands are taken into account in the camera position (p_c) and direction ($\vec{v_c}$). The calculation process is shown in Equation (1).

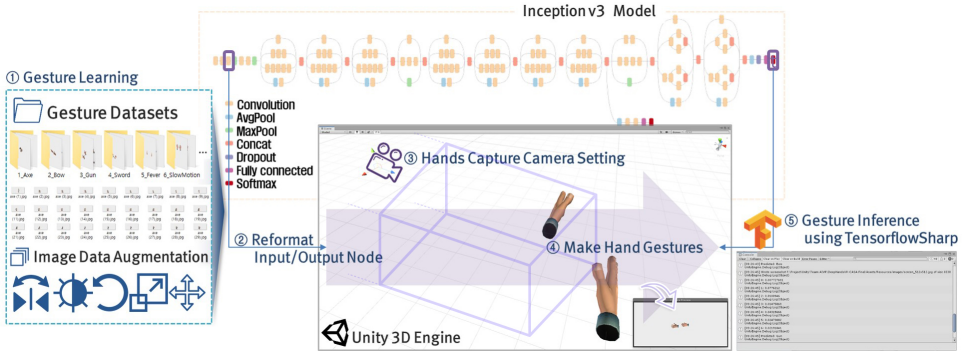

Figure 4. Deep learning model structure for gesture to action interface and process of inference in Unity 3D engine.

The alpha variable (α) is a threshold value adjusted such that it is not too small or too large considering the resolution of the image to be captured. The alpha variable is set to 1 when shooting

with the orthographic projection; however, when capturing with the perspective projection, the user can adjust the value depending on the image resolution and the hand size.

$$\begin{aligned} p_h &= \frac{(p_l + p_r)}{2}, \\ \theta &= \cos^{-1}\left(\frac{\vec{v_l} \cdot \vec{v_r}}{|\vec{v_l}||\vec{v_r}|}\right), \\ d &= |p_l.y - p_r.y| + |p_l.z - p_r.z|, \\ \vec{v_h} &= \theta > 100^\circ \; and \; d > 0.1 \; ? \; \vec{v_l} - \vec{v_r} : \vec{v_l} + \vec{v_r}, \\ \vec{v_h} &= \frac{\vec{v_h}}{|\vec{v_h}|}, \\ p_c &= p_h + \alpha \times \vec{v_h}, \\ \vec{v_c} &= -\vec{v_h} \end{aligned} \tag{1}$$

The process of inferring from the trained data file by applying it to the Unity 3D engine can be summarized into the following three steps. The implementation is performed through the TFGraph class, provided by the TensorFlowSharp plugin.

(a) Generate and train graph objects (TFGraph), and load label data.
(b) Input the gesture image after changing the format according to the set input node.
(c) Calculate the probability result value for each label inferred from the output node.

3.3. Immersive Interaction

The hand interface with a deep learning model allows the users to interact more directly with virtual environments or objects in an intuitive structure. Figure 5 presents experimental actions from this study corresponding to each of the six gestures used for training process. The user takes one of the defined gestures through the controller. Here, the real-to-virtual interface is used to reflect a realistic input process. Next, based on the training process of the gesture to action interface, the gesture result is inferred, and the corresponding action is performed in real-time without going through a separate GUI. The six gestures used herein were defined by considering the characteristics of the application (Section 4). These gestures can be customized according to the characteristics and purposes of the application or the content that the user wants to create.

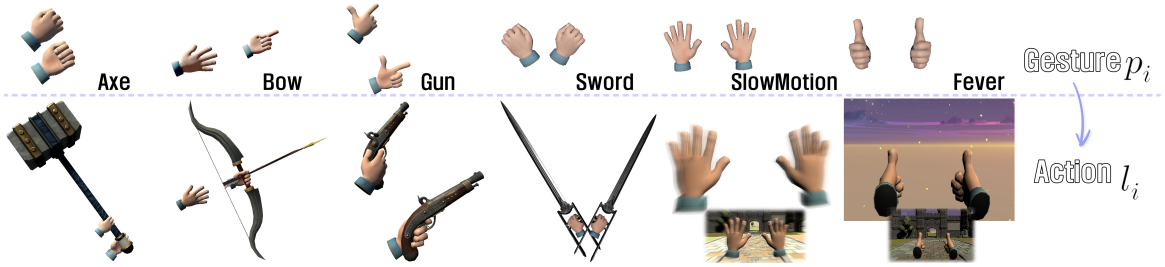

Figure 5. Immersive interaction process of directly reflecting corresponding gestures to the actions in the virtual environment.

Equation (2) is the calculation process of action corresponding to the gesture based on the probability result values inferred for each label through the training process. The six types of gestures introduced in the proposed interface have a probability inference value of p_i. If the maximum probability value (p_f) is greater than 60%, the label index (l_i) of the corresponding gesture is searched for, and the action corresponding to the label index (l_i) is selected and activated in the array (A). The *Active* function activates the action corresponding to the inferred gesture. Here, the 0th action

denotes the default action in which inference does not apply. The 60% range is the threshold value derived through the process of repeating gesture recognition about 100 times.

$$\begin{aligned} &p_f = \max_{i=1}^{6}(p_i), \\ &l_i = p_f > 0.6\,?\,argmin(p_f) : 0, \\ &Active(A[l_i]) \end{aligned} \tag{2}$$

4. Application

The hand interface proposed in this study allows the users to experience the application in a more convenient environment by directly connecting the action from the gesture instead of going through the traditional immersive virtual reality application process of determining the action through the GUI. This method is expected to provide a satisfactory interface experience and ultimately enhance the sense of presence. Thus, this study directly creates a virtual reality application for the user evaluation on the proposed interface.

The application is arranged and designed as an arcade game based on the six gestures and the corresponding actions from the categories defined above. The basic configuration and flow of the application are as follows. Four of the six gestures and actions suggested in this study correspond to attacks, and two monsters are assigned per gesture. Therefore, eight monsters are randomly generated in the application. Basically, in order for the user to remove a monster, he/she must perform an action through the gesture assigned to the monster. The other two gestures are in charge of a special function of the application, and have the function of reducing the movement speed of the monster or increasing the monster removal score. However, to perform a special function, it can only be used after a certain period of time or when more than a specified number of monsters are removed. The application experience time was 5 min, and i a recording method was planned, in which users who obtained many points during a given time occupy a high rank (Supplementary Video S1).

Figure 6 displays the execution process of the application developed in this study, where, if the user presses the designated button while performing a gesture of a desired action, the inferred action is directly performed through the training process. The purpose of this study is to provide improved user satisfaction with the proposed interface compared to the previous GUI method. To compare the satisfaction levels, an additional mode for selecting actions through GUI method is provided in the exact same experience environment. In other words, to remove monsters or perform special functions, the proposed interface directly performs actions through gestures, or the GUI method performs actions by selecting image buttons using a controller (Supplementary Video S1).

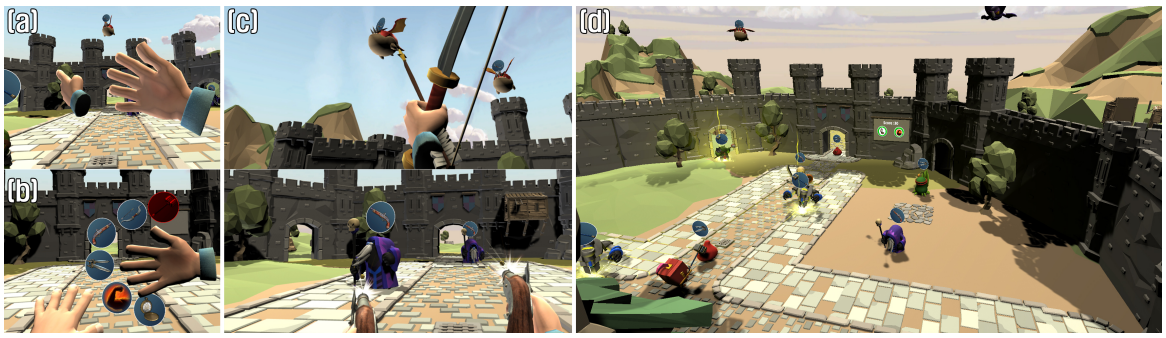

Figure 6. Application development result: (**a**) proposed interface; (**b**) existing GUI control method; (**c**) play scene; and (**d**) overall application scene.

5. Experimental Results and Analysis

The deep learning model applied to the proposed interface was implemented using Anaconda 3, conda 4.6.12, and TensorFlow 1.13.0. The experiment on the learning model in Unity 3D engine was

implemented using TensorFlowSharp 1.15.1 plugin. Further, the virtual reality application for the user survey was created using Unity 3D 2019.2.3f1 (64-bit) and Oculus SDK. The PC configuration for the system implementation and experiments consisted of Intel Core i7-6700, 16 GB RAM, and Geforce GTX 1080 GPU. Figure 7 shows the experience environment where the user can try the virtual reality application after wearing the Oculus HMD and touch controller. As it is a 1.5 m $\times$ 1.5 m environment, the user can try the application comfortably, regardless of sitting or standing.

Figure 7. Immersive virtual reality experience environment: (**a**) sitting; (**b**) standing.

A survey was conducted on the users to compare and analyze the satisfaction level of the proposed interface in the experience environment shown in Figure 7. The survey employed questions from previously verified questionnaires that were used to analyze the user experience and presence in the immersive virtual reality application field. There were 16 participants (male and female) between the ages 22 and 38. To be able to compare interfaces regardless of whether they experience virtual reality or not, it has been configured in a variety of ways, from participants who do not have experience with virtual reality contents to those who frequently experience virtual reality contents. The key purpose of the survey was to verify whether the proposed hand interface provides a convenient and satisfactory interface experience (when compared with the existing GUI method) and provides positive impacts on improving the sense of presence in immersive virtual reality. For an objective comparison experiment, first of all, half of the participants conducted the proposed interface and the other half experienced the existing GUI method first. In addition, other factors (application progress and composition, game elements, experience time, etc.) than the interface were kept the same to increase the accuracy of the comparison experiment.

The first experiment was a comparative questionnaire to compare the satisfaction levels of the hand interface. The immersive interaction using the proposed interface is intended to provide users a convenient experience by directly dealing with virtual environments or objects without using GUIs. Hence, to evaluate this, separate experimental applications were developed using the proposed interface and the existing GUI method. Next, based on the usefulness, satisfaction, and ease of use (USE) questionnaire by Arnold Lund [35], the results were recorded on a seven-point scale for the 30 items of four dimensions of usability. Table 1 shows the statistical data based on the survey results. In each of the four categories (usefulness, ease of use, ease of learning, and satisfaction), the proposed interface showed higher satisfaction levels than the existing GUI. In particular, as the gesture used as input directly expresses the user intended action, the interface was found to be

easier to use. The usefulness was also recorded with significant differences. However, due to the familiarity of the existing GUI, it was confirmed that it is as easy to learn as the proposed interface. At this time, due to the proposed interface being based on training, if the trained results were inferred inaccurately, there was a slight probability of conducting different action than the desired one, resulting in inconveniences. The results of calculating statistical significance through one-way ANOVA analysis showed an improved satisfaction and a significant difference in the overall improvement (usefulness, ease to use, and satisfaction) on the proposed interface.

The second survey was regarding the analysis of the sense of presence. This study focused on eliminating inconvenient processes as much as possible by directly expressing the decision-making process of the user actions. We expected this focus to have a positive effect on the user immersion in virtual reality. Therefore, a survey was conducted to verify this by comparing with the existing GUI. Based on the 19-item presence questionnaire proposed by Witmer et al. [36], the survey participants recorded their responses on a seven-point scale. The items were compared and analyzed in detail based on the recorded values. The results are shown in Table 2, where the proposed interface received higher average scores on the overall items of presence. In particular, significant differences were noticed in realism, possibility to act, quality of interface, and possibility to examine, which are the categories directly related to the action. The participants responded that the proposed interface reflected the user's actions in the virtual environment more directly and realistically, and this was shown to increase immersion by accurately inferring the user's action results. Similar to the USE survey, the results of calculating statistical significance through one-way ANOVA analysis showed significant differences in most items, and the interface was shown to induce an improved sense of presence. The existing GUI is a general interaction method centered on a menu, and the user can use it easily and skillfully. Due to this, similar results were found in the self-evaluation of performance without any significant difference. However, the DeepHandsVR interface, which was able to quickly interact with the virtual environment with an intuitive structure, showed significant differences in all aspects such as realism and possibility to act and examine.

Table 1. Satisfaction analysis results of the proposed hand interface.

	DeepHandsVR	**Existing GUI**
Mean(SD)		
usefulness	5.625(1.048)	4.610(1.467)
ease of use	5.511(1.291)	4.614(1.308)
ease of learning	5.234(1.726)	5.172(1.374)
satisfaction	6.188(0.910)	4.696(1.667)
Pairwise Comparison		
usefulness	$F(1,30) = 4.761$, $p < 0.05$ *	
ease of use	$F(1,30) = 4.512$, $p < 0.05$ *	
ease of learning	$F(1,30) = 0.012$, $p = 0.913$	
satisfaction	$F(1,30) = 9.242$, $p < 0.01$ *	

* indicates statistical significance.

Table 2. Statistical comparative analysis results for presence with the proposed hand interface.

	DeepHandsVR	**Existing GUI**
Mean(SD)		
total	6.138(0.542)	5.102(1.108)
realism	6.308(0.596)	5.174(1.088)
possibility to act	6.101(0.631)	5.109(1.019)
quality of interface	5.938(0.757)	4.833(1.958)
possibility to examine	6.198(0.757)	5.115(0.839)
self-evaluation of performance	5.813(1.579)	5.219(1.262)

Table 2. *Cont.*

	DeepHandsVR	Existing GUI
Mean(SD)		
Pairwise Comparison		
total	$F(1,30) = 10.594$, $p < 0.01$ *	
realism	$F(1,30) = 12.534$, $p < 0.001$ *	
possibility to act	$F(1,30) = 10.438$, $p < 0.01$ *	
quality of interface	$F(1,30) = 4.150$, $p < 0.05$ *	
possibility to examine	$F(1,30) = 13.779$, $p < 0.001$ *	
self-evaluation of performance	$F(1,30) = 1.293$, $p = 0.264$	

* indicates statistical significance.

In addition, the accurate gesture inference in the proposed interface is an important factor in providing the user with improved presence through high satisfaction and immersion. Therefore, the accuracy of gesture inference was recorded during the process of the survey experiment. The average accuracy of the gestures (axe gesture: 73.35%; bow gesture: 90.78%; gun gesture: 76.61%; sword gesture: 74.89%; fever gesture: 79.89%; and slow-motion gesture: 94.71%) were recorded. In the case of some gestures or specific pose with relatively low accuracy, it was analyzed that the performance should be improved by supplementing the learning data, although it did not significantly affect the application experience.

Finally, the frame rate is derived by measuring the speed starting from gesture recognition to action execution. For virtual reality applications, as the frame rate factors, including the number of frames per second (fps), affect user immersion, such as by inducing VR sickness, assuring that gesture recognition does not affect simulation speed is necessary. First, a difference of up to 10 fps was measured as a frame rate difference between the initial screen before gesture recognition and at the point at which the action result appears after the recognition. However, the overall frame rate was not at the level where the user experience in virtual reality would be affected. The recognition time was also less than 0.000001 s, suggesting that the recognition and inference processes do not have a significant impact on the system.

6. Limitation and Discussion

The deep learning model used in the hand interface of this study was CNN; among various CNN methods, transfer learning using Inception v3, which infers efficient structure and accurate results, was chosen. This was because the main purpose of this study was to focus on presenting a new interface with a training method and analyzing it, rather than developing a whole new learning model. However, as the experiment results showed that the interface using deep learning model provides users high satisfaction levels and an improved sense of presence, it may be necessary to design a new specific CNN model that is more optimized for the interface. In addition, as the gestures and actions defined in this study were prototype versions, the variety is very limited. It is necessary to conduct more experiments to define and analyze more diverse number of gestures and corresponding actions.

Additionally, the experimental application was designed and developed as an arcade game; however, it is important to design the interface proposed in this study to be applicable not only to games but also other various fields, such as education, manufacturing, and medicine. Therefore, as shown in Figure 8, we plan to improve the interface to be easily used when implementing various experiences in daily life into virtual reality. In addition, we intend to further investigate the pseudo-haptic approach by comparing experimental results with those of existing studies related to hand interfaces using haptic feedback.

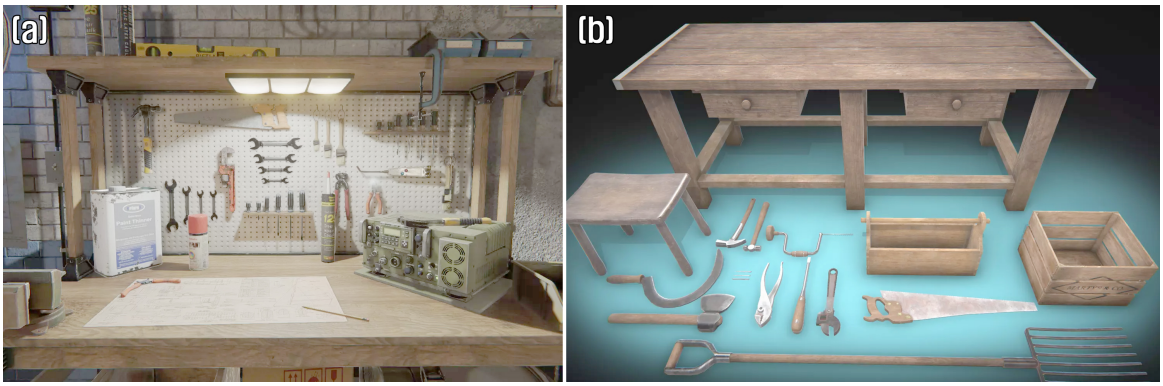

Figure 8. Various application examples of applying the interface proposed in this study: (**a**) DIY workshop; (**b**) woodwork.

7. Conclusions

This study proposed a hand interface using a deep learning method to provide users with a realistic experience in an intuitive structure when interacting with immersive virtual reality (using hands of the users). The proposed real-to-virtual direct-hand interface was designed to feature a structure that can be easily utilized at a low cost while providing seamlessness in the process of rendering the real actions taken using hands to the virtual environment. Additionally, a real-time gesture to action interface was designed to allow the users to intuitively connect gestures to actions and interact with virtual environments and objects without having to use a GUI as in existing interactive applications. This was implemented using a deep learning model (CNN) to enable fast and accurate action inference by applying the process of training and inferring the gesture images. This process allowed the user to become familiar with the interface while intuitively expressing the intended actions, and ultimately increased the user's immersion in virtual reality to provide improved sense of presence. For analysis, comparative surveys (using USE and presence questionnaire) were conducted on the existing GUI method and the proposed interface, and the interface was demonstrated to have positive effects on the sense of presence while yielding a satisfactory interface experience.

Supplementary Materials: The following are available online at http://www.mdpi.com/2079-9292/9/11/1863/s1, Video S1: DeepHandsVR: Hand Interface Using Deep Learning in Immersive Virtual Reality.

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

Funding: This study was supported by the Korea Creative Content Agency (KOCCA) funded by the Ministry of Culture, Sports, and Tourism (MCST) for Mingyu Kim. Also, this research was financially supported by Hansung University for Jinmo Kim.

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

References

1. Pfeiffer, T. Using virtual reality technology in linguistic research. In Proceedings of the 2012 IEEE Virtual Reality Workshops (VRW), Costa Mesa, CA, USA, 4–8 March 2012; pp. 83–84.
2. Sidorakis, N.; Koulieris, 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.
3. Jeong, K.; Lee, J.; Kim, J. A Study on New Virtual Reality System in Maze Terrain. *Int. J. Hum. Comput. Interact.* **2018**, *34*, 129–145. [CrossRef]

4. Joo, H.; Simon, T.; Sheikh, Y. Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UH, USA, 18–23 June 2018; pp. 8320–8329.
5. Metcalf, C.D.; Notley, S.V.; Chappell, P.H.; Burridge, J.H.; Yule, V.T. Validation and Application of a Computational Model for Wrist and Hand Movements Using Surface Markers. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1199–1210. [CrossRef] [PubMed]
6. Zhao, W.; Chai, J.; Xu, Y.Q. Combining Marker-based Mocap and RGB-D Camera for Acquiring High-fidelity Hand Motion Data. In Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Lausanne, Switzerland, 29–31 July 2012; Eurographics Association: Aire-la-Ville, Switzerland, 2012; pp. 33–42.
7. Inrak, C.; Eyal, O.; Hrvoje, B.; Mike, S.; Christian, H. CLAW: A Multifunctional Handheld Haptic Controller for Grasping, Touching, and Triggering in Virtual Reality. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 654:1–654:13.
8. Vasylevska, K.; Kaufmann, H.; Bolas, M.; Suma, E.A. Flexible spaces: Dynamic layout generation for infinite walking in virtual environments. In Proceedings of the 2013 IEEE Symposium on 3D User Interfaces, Orlando, FL, USA, 16–17 March 2013; pp. 39–42.
9. Lee, J.; Jeong, K.; Kim, J. MAVE: Maze-based immersive virtual environment for new presence and experience. *Comput. Anim. Virtual Worlds* **2017**, *28*, e1756. [CrossRef]
10. Carvalheiro, C.; Nóbrega, R.; da Silva, H.; Rodrigues, R. User Redirection and Direct Haptics in Virtual Environments. In Proceedings of the 2016 ACM on Multimedia Conference, Amsterdam, The Netherlands, 15–19 October 2016; ACM: New York, NY, USA, 2016; pp. 1146–1155.
11. Pusch, A.; Martin, O.; Coquillart, S. HEMP-Hand-Displacement-Based Pseudo-Haptics: A Study of a Force Field Application. In Proceedings of the 2008 IEEE Symposium on 3D User Interfaces, Reno, NE, USA, 8–9 March 2008; pp. 59–66.
12. Achibet, M.; Gouis, B.L.; Marchal, M.; Léziart, P.; Argelaguet, F.; Girard, A.; Lécuyer, A.; Kajimoto, H. FlexiFingers: Multi-finger interaction in VR combining passive haptics and pseudo-haptics. In Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, 18–19 March 2017; pp. 103–106.
13. Lung-Pan, C.; Thijs, R.; Hannes, R.; Sven, K.; Patrick, S.; Robert, K.; Johannes, J.; Jonas, K.; Patrick, B. TurkDeck: Physical Virtual Reality Based on People. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Daegu, Kyungpook, Korea, 8–11 November 2015; ACM: New York, NY, USA, 2015; pp. 417–426.
14. 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. [CrossRef]
15. Choi, I.; Hawkes, E.W.; Christensen, D.L.; Ploch, C.J.; Follmer, S. Wolverine: A wearable haptic interface for grasping in virtual reality. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 986–993.
16. Sebastian, M.; Maximilian, B.; Lukas, W.; Cheng, L.-P.; 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.
17. Jeong, K.; Kim, J.; Kim, M.; Lee, J.; Kim, C. Asymmetric Interface: User Interface of Asymmetric Virtual Reality for New Presence and Experience. *Symmetry* **2019**, *12*, 53. [CrossRef]
18. Kim, M.; Lee, J.; Jeon, C.; Kim, J. A Study on Interaction of Gaze Pointer-Based User Interface in Mobile Virtual Reality Environment. *Symmetry* **2017**, *9*, 189. [CrossRef]
19. 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. [CrossRef]
20. Kim, J.; Lee, J. Controlling your contents with the breath: Interactive breath interface for VR, games, and animations. *PLoS ONE* **2020**, *15*, e241498. [CrossRef] [PubMed]
21. Jayasiri, A.; Ma, S.; Qian, Y.; Akahane, K.; Sato, M. Desktop versions of the string-based haptic interface—SPIDAR. In Proceedings of the 2015 IEEE Virtual Reality (VR), Arles, France, 23–27 March 2015; pp. 199–200.

22. 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. [CrossRef] [PubMed]
23. Prattichizzo, D.; Chinello, F.; Pacchierotti, C.; Malvezzi, M. Towards Wearability in Fingertip Haptics: A 3-DoF Wearable Device for Cutaneous Force Feedback. *IEEE Trans. Haptics* **2013**, *6*, 506–516. [CrossRef] [PubMed]
24. Andreas, P.; Anatole, L. Pseudo-haptics: From the Theoretical Foundations to Practical System Design Guidelines. In Proceedings of the 13th International Conference on Multimodal Interfaces, Alicante, Spain, 14–18 November 2011; ACM: New York, NY, USA, 2011; pp. 57–64.
25. 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. [CrossRef]
26. 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.
27. Lalonde, J. Deep Learning for Augmented Reality. In Proceedings of the 2018 17th Workshop on Information Optics (WIO), Quebec City, QC, Canada, 16–19 July 2018; pp. 1–3.
28. Yang, J.; Liu, T.; Jiang, B.; Song, H.; Lu, W. 3D Panoramic Virtual Reality Video Quality Assessment Based on 3D Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 38669–38682. [CrossRef]
29. Hu, Z.; Li, S.; Zhang, C.; Yi, K.; Wang, G.; Manocha, D. DGaze: CNN-Based Gaze Prediction in Dynamic Scenes. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 1902–1911. [CrossRef]
30. Kim, J. VIVR: Presence of Immersive Interaction for Visual Impairment Virtual Reality. *IEEE Access* **2020**. [CrossRef]
31. Mel, S.; Sanchez-Vives, M.V. Transcending the Self in Immersive Virtual Reality. *Computer* **2014**, *47*, 24–30.
32. Han, S.; Kim, J. A Study on Immersion of Hand Interaction for Mobile Platform Virtual Reality Contents. *Symmetry* **2017**, *9*, 22. [CrossRef]
33. Kim, M.; Lee, J.; Kim, C.; Kim, J. TPVR: User Interaction of Third Person Virtual Reality for New Presence and Experience. *Symmetry* **2018**, *10*, 109. [CrossRef]
34. 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.
35. Lund, A. Measuring Usability with the USE Questionnaire. *Usability Interface* **2001**, *8*, 3–6.
36. Witmer, B.G.; Jerome, C.J.; Singer, M.J. The Factor Structure of the Presence Questionnaire. *Presence Teleoper. Virtual Environ.* **2005**, *14*, 298–312. [CrossRef]