2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 978-1-6654-5325-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISMAR55827.2022.00032

NailRing: An Intelligent Ring for Recognizing Micro-gestures in Mixed Reality

Tianyu Li Yue Liu* Shining Ma Mingwei Hu Tong Liu Weitao Song†

Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology

Figure 1: NailRing is an intelligent finger ring that recognizes micro gestures based on fingertip physiological characteristics to achieve interactive input in MR. (a) The user wears the NailRing on a flat surface to perform interaction tasks by fingertip micro-gestures; (b) In the walking state, the user responds quickly to interaction requirements by one-handed gestures; (c) The user interacts with the MR application through the NailRing with the back of the auxiliary hand as the interaction plane.

Abstract

Gesture interaction is currently a main interaction technology in the field of mixed reality. However, long-term and large-scale gesture in mid-air will lead to muscle fatigue and privacy problems, which cannot meet the comfort requirements of continuous interaction and inevitably hinder the development of mixed reality systems. To solve this problem, we propose NailRing, an intelligent ring to recognize fingertip micro-gestures using a micro-close-focus camera on a fingertip bracket. Such fingertip physiological characteristics as the changes in fingertip color distribution and muscle shape changes caused by fingertip pressure have been studied. According to the recognition principle, ten types of micro-gestures have been designed and used for contact interaction and one-hand interaction respectively. The accuracy of gesture recognition (cross-session $F_{Macro} = 98.3\%$; cross-person $F_{Macro} = 86.4\%$) in user studies verifies the performances of NailRing under different interaction conditions. Finally, the capability of NailRing in a series of potential application scenarios has also been discussed and analyzed.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction techniques—Gestural input

1 Introduction

In recent years, more and more mixed reality (MR) devices (e.g. Hololens 2) and virtual reality (VR) devices (e.g. Quest 2) have begun to adopt gesture recognition technology as system input [16]. Gestures have gradually replaced traditional controllers and been used in the interaction of MR systems due to their simplicity and naturalness. At present, the main gesture interaction solution in the MR field is to first use the built-in camera of the HMD to obtain the gesture in front of the user's line of sight, and then use the detection algorithm to track and recognize the gesture. However, this kind of long-term hovering gesture will cause fatigue of arm muscles, and the large-scale gesture action will also bring unnecessary interaction burden to the user [13], which has become a key obstacle to the interaction of MR scenes.

In order to improve interactive experience in MR applications, it is necessary to simplify interaction scenarios of macro gestures to be mainly used in the interaction with virtual objects, such as "grasping" and "moving" virtual objects. Non-essential operations such as command input and UI control should be performed in a more convenient interactive way. Current recognition schemes use voice interaction to replace part of the function of gestures, but which is generally regarded as indecent and privacy affecting in many occasions [12]. At the same time, users have requirements for fast and eyes-free interaction [22]. For example, when dialing in happens during a meeting or in public, users may want to hang up or mute the call immediately instead of reaching out and clicking the corresponding function button in mid-air [23]. Therefore, there is an urgent need for a private, fast and eyes-free interaction method to replace some of the macro-gesture functions to improve the interactive experience of MR systems under various scenarios.

Existing researches have explored the use of micro-gestures as a new interaction method for MR applications. Compared with the interaction scheme of macro-gestures, micro-gestures enable users to interact with MR systems in a more private and convenient way. Tung [26] found that users wearing MR head-mounted displays prefer to use finger-based and less obvious interactive gestures in public. Though micro-gestures have many advantages in the field of interaction, such as naturalness and convenience. However, due to the complexity of micro-gesture features and the serious occlusion problems in the interaction process, it is not feasible to use the built-in camera of HMD to obtain and identify data. The state-of-art work focuses on exploring suitable wearable devices to realize

*Corresponding author: liuyue@bit.edu.cn
†Corresponding author: swt@bit.edu.cn

978-1-6654-5325-7/22/$31.00 ©2022 IEEE
DOI 10.1109/ISMAR55827.2022.00032

micro-gesture interaction, but majority of the current solutions has such problems as uncomfortable interaction posture [10], the detection method's prone to environmental interference [8, 32] and device's possibly causing interference to the user [25], etc. During the research, we found that most of the research works focused on sensing devices and recognition algorithms for the detection of micro-gestures, thus ignoring the characteristics of fingers due to the unique physiological structure during the interaction process, which is a direction worth exploring.

In order to solve the above interaction problems in MR applications, we propose NailRing, an eyes-free intelligent ring that recognizes fingertip micro-gestures. NailRing obtains fingertip images through a miniature close-focus camera on the fingertip bracket. Fingertip micro-gestures are identified according to the color distribution change of the nail bed caused by blood perfusion mechanism (formed by the interaction of the phalanx, nail cap and muscle when the finger pulp is compressed). The muscle shape and color changes of fingertip are also used as a recognition feature for joint detection.

Figure 1a shows the application scenario of NailRing. When the user has interaction requirements in MR environment, the finger wearing the ring can be used to slide on the desktop to complete micro-gesture command input. The interaction scope of NailRing covers most surfaces in our daily life. In addition to the commonly used flat surfaces such as desktops and walls, users can interact with narrower surfaces such as seat armrests and clothing surfaces on the legs in a sitting state with NailRing. The user can also use the back of the auxiliary hand as the plane of interaction. As shown in Figure 1b, NailRing can recognize the micro-gestures of one hand for urgent operation requirements when there is no interactive plane, and perform fast and conceal interactive operations in special states as walking. Figure 1c shows the application example of NailRing in the MR system. The micro-gesture recognition algorithm processes the gesture data, and the recognition results are used as input control function buttons to interact with the UI panel.

In user experiment, we evaluated the recognition performance of NailRing with 10 participants. Experimental results show that NailRing can reliably recognize micro-gestures (cross-session $F_{Macro} = 98.3\%$, cross-person $F_{Macro} = 86.4\%$), and the recognition deviation among different subjects is within an acceptable range (SD=0.05). In our experiments, NailRing exhibits recognition robustness under different colors, material interaction planes, and different lighting conditions. Under various situations, users can complete an input task in an average of 1.14 seconds. We also evaluate the performance of NailRing in different poses (Accuracy: Sitting=0.875, Standing=0.816) using the back of the hand as the interaction plane.

The main contributions of this work are summarized as follows:

- For the first time, we proposed an intelligent ring based on the characteristics of fingertip blood perfusion and muscle shape change as a human-computer interaction device in MR. Our solution is eyes-free and natural, which can realize relatively hidden input.
- We implemented a prototype of NailRing with 10 targeted micro-gestures with clear meanings and achieved high recognition accuracy.
- A 10-participant user study was conducted to evaluate system performance under multiple scenarios, possible applications and potential of the technology in MR systems were also discussed.

2 Related Work:

2.1 Wearable devices that recognize finger interactions

Wearable devices have become the main technical solutions in the field of micro-gesture recognition. In recent years, researchers have proposed a variety of wearable devices for finger interaction based on different sensing principles. Among them, inertial measurement units (IMUs) has become one of the most commonly used sensing methods due to its small size and low power consumption. Inertial sensing technology is widely used in finger tracking and pointing tasks [7, 33], as well as touch detection tasks [6]. It can also be combined with other sensing methods [11] for micro-gesture interaction tasks. In addition, various other sensing methods are also used to recognize micro-gestures. For example, Chan [2] used the magnetic sensor installed on the fingernail to recognize the interactive micro-gestures of the fingertip and realized the hidden input. Based on this research, Chen [3] proposed a multi-point tracking system using magnetic field sensing to identify the subtle movements of the fingertips to realize interactive tasks such as air writing; Zhang [34] used the thermal imaging sensor worn on the auxiliary hand to recognize the sliding, clicking and other movements of the interactive hand. In recent years, researchers have also carried out a lot of explorations in the micro-gesture recognition work based on wearable devices. Among these solutions, ring-based recognition approaches are widely adopted, and a summary of ring-based gestures and recognition devices is presented by Vatavu [27] et al. Still and all, the current solutions are still flawed in some aspects.

2.2 Interactions based on fingertip physiology

The research on the principle of using blood infusion was first proposed by Mascaro in 2001 [14], in which miniature light-emitting diodes and photodetectors were used to measure the change in the reflected intensity of the nail surface when pressed, and then estimated the degree of finger pressure based on the change in intensity called "reflected light volume scanning". In subsequent work, Grieve [4] et al. verified in a stationary platform with a good light source that the blood infusion mechanism was also available for other fingers, and Sato [20] et al. explored the possible application of this recognition mechanism to patients with absolute rest. In 2019, Saito et al. [18] proposed a wearable device that estimated the contact force of the fingertip according to the deformation of the fingertip skin when the force was applied. The device uses multiple photo-reflective sensors (PRSs) to measure the distance to the side of fingertip. However, it is unstable to judge the pressing situation only based on the distance, and calibration is required each time of wearing. The work in this paper is mainly inspired by the above work. We fabricate an intelligent ring to obtain real-time images of fingertips, and judge the user's input instructions more accurately according to the principle of blood infusion of fingertips and changes in the shape of fingertips muscles. The unique perspective allows us to recognize the interactive gestures of thumb and fingers. We also explored the usability of the proposed wearable device in the MR field through multiple aspects. In recent years, other uses of human physiological structure as identification features are also worthy of attention. K Sakuma et al. [19] proposed a wearable strain sensor, which used the feature of fingernail deformation under the conduction of complex human structure when the finger pulp was under force to detect input gestures, so as to perform human-computer interaction.

2.3 Micro gesture interaction in MR domain

Gesture recognition is currently the main interaction scheme in the MR field, and micro-gesture interaction has received quite a few attentions as one of the main directions. For example, Weng et al [28] used the infrared camera installed on the HMD to recognize hand-to-face gesture, which can be used as a gesture interaction input in the MR system, but the proposed scheme is not in line with the user's interaction habits. Meier et al. proposed TapID [16], a wrist-based inertial sensing system that recognizes fingertip clicks and complements headset-tracked hand poses. It is worth mentioning that we believe many micro-gesture recognition schemes based on wearable devices proposed in recent years also have potential as MR interaction inputs [11, 22], but these schemes need to be verified in

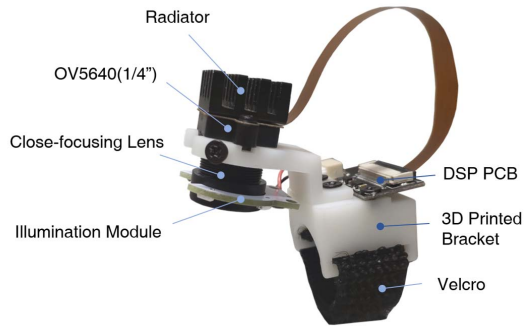

Figure 2: Hardware for the NailRing prototype

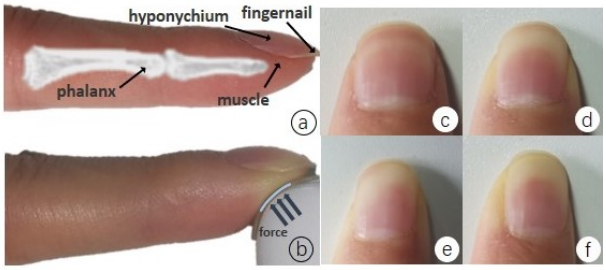

Figure 3: The color distribution of the nail bed and the shape of the muscle color change due to the blood perfusion mechanism formed by the interaction of the finger bone, nail cap and muscle when the finger belly is compressed. (a) Schematic of fingertip structure; (b) Change in fingertip color distribution due to contact force; detail diagram of fingertip color distribution due to different pressing forces: (c) Finger lightly touching the table; (d) Fingertip pressing downward; (e) Fingertip sliding to the left; (f) Fingertip sliding to the right;

real situations for usability in MR environments.

3 Design And Implementation Of NailRing:

3.1 Hardware design

Figure 2 shows a prototype of NailRing, which uses a 3D-printed bracket as its main support and then is secured to the front of the user's index finger by a flexible elastic band. NailRing uses OV5640(1/4 ") as an image sensor to obtain image data (640*480) at an operating frequency of 30Hz. Due to size constraints, we used a close focal length optical lens with structure 2G2P (F/No=2.4, EFL=3.7mm, FOV=100°). The lens and sensor module mounted on a bracket have been modified several times, making it possible to obtain a clear image of the fingertip when pressed. In order to make the NailRing adapt to various lighting environments, we designed a lighting module symmetrically distributed with 4 patch LEDs. The raw data is digitally signal processed on the PCB and then transferred via Universal Serial Bus to a python program running on a PC for online recognition, and then transferred over the network for input on the MR device. In the process of testing the equipment, we found that the NailRing consumed about 930mW of power.

3.2 Recognition principle

Figure 3 depicts the color distribution of the nail bed and the changes in muscle color and shape caused by the blood perfusion mechanism when the finger pulp is compressed. As shown in Figure 3(a), the nail bed and fingernail are tightly combined with the fingertip muscles and distal phalanx in a complex manner [15]. As shown in Figure 3(b), when the finger pulp contacts the plane, the contact plane interacts with the complex physiological structure composed of the fingernails and bones, resulting in the compression of the capillaries in the nail bed and the muscle, and the blood volume flowing through the fingertip is affected. The effect is manifested as obvious changes in the color of the nail bed and fingertip muscles, and at the same time, the muscles of the fingertips will show obvious shape changes under the action of nails, bones and interaction planes. As shown in Figure 3(c), when the pressure reaches about 0.3N, the venous vascular return constricts, and the arterial blood accumulates in the capillaries of the fingertips. Because the arterial blood is rich in hemoglobin, the color of the nail bed becomes red. When the contact force reaches about 1N, as shown in Figure 3d, the vein is completely blocked. With the increase of contact pressure, the blood in capillaries is gradually squeezed out of the fingertip area, resulting in the white area at the free end of fingertip, and the muscle shape begins to change. The variation increases with increasing pressure until it reaches a limit when the contact pressure reaches approximately 4N [14]. Figures 3(e) and (f) show the color pattern of the fingertip when swiping left and right, respectively. The nail bed and muscles also exhibit complex, regular color patterns under different contact pressure patterns. The degree of variation in these patterns may vary from person to person, but the fingertip characteristics determined by its physiology apply to most healthy nails.

3.3 Microgestures Category

To improve the usability of NailRing, we targeted 10 finger-based micro gestures as shown in Figure 4. Detecting touch events is a long-standing challenge in MR interaction [29,31], but NailRing can solve this problem well based on the blood perfusion mechanism. We define the finger contact state as two micro gestures, with good touch detection to optimize the interaction experience, and non-touch state for gesture segmentation and elimination of interference. The different directions of force on the finger pulp will lead to changes in the color distribution of the fingertips. We define the natural sliding of the index finger on the plane in four directions as different micro-gestures, and the clicking of the index finger on the plane as a micro-gesture with the meaning of "Enter". Through the above five gestures, the user can use NailRing to naturally complete the operation of movement and selection in various tasks. NailRing can obtain the relative position of the index finger and thumb. According to this principle, we define three one-handed micro-gestures for fast interaction without interaction plane. And such gestures can complement the interaction mode of planar interaction (such as switching between upper and lower menus in the UI). These micro-gestures mean exactly what the user is used to and are smaller in magnitude than traditional gestures, allowing for faster and more comfortable interactions.

4 Micro Gesture Recognition

4.1 Recognition Algorithm

NailRing transmits the fingertip images to a processing program, which is normalized and fed to a neural network classifier. The classifier estimates the probability of occurrence of each micro-gesture. Using supervised learning, we train the network by adjusting the network weights through backpropagation to minimize the cross-entropy loss on the training dataset. The dataset that we construct in our subsequent work includes data for each micro-gesture as well as the labels.

Our classifier is mainly implemented as a deep residual network with DenseNet [9] as the backbone framework. To improve the recognition of fingertip physiological features, the extracted feature maps are subjected to adaptive feature refinement. We improve the network structure by integrating the CBAM [30] module into the

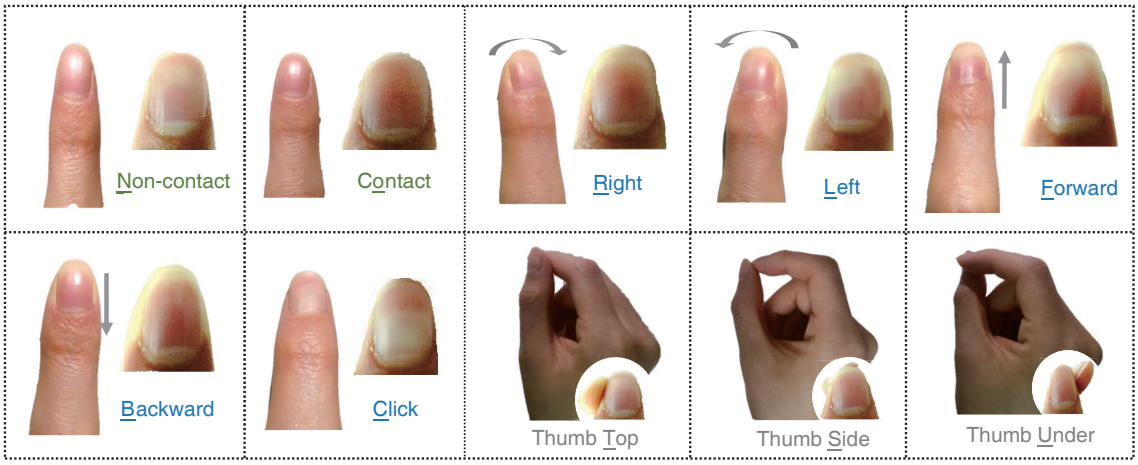

Figure 4: Ten micro gestures are designed based on fingertip interaction features. A schematic diagram of the micro gestures is shown on the left of the figure, and a detailed view of the fingertips during interaction is shown on the right, with the names of the gestures labeled at the bottom. The different colors of the labeled text in the figure represent different categories of gestures: (green) fingertip contact state (blue) flat interaction gestures (black) one-handed interaction gestures. The text checked with '_' in the labeled text is the code for the micro gesture.

layer of the last dense block, which is spatially and channel-wise optimized for high-dimensional features.

Figure 5 illustrates our micro-gesture recognition network structure, which is designed to use small convolutional kernels with a deeper network structure considering the complexity of the physiological features of the fingertips. Compared to shallow networks with large convolutional kernels, deeper networks have stronger fitting ability due to the additional nonlinearity between layers [21], while the successive convolution of multiple small-sized kernels works similarly to large convolutional kernels and also can reduce the computational effort [24]. We use DenseNet as the backbone of the network, which achieves a deeper network structure through residual connectivity, while mitigating the gradient disappearance problem and reducing the number of parameters through feature reuse. In the dense block, each layer takes the output of all previous convolutional layers as input to achieve residual connectivity and feature reuse. In the transition layer, dimension aggregation is performed by convolution and pooling. To perform adaptive optimization of the high-dimensional features of the fingertip information in channel and space, we aggregate CBAM into the last dense block. Figure 5 shows the network structure of the optimized Attention dense block. As a result of the unique residual connection, the Nth layer connects the outputs of all previous layers in dimension $[X_0, X_1, \cdots, X_{N-1}]$ as input, and the network layers are computed as follows:

$$X_N = H_N[X_0, X_1, \cdots, X_{N-1}] \tag{1}$$

$$X_N^{'} = M_c(X_N) \otimes X_N \tag{2}$$

$$X_N^{''} = M_s(X_N^{'}) \otimes X_N^{'} \tag{3}$$

where H_N is defined in the same way as DenseNet-BC, the difference with the previous network is that we pass the output of each layer first through the channel attention module M_c for feature weight proportion optimization, and then through the spatial attention module M_s for feature map location focus, the attention weights are calculated as follows:

$$M_c(X_N) = \sigma(MLP(AvgPool(X_N)) + MLP(MaxPool(X_N))) \tag{4}$$

$$M_s\left(X_N^{'}\right) = \sigma\left(f^{7\times7}\left[AvgPool\left(X_N^{'}\right), MaxPool\left(X_N^{'}\right)\right]\right) \tag{5}$$

The attention module performs AvgPool and MaxPool operations on the input features to aggregate the spatial information of the feature map. The pooled features are input to the multilayer perceptron and then output and summed to obtain the channel attention weight coefficients. The spatial attention module performs AvgPool and MaxPool operations along the feature map dimension to generate two 2D maps, and then passes through a convolution layer to obtain the spatial attention weight coefficients. In the aggregation of high-dimensional features of fingertip information, the channel attention module adaptively selects the weights of different features, while the spatial attention module enhances the important information in the feature map. Through feature extraction in the deep network and feature optimization in the attention module, the network can accurately classify the fingertip information through the output layer.

The complete network includes less than 1.1 million trainable parameters. We implemented the model in Pytorch and trained 300 iterations using the Adam optimizer with a learning rate of 10^{-3}, $\varepsilon = 10^{-8}$ and $\beta = (0.9, 0.999)$, and reduced the learning rate to 10^{-4} starting from the 200th epoch, we optimized these hyperparameters over a micro-gesture dataset spanning multiple users and sessions.

4.2 Dataset Acquisition

4.2.1 Participants

In order to collect the dataset for training the micro-gesture classification network, we recruited 10 volunteers whose ages ranged from 22 to 26 years (MEAN=23.9, SD=1.23). Four of them were female and six were male, and half of them had experience with wearable interactive devices. All were right-handed, and thus, the NailRing can be worn on the participants' right index finger. Each participant was informed of the methodology, and the experiments were processed according to the requirements from the ethics committee in our university.

4.2.2 Acquisition Procedure

We performed data acquisition using a NailRing worn on the index finger. Prior to the acquisition of each micro-gesture, the participant was asked to familiarize with the corresponding gesture, which was then completed on the back of the participant's left hand as well as on nine other contact planes of different colors, materials, and friction coefficients. Under each plane of interaction, the participant was

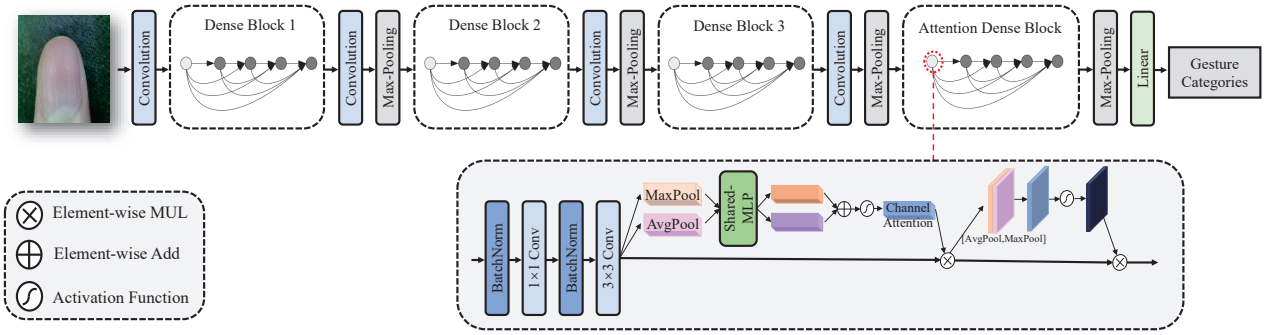

Figure 5: Fingertip micro-gesture recognition network structure. NailRing feeds the acquired fingertip information into the network to recognize gesture categories. The deep neural network is based on the DenseNet backbone, and the CBAM module is integrated into the network to form the Attention Dense Block for spatial and channel optimization of high-dimensional features.

asked to complete the micro-gesture six times at different locations. When participants were fatigued, the experiment could be paused at any time until the participants were ready to continue with the collection. We eliminated ineligible data from the acquisition process and performed a re-acquisition. The total experimental time for each participant was approximately 40 minutes. After completing the acquisition, we obtained a total of 10 micro-gestures * 10 participants * 10 interaction planes * 6 attempts = 6000 micro-gesture datasets with 600 attempts of each gesture.

4.3 Identification Effect Evaluation

In order to evaluate the effectiveness of micro-gesture recognition, we conducted an evaluation experiment using the acquired data. The evaluation experiment is divided into two parts, the first part is called Cross-session test, which focuses on the overall recognition effectiveness of the system, and the second part is called Cross-person test, which tests the recognition effectiveness of micro-gesture recognition for unused users and the usability of the device for new users.

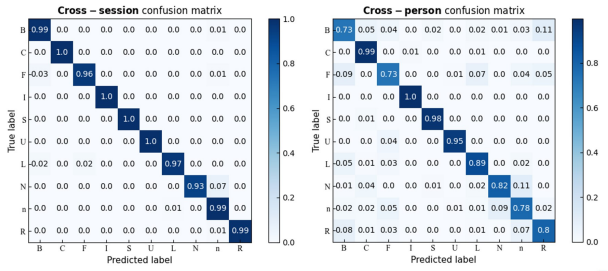

Figure 6: Left: Normalized confusion matrix for 3-fold cross-validation for cross-session identification. Right: 10-fold cross-validation for cross-person identification.

Cross-session Testing In this experimental part, we evaluated the data from all participants as a whole, and we took a 3-fold cross-validation approach, using two blocks for training and one block for validation. Across all participants, the average accuracy for all poses was 98.25% and the F_{macro} score was 98.33%.The confusion matrix on the left of Figure 6 shows the average recognition accuracy of all participants for each gesture. It can be seen from Figure 6 that almost all micro gestures are correctly recognized, which proves that our recognition method is effective and accurate. However, in a real-world scenario, when a new user uses our device without calibration, it is impossible for the recognition algorithm to learn the user's features in advance, so we conducted experiments to test the system's recognition effect on new users to investigate the usability of the micro-gesture recognition system in a real-world scenario.

Cross-person Testing In this experimental part, we performed 10-fold cross-validation, testing on each participant's events and training on all other participants, and the experimental results showed that the average recognition accuracy of all micro gestures was 86.45% and the F_{macro} score was 86.38%. The confusion matrix on the right side of Figure 6 shows the average recognition accuracy of each gesture in the Cross-person case in which one-handed interactive gestures have the highest recognition accuracy (Code I, S, U, average accuracy = 97.6%), because this type of gesture recognition relies on the relative position of the fingers and therefore varies less between users. The gestures with the lowest recognition accuracy were forward and backward swipes (Code B, F, accuracy = 73%), and we hypothesize that this result is caused by the differences in fingertip physiology between users, including external shape, fat content, and nail length. Among the planar interaction gestures, the click operation has the highest accuracy rate (Code C, Accuracy = 99%) because it results in a distinct change in fingertip features (touching the table with the nail causes most of the nail bed to turn white and a certain degree of perspective change occurs during the operation) and is easy to recognize. The highest accuracy of different participants in the Cross-person case is 93.37%, the lowest is 77.8%, the median is 86.5%, and the standard deviation is 0.05. According to the experimental results, the accuracy of micro-gesture recognition in the Cross-person case is high, which proves that it has good usability in the addition of new users and can meet the requirements of interaction.

In this section, we demonstrate the effectiveness of the NailRing prototype for micro-gesture recognition. In a real-world scenario, the model parameters can be optimized by user calibration, i.e., using a small amount of user data to further improve the recognition accuracy. We also indirectly demonstrate the validity of the recognition principle that "the degree of variation in fingertip information because pressure may vary from person to person, but the similar physiological structure makes it applicable to most healthy nails".

5 User Study

The purpose of this user experiment is to evaluate the effectiveness of NailRing recognition in different usage scenarios. In our previous work, we evaluated the effectiveness of micro-gesture recognition, but in practice, users may face different interaction environments, such as different colors of interaction planes and friction, as well as changes in the brightness of the interaction environment lighting. Therefore, we selected four main influencing factors and conducted targeted user experiments to evaluate the usability of NailRing in

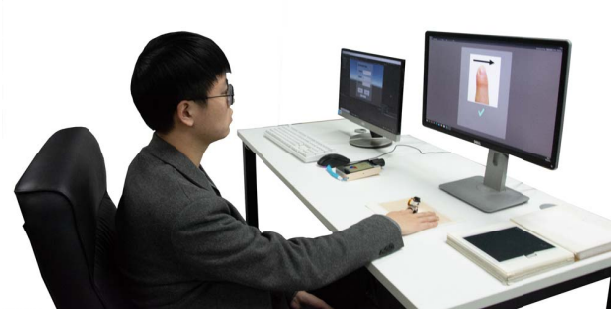

Figure 7: Experimental environment

different environments. The experiments test the recognition of flat interaction micro gestures as well as one-handed interaction micro gestures. We invited participants who performed data acquisition to participate in this part of the experiment and used the previously collected data to train a micro-gesture recognition network to obtain a fingertip micro-gesture classifier. Throughout the experiment, each participant used the same classifier for the experiment.

5.1 Task

Figure 7 shows our experimental scenario, where participants sit comfortably in an office chair in front of a table, place right arm on the table and wear NailRing on right index finger. In each task, the participant is asked to complete the corresponding micro-gesture according to the instructions of the monitor. The experiment is divided into 4 parts as a result of the following influencing factors:

- Color of interaction planes: We measured the average spectral tri-stimulus values (CIE 1931 X: 15.48, Y: 16.06, Z: 8.76) of the fingertip areas of five users using a spectral luminance meter (model: OHSP-350L) and converted them to CIE 1976 L*a*b* color space. The interaction planes were then measured using the same method and seven interaction planes ($\Delta E_{min} = 6.10, \Delta E_{max} = 55.35$) were selected based on the magnitude of the color difference with the fingertip, where the color difference was calculated as follows:

$$\Delta E^* = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \tag{6}$$

  where ΔL^* represents the difference in lightness between the two colors, and Δa^* and Δb^* represent the difference in chromaticity. The selected seven interaction planes are consistent in all parameter indices except color.

- The luminance of the interaction plane: We obtained different lighting environments with adjustable light sources and measured the reflected luminance of the table using an apparatus. Four luminance ($[L_1 = 1.7, L_2 = 45.5, L_3 = 81.6, L_4 = 120.3]$ cd/m^2) were selected for the experiment.

- Materials of the interaction plane: Six different materials of the interaction plane were selected for the experiment, namely Desktop, Reflective material, Stretchy, Smooth fabric, Coarse fabric and Plush fabric.

- Interaction posture: Experiments were performed in sitting and standing positions, and the back of the left hand was used as the plane of interaction in this part of the experiment.

5.2 Procedure

At the beginning of the user study, an experimenter introduced our project and outlined the study's procedure. The experimenter then helped the user wear the ring, ensuring that it was securely and comfortably placed on the correct position on the index finger to avoid unnecessary rotation. Next, the experimenter demonstrated each of the micro-gestures and asked participants to be familiar with them. During the first three parts of the experiment, participants were seated. Each participant was requested to complete a total of 304 random micro-gestures under different experimental conditions. After completing each micro-gesture, the results were displayed on the screen in real time. If the results matched the requested gesture, a correct mark appeared. Otherwise, an incorrect sign was displayed with the incorrect predicted gesture. When recognition was completed, participants were asked to return to the "no-touch" state within two seconds and prepared to recognize the next micro-gesture. Between each test phase, participants had a short break. After the completion of four parts of the experiment, all participants were asked to fill out a questionnaire to collect information about gender, preferred hand use and use of the wearable interaction device. Participants also provided subjective feedback on system usability, comfort and specific comments on the micro-gesture design.

5.3 Result

Throughout the experiment, we recorded the recognition accuracy of NailRing under different experimental conditions as well as the total time for the user to execute the micro-gesture and the micro-gesture recognition pipeline to complete the recognition. During the experiment, participants can complete an input task in 1.14 seconds on average, and the inference time is about 48.3ms.

Color of interaction planes: As shown in Figure 8(a), the average accuracy of micro-gesture recognition was 86.9% (SD=0.04) in the experiments of seven interaction planes with various values of color difference of skin, and the recognition module showed better performance in such case. We found that as the color difference between the interaction plane and the fingertip increased, the recognition accuracy increased to a certain extent and the gesture recognition time also showed a decreasing trend. We believe that the large color difference makes it easier for NailRing to extract fingertip feature information and achieve faster recognition of relatively obscure features earlier in the micro-gesture execution process. We found that anomalous recognition occurred when $\Delta = 30.96$, the reason is that the green channel in the RGB image has a higher weight for fingertip information extraction [5], and the green interaction plane corresponding to $\Delta = 30.96$ interferes with feature extraction to a certain extent. The data in the green interaction plane can be collected as training data to improve the model's and we will optimize this in our subsequent work.

The luminance of the interaction plane: Figure 8(b) shows the experimental results under different luminance conditions, with an average accuracy of 89.5% (SD=0.04) for micro-gesture recognition, and NailRing demonstrates its usability under different luminance conditions. We found the highest recognition accuracy at $L = 41.7 cd/m^2$, a lighting condition that is closest to the reflected light brightness of a desktop in everyday situations. NailRing also recognizes micro gestures well both in bright and dark cases. However, when the ambient brightness is too high, the lighting module of the implemented prototype cannot automatically adjust the brightness, making the contrast of the information collected by NailRing lower and the recognition effect worse. This problem can be solved by hardware optimization.

Materials of the interaction plane: Figure 8(c) shows the results of the experiments with different contact materials. The average accuracy of micro-gesture recognition is 89.4% (SD=0.01), and NailRing can accurately recognize micro-gestures on various materials. In this set of experiments, participants showed relatively larger differences in recognition results, which may be attributed to the different execution of micro-gestures of some participants on different material of interaction planes. For example, higher friction

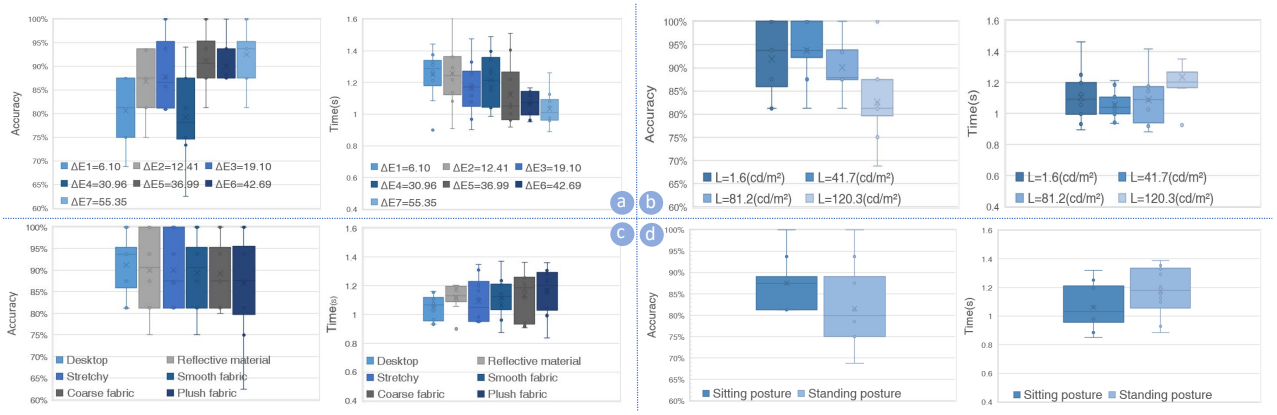

Figure 8: Results of user experiments with four different experimental conditions: (a) Color of interaction planes; (b) The luminance of the interaction plane; (c) Materials of the interaction plane; (d) Interaction posture. The results in each section contain the accuracy and the experiment time. The error line in the figure shows the upper and lower edges of the data, the horizontal line in the box is the median, and the mean of the experimental data is marked by the symbol "X".

sometimes caused some participants to be hindered in the execution of sliding gestures, etc. We also do not recommend using NailRing on Plush fabric because the excessively long lint will obscure the fingertips and affect the recognition effect.

Interaction posture: Figure 8(d) shows the results of the different experiments in standing and sitting postures, with an average accuracy of 84.5% (SD=0.03) for micro-gesture recognition, which is lower because the skin color of the back of the hand being close to the fingertips and unevenness of the back of the hand. The recognition accuracy decreased and became more unstable in the standing situation, which may be caused by the instability of the left arm in the standing situation. The longer total recognition time in standing can be attributed to the longer gesture execution time. This experimental result demonstrates that users can use NailRing to perform interaction tasks on the back of the left hand in the absence of an interaction plane.

Subjective feedback: All participants were willing to interact with NailRing in both MR (MEAN=4.4, MEDIAN=4.5, SD=0.66) and VR (MEAN=4.2, MEDIAN=4.5, SD=0.87) environments, and participants felt that the meaning of the system's micro-gesture categories was clear (MEAN=4.8, MEDIAN=5, SD=0.4). NailRing had good usability according to the System Usability Scale (MEAN=89). Participants completed a seven-point Likert scale on four indicators of micro-gesture, and the results are shown in Figure 9. Participants preferred the "Click" gesture the most, probably because of the large number of repetitive swipes, and the "comfortable" option of "Forward" and "backward" gestures was the least rated. The average score for all gestures was 6.46 (Learnability=6.48, Effective=6.51, Intuitive=6.64, Comfortable=6.19).

6 Application

NailRing can be used in many MR scenarios to provide micro-gesture inputs. In this section, the interaction advantages and application potential of NailRing are first described, and then several applications are implemented as demonstrations to prove the usefulness and scalability of NailRing.

6.1 Application Forms

NailRing's micro gesture recognition can be used as a complementary input to MR interaction systems because of its simplicity, speed and eyes-free, releasing unnecessary functions such as command input from traditional macro gestures and improving the usability and comfort of MR interaction systems.

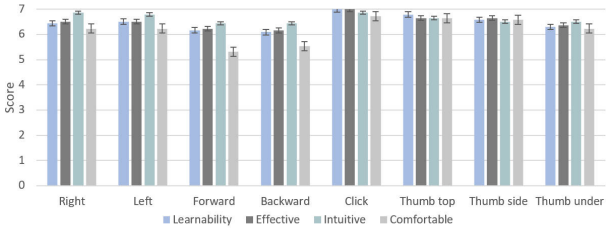

Figure 9: Results of participants' evaluation of the four aspects of each gesture on a 7-point Likert scale (7 being the best and 1 being the worst) and the error line is the standard error.

The proposed NailRing can be used as an interaction area on any surface that the finger with the device touches, such as a desktop, wall or MR device surface. In the walking state, NailRing can be used for interactive input on the back of the assisted hand or on the worn watch, and for sudden interaction needs, such as phone dialing, NailRing can respond quickly with one-handed gestures. What's more, NailRing is easy to learn and intuitive, its gesture meaning basically matches the user's interaction cognition. And the gesture content covers the micro-gesture interaction required to move (Code: "R", "L", "F", "B"), switch (Code: "T", "U"), confirm (Code: "C") and other different functions which can be implemented for a variety of applications such as listening to music, communication and so on. The gestures are also redefined to enable more complex functions, such as activating the voice function and prompting the user when a continuous "T" gesture is detected, and performing voice input and recognition during the gesture duration. Users can use NailRing to achieve natural and comfortable micro gesture interaction.

6.2 Application Scenarios

Based on the set of gestures designed in this paper, we designed and developed several MR applications to demonstrate the potential use of NailRing. The applications include home, video player and contacts, which we chose because they are representative of tasks on MR devices. We used a client-server architecture to implement our application, where the detection algorithm runs on a local server, recognizes real-time information as it becomes available, and sends the results to the application running on the MR HMD via Wi-Fi.

Home: It contains a user interface that is the entry point to other

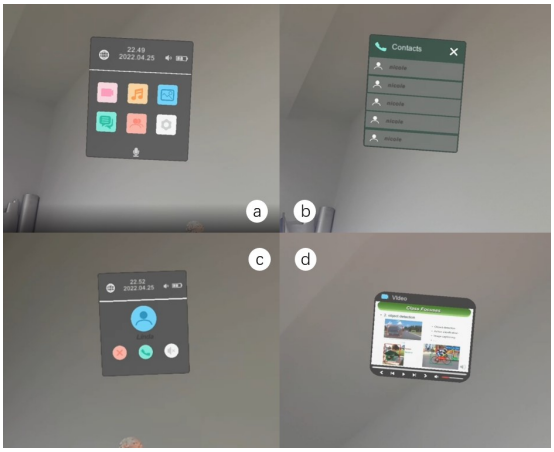

Figure 10: MR applications based on NailRing and micro gestures: (a) home application; (b) contact application; (c) dial-in call application (d) video playback application

applications. We invoke the Home interface with the symbol "S" gesture, and then the user can use the move gesture to select the application icon and enter the application with the confirmation gesture. Our implementation of the Home application contains six icons, of which we will present two applications.

Contacts: The user uses the up and down movement function to browse and select a contact after entering the program, and uses the confirmation gesture to dial out the call. For dial-in calls, we have implemented three one-handed gestures as connect, hang up, and mute functions for quick processing. The volume can be adjusted by swiping the gesture up and down after entering the call screen.

Video Player: We run the user-controlled video player through NailRing to complete operations such as pausing and switching. Specifically, the user starts and pauses video playback by tapping ("C") gestures, switches between videos with "T" and "U" gestures, swipes left and right "R" and "L" are used to fast forward and fast rewind the video, and swiping "F" and "B" up and down is used to adjust the video volume. Unlike the current finger tapping or eye movement method, NailRing supports a simpler eyeless operation.

7 Discussion

NailRing has driven research on wearable interactive devices based on physiological information, however there are a number of issues and directions that need further discussion.

7.1 Potential Application Platform

With the development of IoT, the number of smart home devices has increased dramatically, but these devices use different input methods, thus increasing the complexity of interaction [1, 17]. NailRing can also serve as a unified human-computer interaction interface, providing a simple, consistent interaction experience. NailRing can be connected by scanning the device's identification code, and then directly controls the smart devices through micro gestures. It can also work with MR devices to achieve a richer IoT device interaction experience.

7.2 Limitations and Optimization

In this paper we have implemented a prototype of NailRing, our current implementation still has some limitations. In terms of hardware, the current prototype is limited by the hardware solution, processing and manufacturing, the size of the form can still be optimized. We will continue to optimize the hardware to better meet the interaction needs. In the optimized solution NailRing can integrate WiFi module to communicate with HMD and perform micro-gesture recognition on MR devices. The necessary control of the lighting module could effectively improve NailRing's performance under certain lighting conditions. In order to further enhance the interactive experience, feedback from the interactive device is necessary, and vibration motors can be added to the future implementation to prompt and provide feedback to the user, increasing the user's confidence in using the device. In the next step we will explore dynamic micro gesture sequences for multiple situations to increase the richness of the interaction as well as to implement more functions to further optimize the interaction experience.

7.3 Differences Among Individuals

NailRing relies on fingertip physiological features resulting from interactions to recognize micro-gestures, and each user's fingertip physiology is not exactly the same, such as nail size and fingertip muscle shape. We have demonstrated through previous experiments that fingertip feature information varies among users, but still has uniform and recognizable features. In order to improve the recognition efficiency for a specific individual, a calibration operation can be used to improve the generalization performance of NailRing across users by fine-tuning the parameters of the recognition model with a small amount of data at the first use. We also noticed in our experiments that NailRing does not work properly for users with excessively long nail lengths, because the excessive nail will be in direct contact with the plane causing a change in the force state of the fingertip.

8 Conclusion

In this paper, we introduce NailRing, a wearable intelligent ring that uses fingertip feature information generated during interaction for micro-gesture recognition. It enables eyeless, fast and natural interaction and replaces some of the functions of macro gestures in existing MR interaction schemes to optimize the interaction experience. We design ten targeted micro-gestures, mainly divided into touch detection, flat interaction and one-handed interaction, and the rich gesture meanings increase the usability of NailRing in various situations. We propose a deep neural network-based recognition algorithm and validate the recognition performance and usability across users (F_{Macro}: Cross-session=98.3%, Cross-person=86.4%). We demonstrate the performance of NailRing in various scenarios through a series of user experiments. We then demonstrate the practical usability of NailRing through a custom MR application. Finally we discuss the optimization directions and potential application platforms for NailRing. In summary, our exploration shows that NailRing has great potential for MR HMD implementation and for driving interaction using fingertip physiological information.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61960206007, 61731003, 62002018), the Key-Area Research and Development Program of Guangdong Province (2019B010149001) and the 111 Project (B18005).

References

[1] G. Alce, A. Espinoza, T. Hartzell, S. Olsson, D. Samuelsson, and M. Wallergård. Ubicompass: an iot interaction concept. *Advances in Human-Computer Interaction*, 2018, 2018.

[2] L. Chan, R.-H. Liang, M.-C. Tsai, K.-Y. Cheng, C.-H. Su, M. Y. Chen, W.-H. Cheng, and B.-Y. Chen. Fingerpad: private and subtle interaction using fingertips. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 255–260, 2013.

[3] K.-Y. Chen, S. N. Patel, and S. Keller. Finexus: Tracking precise motions of multiple fingertips using magnetic sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1504–1514, 2016.

[4] T. R. Grieve, C. E. Doyle, J. M. Hollerbach, and S. A. Mascaro. Calibration of fingernail imaging for multidigit force measurement. In *2014 IEEE Haptics Symposium (HAPTICS)*, pp. 623–627. IEEE, 2014.

[5] T. R. Grieve, J. M. Hollerbach, and S. A. Mascaro. Force prediction by fingernail imaging using active appearance models. In *2013 World Haptics Conference (WHC)*, pp. 181–186. IEEE, 2013.

[6] Y. Gu, C. Yu, Z. Li, W. Li, S. Xu, X. Wei, and Y. Shi. Accurate and low-latency sensing of touch contact on any surface with finger-worn imu sensor. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1059–1070, 2019.

[7] A. Gupta, C. Ji, H.-S. Yeo, A. Quigley, and D. Vogel. Rotoswype: Word-gesture typing using a ring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[8] C. Harrison and S. E. Hudson. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pp. 205–208, 2008.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[10] H.-L. Kao, A. Dementyev, J. A. Paradiso, and C. Schmandt. Nailo: fingernails as an input surface. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3015–3018, 2015.

[11] W. Kienzle, E. Whitmire, C. Rittaler, and H. Benko. Electroring: Subtle pinch and touch detection with a ring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.

[12] M. Li, M. Fan, and K. N. Truong. Brailleskatch: A gesture-based text input method for people with visual impairments. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 12–21, 2017.

[13] K. Masai, K. Kunze, D. Sakamoto, Y. Sugiura, and M. Sugimoto. Face commands-user-defined facial gestures for smart glasses. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 374–386. IEEE, 2020.

[14] S. A. Mascaro and H. H. Asada. Photoplethysmograph fingernail sensors for measuring finger forces without haptic obstruction. *IEEE Transactions on robotics and automation*, 17(5):698–708, 2001.

[15] D. McGonagle, A. L. Tan, and M. Benjamin. The nail as a musculoskeletal appendage–implications for an improved understanding of the link between psoriasis and arthritis. *Dermatology*, 218(2):97–102, 2009.

[16] M. Meier, P. Streli, A. Fender, and C. Holz. Tapld: Rapid touch interaction in virtual reality using wearable sensing. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 519–528. IEEE, 2021.

[17] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan. Rio: A pervasive rfid-based touch gesture interface. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pp. 261–274, 2017.

[18] A. Saito, W. Kuno, W. Kawai, N. Miyata, and Y. Sugiura. Estimation of fingertip contact force by measuring skin deformation and posture with photo-reflective sensors. In *Proceedings of the 10th Augmented Human International Conference 2019*, pp. 1–6, 2019.

[19] K. Sakuma, A. Abrami, G. Blumrosen, S. Lukashov, R. Narayanan, J. W. Ligman, V. Caggiano, and S. J. Heisig. Wearable nail deformation sensing for behavioral and biomechanical monitoring and human-computer interaction. *Scientific reports*, 8(1):1–11, 2018.

[20] Y. Sato, J. Inoue, M. Iwase, and S. Hatakeyama. Contact force estimation based on fingertip image and application to human machine interface. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 4219–4224. IEEE, 2019.

[21] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[22] W. Sun, F. M. Li, C. Huang, Z. Lei, B. Steeper, S. Tao, F. Tian, and C. Zhang. Thumbtrak: Recognizing micro-finger poses using a ring with proximity sensing. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, pp. 1–9, 2021.

[23] W. Sun, F. M. Li, B. Steeper, S. Xu, F. Tian, and C. Zhang. Teethtap: Recognizing discrete teeth gestures using motion and acoustic sensing on an earpiece. In *26th International Conference on Intelligent User Interfaces*, pp. 161–169, 2021.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[25] R. Takada, J. Kadomoto, and B. Shizuki. A sensing technique for data glove using conductive fiber. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2019.

[26] Y.-C. Tung, C.-Y. Hsu, H.-Y. Wang, S. Chyou, J.-W. Lin, P.-J. Wu, A. Valstar, and M. Y. Chen. User-defined game input for smart glasses in public space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3327–3336, 2015.

[27] R.-D. Vatavu and L.-B. Bilius. Gesturing: A web-based tool for designing gesture input with rings, ring-like, and ring-ready devices. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 710–723, 2021.

[28] Y. Weng, C. Yu, Y. Shi, Y. Zhao, Y. Yan, and Y. Shi. Facesight: Enabling hand-to-face gesture interaction on ar glasses with a downward-facing camera vision. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021.

[29] A. D. Wilson. Using a depth camera as a touch sensor. In *ACM international conference on interactive tabletops and surfaces*, pp. 69–72, 2010.

[30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[31] R. Xiao, J. Schwarz, N. Throm, A. D. Wilson, and H. Benko. Mrtouch: Adding touch input to head-mounted mixed reality. *IEEE transactions on visualization and computer graphics*, 24(4):1653–1660, 2018.

[32] X. Xu, H. Shi, X. Yi, W. Liu, Y. Yan, Y. Shi, A. Mariakakis, J. Mankoff, and A. K. Dey. Earbuddy: Enabling on-face interaction via wireless earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[33] H.-S. Yeo, J. Lee, H.-i. Kim, A. Gupta, A. Bianchi, D. Vogel, H. Koike, W. Woo, and A. Quigley. Wrist: Watch-ring interaction and sensing technique for wrist gestures and macro-micro pointing. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–15, 2019.

[34] T. Zhang, X. Zeng, Y. Zhang, K. Sun, Y. Wang, and Y. Chen. Thermalring: Gesture and tag inputs enabled by a thermal imaging smart ring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.