CheekInput: Turning Your Cheek into an Input Surface by Embedded Optical Sensors on a Head-mounted Display

Koki Yamashita
Keio University
Yokohama, Japan
kouki@imlab.ics.keio.ac.jp

Takashi Kikuchi
Keio University
Yokohama, Japan
tkiku393760@gmail.com

Katsutoshi Masai
Keio University
Yokohama, Japan
masai@imlab.ics.keio.ac.jp

Maki Sugimoto
Keio University
Yokohama, Japan
sugimoto@imlab.ics.keio.ac.jp

Bruce H. Thomas
University of South Australia
South Australia, Australia
bruce.thomas@unisa.edu.au

Yuta Sugiura
Keio University
Yokohama, Japan
sugiura@keio.jp

ABSTRACT

In this paper, we propose a novel technology called "CheekInput" with a head-mounted display (HMD) that senses touch gestures by detecting skin deformation. We attached multiple photo-reflective sensors onto the bottom front frame of the HMD. Since these sensors measure the distance between the frame and cheeks, our system is able to detect the deformation of a cheek when the skin surface is touched by fingers. Our system uses a Support Vector Machine to determine the gestures: pushing face up and down, left and right. We combined these 4 directional gestures for each cheek to extend 16 possible gestures. To evaluate the accuracy of the gesture detection, we conducted a user study. The results revealed that CheekInput achieved 80.45 % recognition accuracy when gestures were made by touching both cheeks with both hands, and 74.58 % when by touching both cheeks with one hand.

Figure 1: Concept of CheekInput: The HMD recognizes the input made by the user pulling cheeks.

CCS CONCEPTS

• **Human-centered computing → Interaction devices**;

KEYWORDS

Skin Interface, OST-HMD, Photo-reflective Sensor

ACM Reference Format:
Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: Turning Your Cheek into an Input Surface by Embedded Optical Sensors on a Head-mounted Display. In *Proceedings of VRST '17*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3139131.3139146

1 INTRODUCTION

Optical see-through head-mounted displays (OST-HMDs) allow us to interact with augmented reality information in our daily lives. Due to the increasing availability of OST-HMDs, designing input methods for interacting with HMD systems has become important, and various methods have been proposed [Microsoft 2016][Google 2013].

One remarkable technique is detecting gestures by camera devices. For example, HoloLens has a built-in camera system that allows users to input information by recognizing a mid-air hand gesture [Microsoft 2016][Metzger et al. 2004] However, it is difficult to operate if used in a small space since this technique requires a certain distance between the camera system and the user's hands. Google Glass supports interactions by touching the frame of the eyeglasses [Google 2013]. Since this requires users to directly touch the OST-HMD itself, there is a possibility that blurring will occur in the video presented to the user from the accidental movement of the frame.

As a novel input surface, we focus the user touching their face cheek as a mean of interaction with information presented on the OST-HMD. There are several advantages of using cheeks as input surfaces. First, the soft nature of the human cheek provides natural tactile feedback to the movement of the user's fingers on the check surface; this provides an intuitive input modality. Secondly, people naturally touch their faces. Therefore, users may perform interactions without being noticed by other people. Furthermore,

VRST '17, November 8–10, Gothenburg, Sweden

ACM ISBN 978-1-4503-5548-3/17/11…$15.00
https://doi.org/10.1145/3139131.3139146

compared with other skin surfaces, the cheeks are often not covered with clothing.

In this paper, we propose a method of using the cheeks as an input interface (Figure 1). To make an input by touching the cheeks, we used several photo sensors attached to an OST-HMD. We attach sensors to the bottom and side of an OST-HMD that measure the distance to defined cheek points. These sensors with machine learning recognize various gestures by sensing the deformation of the skin. Since they can be integrated into the OST-HMD, the device does not look unfashionable, and there is no need for the user to wear additional sensing devices. Here is a summary of our contributions:

- We implement a device packaged into an HMD to measure the deformation of cheeks by using photo sensors.
- We evaluate the accuracy of a prototyped device under various conditions.
- We design several example application scenarios that use our sensing system.

2 RELATED WORK

2.1 Human Skin as Input Interface

We propose a method of using the human skin as an input interface. There have been a number of investigations into the use of a person's skin as part of an input device. Palms can be an input surface for controlling information [Mistry and Maes 2009]. Gustafson et al. investigated palm-based interaction for future mobile-device interaction [Gustafson et al. 2013]. Liu et al. investigated how people can accept skin interfaces and design new types of devices that are attached on the skin [Liu et al. 2016]. Nicas et al. analyzed how user place their hands on their face [Nicas and Best 2008].

Many methods using human skin as an input interface have been proposed. Nakatsuma et al. devised a method that uses the back of the hand as an input interface. In their wristband device, photo sensors are installed in an array, and the position of the fingertips touching the back of the hand is recognized [Nakatsuma et al. 2011]. Makino et al. proposed a band device equipped with multiple photo sensors. By wrapping this around the userĄfs forearm, the deformation of the skin of the forearm can be measured [Makino et al. 2013]. Ogata et al. enhanced Makino's method to input gestures generated on the forearm such as by pinching or pressing the skin [Ogata et al. 2013]. Harrison et al. developed a system that can recognize tapping on the skin by measuring the vibration propagating on the skin surface [Harrison et al. 2010]. Koyama et al. proposed a multi-touch steering wheel that can recognize hand gestures while driving [Koyama et al. 2014]. Weigel et al. developed a film sensor sheet that is stretchable and adapts to the skin [Weigel et al. 2015]. Weigel et al. also designed an interaction method for mobile devices that use the skin surface of the forearm [Weigel et al. 2014]. A subtle input method that uses hair is also proposed [Vega et al. 2015]. In our previous study, we proposed EarTouch, a sensing technology for ear-based input for controlling applications by slightly pulling the ear and detecting the deformation by an enhanced earphone device [Kikuchi et al. 2017]. EarPut is a device that can detect natural interactions around the ear [Lissermann et al. 2013]. Compared with these pieces of research, our research is focused on the cheeks.

Serrano et al. explored the use of hand-to-face interaction for controlling HMDs [Serrano et al. 2014]. In their studies, ears were also defined as a gesture area. Their studies use optical cameras and markers for tracking, and compared with theirs, we propose a more compact device.

2.2 Detecting Cheek Movement

Measuring cheek movement is particularly important when measuring chewing. Koizumi et al. measure the movement of the jaw by using a photo sensor [Koizumi et al. 2011]. Tongue-in-Cheek is a sensor system that recognizes the motion of the cheek being pushed out by the tongue[Goel et al. 2015]. These systems are not packaged in eyeglasses-type equipment. Also, our research is focused on input methods that deform the skin with human fingers.

2.3 Input Method for HMDs

Various methods of operating HMD have been proposed. With Google Glass, a touch input sensor makes it possible to make inputs with the fingers [Google 2013]. There is also research in which a touch sensor is placed on the front side or on the side of the HMD, and the user touches these sensors with his or her finger to manipulate information [Gugenheimer et al. 2016][Kato and Miyashita 2015]. There is a problem in that the image projection is disturbed in some cases since these methods directly touch the HMD.

HoloLens allows users to manipulate information by using mid-air gestures [Microsoft 2016]. Anusha et al. proposed a method to recognize aerial gestures made around the HMD that are picked up by several photo sensors [Withana et al. 2015]. Users can track the position of the hands by attaching LeapMotion to the front of an HMD [LeapMotion 2012]. Ishii has proposed the input method of the air gesture for the HMD using a smartphone. The system measures the position and gesture of a finger using a camera incorporated in a smartphone [Ishii et al. 2017]. In these systems, occlusion problems occur, and there are situations where it is difficult to operate in a narrow space.

In addition, some studies have applied eye movement as an input method. Tag et al. use blinking to control visual information [Tag et al. 2016]. Špakov proposed an input method that combines gaze and head motion [Špakov and Majaranta 2012]. With this type of method, the direction of the eyes is restricted.

2.4 Photo Sensing on Head-mounted Display

There is an attempt to reconstruct human facial expressions using camera and several sensors [Li et al. 2015]. Also, methods have been proposed that measure facial expressions from human skin deformation by incorporating multiple photo sensors in an HMD or spectacles. Masai et al. proposed a wearable device that can identify human facial expressions by placing several photo sensors in the frame of glasses[Masai et al. 2016]. Suzuki et al. applied Masai's method to detect the facial expressions of a user inside an HMD [Suzuki et al. 2017]. Nakamura et al. measured the deformation between the eyebrows by using a light sensor and applied it to information manipulation [Nakamura and Miyashita 2010]. In addition, another developed system utilizes photo-reflective sensors attached to a set of earphones, and the sensors recognize which

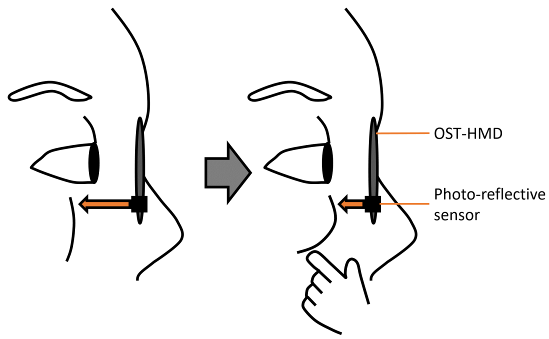

Figure 2: Principle of the proposed method. The sensors attached to the HMD measure the skin deformation of the cheek.

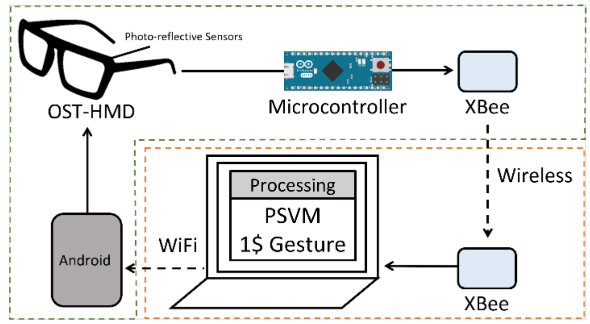

Figure 3: System configuration.

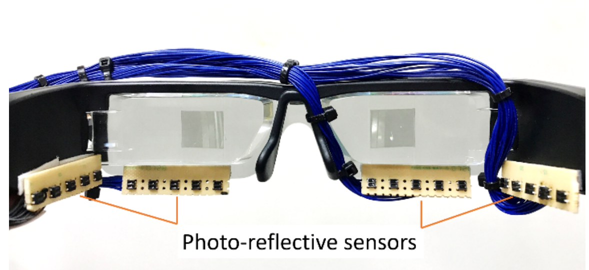

Figure 4: CheekInput device.20 photo-reflective sensors are attached to the frame of glasses.

earphones are fitted to which ear and appropriately provide sounds to the left [Matsumura et al. 2012]. Saponas et al. proposed a tongue input device for controlling information [Saponas et al. 2009]. Several photo sensors are embedded into the device and detect the motion of a user's tongue. Although our proposed device is similar to these studies as a system configuration, it accepts explicit input like touches made on the cheeks.

3 CHEEKINPUT

3.1 Principle

In this study, cheek deformation is measured by an array of photo sensors mounted on an HMD. We employ photo-reflective sensors, and they are composed of an infrared LED light source and a phototransistor light detector. The sensors are generally used for measuring the distance between a sensor and an object. Inspired by the work of Masai et al. [Masai et al. 2016], we embedded these sensors in the HMD frame to measure the distance from the frame to the cheek. When a user touches his/her cheek, force is applied to the skin surface, and it makes a deformation (Figure 2). As a result, the distance between the skin surface and a sensor is changed, and by applying a machine learning technique for multiple photo sensor values, it is possible to detect hand gestures on the cheek. The detection result can be used for human computer interactions.

3.2 Hardware

Our system consists of a head-mounted device and laptop PC (Figure 3). They are connected wirelessly through a wireless module XBee and WiFi. We built a device integrated with an OST- HMD (Figure 4), the EPSON MOVERIO BT-200. Five photo sensors were placed under each image projection plane, and five photo sensors were placed to the sides of the eyes. Since we installed sensors on both sides of the head, we could measure the deformation of both the left and right cheeks. We used the Kodenshi SG-105 photo sensor. The sensors were connected to an Arduino Pro Mini 3.3 V microcontroller, and data was sent to the PC through XBee. Three multiplexers were used to switch the reading of the 20 photo-reflective sensors. The data of the projection images for the OST-HMD were generated by a laptop PC and transmitted by WiFi to an Android device connected to the OST-HMD. Since the battery is integrated, there was no need to extend the wiring to the outside.

3.3 Directional Gesture Recognition

The system recognizes gesture by using the data obtained from the photo sensors. We used a Support Vector Machine (SVM), which is one of the most common supervised machine learning methods for gesture recognition. For the implementation, we used the SVM for Processing (PSVM) [Makematics 2012]. The first step is a direction data set is prepared. The user wears the device, and learning data is accumulated by recording the data of the 20 photo-reflective sensors a hundred times when the cheek is dragged up and down, right and left with the fingers, respectively (Figure 5). After learning, by recording the same gesture, it becomes possible to recognize the gesture of pulling the cheek upward, downward,leftward and rightward. This system provides the probability of how similar the input is to each basic direction. On the basis of the probability, each direction is weighted. This enables the system to recognize not only the four basic directions but also intermediate directions on a 2D plane.

As previously mentioned, sensors are placed on both sides of the OST-HMD so that the deformation of both cheeks can be measured. Therefore, it is possible to input various gestures by using both cheeks (Figure 6). For both cheeks, the user makes inputs with both hands. Alternatively, the user can input commands while touching both cheeks by using the index finger and thumb of one hand. Since it is possible to input four directions on each of the left and right

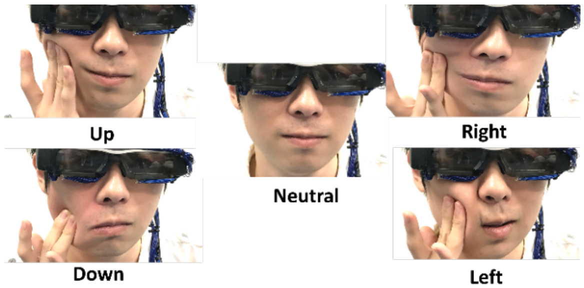

Figure 5: Five directional gesture.

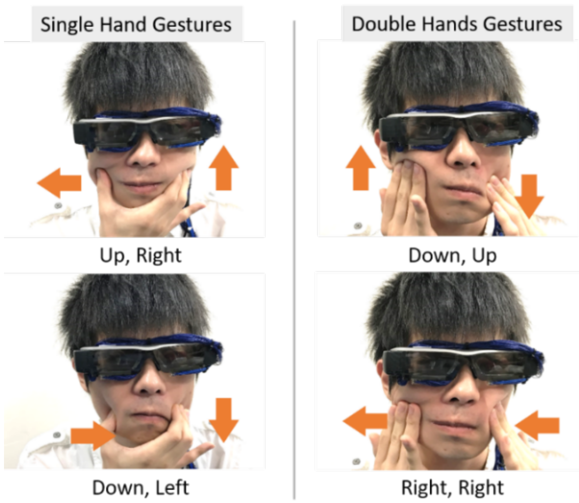

Figure 6: Example of double-side gestures. We designed two types of gestures: gestures generated by single hand (left) and gestures generated by double hands (right).

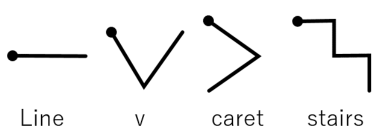

Figure 7: Four symbolic gestures.

cheeks, it is possible to input a total of 16 gestures in combinations. Figure 6 depicts examples of the two cheek input method.

3.4 Symbolic Gesture Recognition

Using the directional gestures as a base, the user can input more complicated symbolic gestures (Figure 7). The amount and the direction in which the cheek is pulled from the origin can be calculated by sampling direction data. From this, the user can draw symbolic gestures that can be used as input.

We implemented a system that can recognize several symbol gestures, with the SVM and $1 Recognizer for gesture recognition [Wobbrock et al. 2007]. The SVM is first used to recognize the direction in which the cheek is pulled. Then, from the directional input, a stroke input is created by plotting 2D points. The 2D points are calculated by adding the unit vector of the direction to the previous 2D point. The end of a gesture can be determined by recognizing a ĄgneutralĄh condition for a set of duration. Sending all the 2D points to the $1 Recognizer makes it possible to use the sensor data chronologically, which expands the variation of gestures. In this paper, we recognized four symbolic gestures.

4 EVALUATING SINGLE-SIDE GESTURES

4.1 Accuracy Concerning Directional Gestures

A user study was conducted to investigate the recognition accuracy of the four directional gestures. Our study aimed to understand the accuracy of the system in different daily scenes. Participants were instructed to hold their cheek and drag it in the four directions (up, down, right, and left) and just hold it (neutral direction) without deforming the cheek itself.

The participants performed the experiment in three conditions (sitting, walking, and re-wearing). The walking experiment was completed indoors in a room, in which the participants walked around the room in random directions and velocities. They were instructed to walk as usual without remaining in a particular place. In the re-wearing condition, after each trial, the participants were instructed to momentarily remove the CheekInput from their face and re-wear the device.

Participants were asked to use their dominant hand to touch the cheeks for input. We collected data when their dominant hand touched the cheek on the same side as the dominant hand and opposite side of the dominant hand.

Before collecting the data, the participants were instructed as to where to touch their cheek and were told to practice moving their cheek in each direction only once. They were not instructed to pull their cheek with a certain force, and they informed wear CheekInput as they would any glasses. They were told to drag the cheek in one of the directions for 5 seconds. The sampling rate of the sensors was set to 30 fps, and sensor data was collected for 100 frames per subject. The 100 frames in the middle of the 5 seconds were used for training. These procedures were defined as one trial. In total, 15,000 sets of data were collected per participant, i.e., 5 directions $\times$ 100 frames $\times$ 5 trials $\times$ 3 situations $\times$ 2 cheeks = 15,000 sets of data. The order of conditions was randomized per participant.

The participants included seven men and one woman in their 20s. One dataset collected from each participant was subjected to five-fold cross validation. The sensor data of four trials was used as a training dataset, and one trial was used as a test dataset. The training dataset for each participant was subjected to the SVM with a linear kernel. The training and the cross-validation were performed for each participant. It was necessary to perform training for each participant because the shape of peoples' cheeks differs.

According to the results for the dominant-hand-side cheek (see Figures 8 and 9), the average recognition accuracy was 89.86% (SD: 7.98%) when the participants were sitting. This accuracy was higher than that in the case of walking (Average: 82.64%, SD: 16.08%). One reason for this discrepancy is the decreased recognition accuracy in the case of the neutral condition while walking compared with that in the case of sitting. This was because the cheeks of the users

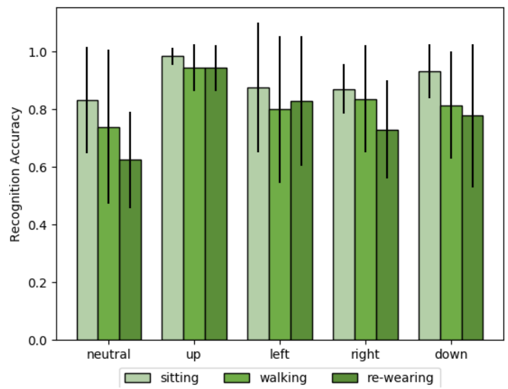

Figure 8: Recognition accuracy when a dominant hand touched cheek on the same side as dominant hand.

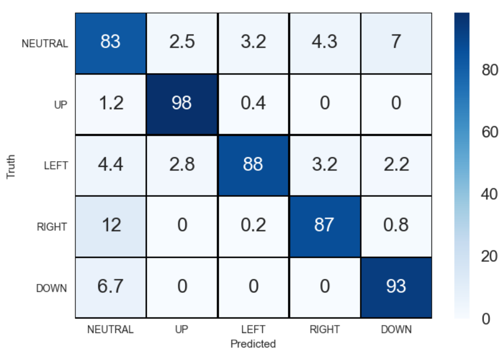

Figure 9: Confusion matrix of recognition accuracy for sitting when a dominant hand touched cheek on the same side as dominant hand.

tended to move through the vibration caused by the movement of the body. Per the results for re-wearing, the average recognition accuracy for directional gestures was 78.07% (SD: 8.77%). This result shows that it is not required to collect more data for the training dataset after CheekInput is repeatedly re-worn, and this demonstrates that the position in which CheekInput feels comfortable is constant. For some participants, the recognition accuracy after they re-worn CheekInput was higher than conditions of sitting and walking. For those participants, CheekInput often tended to slip on their face. This problem was caused by the weight of the device, which can be solved by building the sensors into the HMD in the future.

According to the results for the non-dominant-hand-side cheek (see Figure 10), the average recognition accuracy was 90.76% (SD: 7.26%) when the participants were sitting. The recognition accuracy was higher than that in the case of walking (Average: 83.58%, SD: 11.63%). Per the results for re-wearing, the average recognition accuracy for directional gestures was 91.50% (SD: 7.21%). There

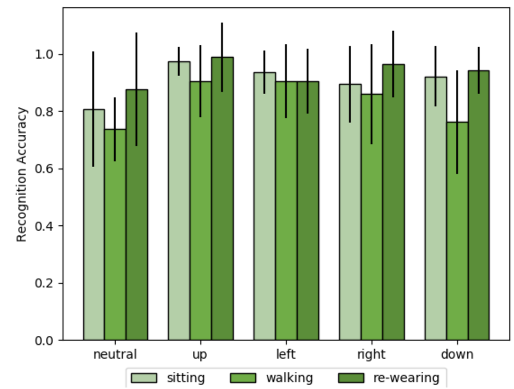

Figure 10: Recognition accuracy when a dominant hand touched cheek on the same side as dominant hand.

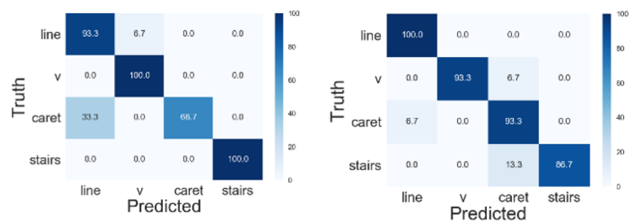

Figure 11: Confusion matrix of recognition accuracy of symbolic gestures with visual aids (left) and without visual aids (right).

was no large difference in accuracy between the conditions of the dominant side and non-dominant side.

4.2 Accuracy Concerning Simbolic Gestures

Another user study was conducted to investigate the recognition accuracy for symbolic gestures. Participants were instructed to input the symbolic gestures shown in Figure 7. They sat in front of a display and performed the same gestures as those shown randomly on the display. This experiment was performed under two conditions: with visual aid and "eyes free". For the first condition, the trajectory of the participants' gesture input was showed, and the participants could redo their gesture input if they were not satisfied. For the second condition, the visual aid was hidden, and the participants had to trust their intuition and were not allowed to repeat their gesture. Before the experiment, the participants practiced for a couple of minutes without being shown the recognition result of their input. For this experiment, the SVM was first trained with 100 sensor data per direction. In the training step, the participants were to push the cheek slightly in the direction instructed and stay in the position while collecting the sensor data. The frame rate of the system was set to 30fps. The participants included 3 men in their 20s.

Mean recognition accuracy for symbolic gestures was 91.66% (SD: 8.98%) (see Figure 11). Recognition accuracy was higher without

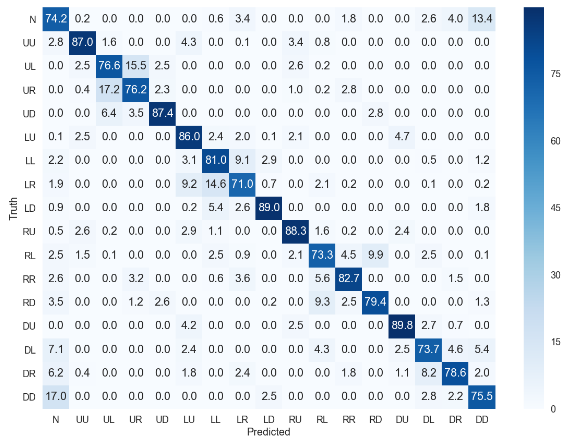

Figure 12: Confusion matrix of recognition accuracy of double side gestures with double hands when sitting.

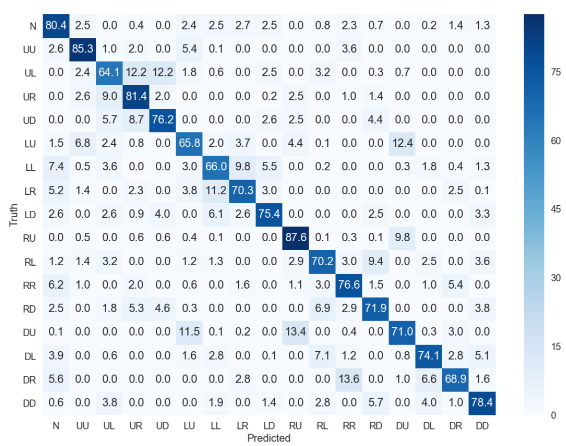

Figure 13: Confusion matrix of recognition accuracy of double side gestures with a dominant hand when sitting.

visual aids (Average: 93.33%, SD: 6.24%) compared to that with visual aids (Average: 90.00%, SD: 10.80%). This discrepancy occurs because many of the participants were trying to improve their gesture input and were going back and forth making the input more complicated. On the other hand, in the eyes-free condition, the user could not see the trajectory that made them want to create a simpler input.

5 EVALUATING DOUBLE SIDE GESTURES

In this study, the precision when gestures were input by using both the left and right cheeks at the same time was obtained. Gestures were made by touching both cheeks with both hands and by touching both cheeks with one hand. The participants included seven

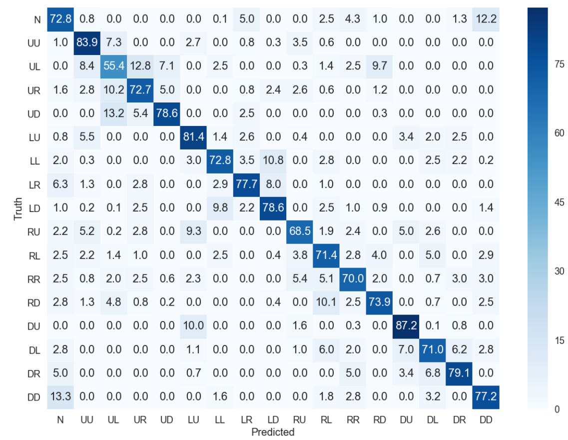

Figure 14: Confusion matrix of recognition accuracy of double side gestures with a non-dominant hand when sitting.

men and one woman in their 20s. One dataset collected from each participant was subjected to five-fold cross validation. The sensor data of four trials was used as a training dataset, and one trial was used as a test dataset. The training dataset for each participant was subjected to the SVM with a linear kernel. The training and cross-validation were performed for each participant.

The confusion matrix on inputs for both cheeks by both hands in sitting condition is shown in Figure 12. The total accuracy of all participants was 80.45% (SD: 10.12%). This result shows a good recognition rate.

The confusion matrix for when inputs were made on both cheeks with one dominant hand is shown in Figure 13. The total accuracy of all participants was 74.33% (SD: 4.80%). The confusion matrix for when inputs were made on both cheeks with one non-dominant hand is shown in Figure 14. The total accuracy of all participants was 74.83% (SD: 6.65%). There was no large difference in accuracy in between the conditions for the dominant hand and non-dominant hand. The total accuracy of both conditions one dominant hand and one non-dominant hand was 74.58%.

Based on the results, the total accuracy of the two hands gestures is higher than the total accuracy of the one hand gestures. This is because the manipulability of fingers in one hand gesture is difficult in some specific gestures.

6 APPLICATIONS

In this section, we show four example applications to demonstrate the benefit of our prototype system. Photo viewing and Music are example apps controlled by one hand gesture. Map and Character Input are example apps controlled by double-cheek gestures.

6.1 Photo Viewing Application:

We show a concept of photo viewing application that can be operated by using our device (see Figure 15). For the application, we used directional left and right gestures to flip through photos.

6.2 Music Application:

We created a music application that can be operated by using our device (Figure 16). For the application, we used the same gestures as shown in Figure 16. Users can change the music and volume. For the volume, users can adjust the gain by pulling the cheek up and down. Our method allows the user to operate the application eyes-free.

6.3 Map Application:

Our device allows users to control applications such as controlling the HMD. We created a map application that can be operated by using our device (Figure 17). Users can scroll the map by pushing the cheek in the relative direction of the cheek movement. A zoom in and out function for the map is performed by pinching both cheeks in and out. Compared with the mid-air-gesture input technique, our method does not require the user to hold their hand in front of their body; therefore, it is suitable for outdoor use in crowded areas.

6.4 Character Input:

Our proposed method can be used to enter letters by making flicking gestures. The user roughly chooses letters on one side of the cheek (Figure 18 left) by the directional gesture. Then, the user selects the character which he/she want to actually input on another side of the cheek (Figure 18 right). This form of interaction requires further investigation. This technique potential could be employed for hierarchical menu navigation.

7 LIMITATIONS AND FUTURE WORK

From the user study, the precision accuracy for participants who had a slim face appeared to be low compared with the other participants. Only a slight change in skin deformation was measured when the participants touched their cheeks. This problem can be solved by adjusting the sensitivity of the photo sensors. A second limitation that came up in the user study was that the women participants wearing makeup tended to dislike touching their cheeks. This is because their makeup will smudge when they touched their cheek for a long time. Therefore, our device might not be suitable for women wearing makeup. In the user study section, we showed that our device can recognize gestures when walking. However, when users moved with intensity, such as running or jumping, the frequency of false positives increased. For these situations, it is possible to use other sensors, for example, acceleration sensors, to recognize the state of the user.

An additional limitation is that the photo sensors measure distance by using infrared light, so the recognition accuracy might decrease in places under intense sunlight. In this research, when facial expressions were changed, there was a low rate of facial expressions falsely recognized as gesture inputs. However, when registering a gesture in such a state that the skin does not move so much, a change in facial expression is erroneously recognized as a gesture.

In this paper, we introduced a technique for operating HMDs by using the cheeks as an input surface. One reason for this is that the cheek is a place where people commonly touch, for example, when thinking about something. However, we have not investigated the incompatibility of touching the cheek for a long time with input gestures. To investigate social acceptance, we will refer to an investigation carried out by Julie et al. [Rico and Brewster 2010]. We would also like to investigate private and subtle interactions.

Figure 15: Photo viewing application.

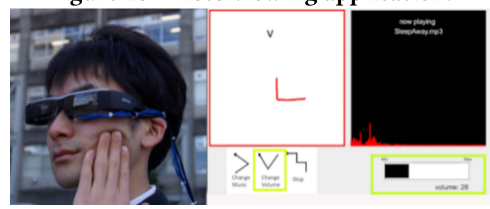

Figure 16: Music Application controlled by our device.

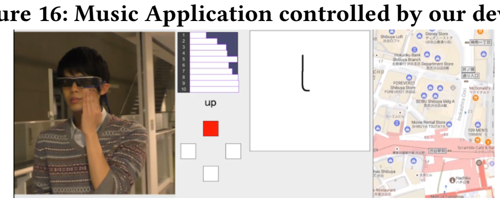

Figure 17: Map application controlled by making a gesture on the side.

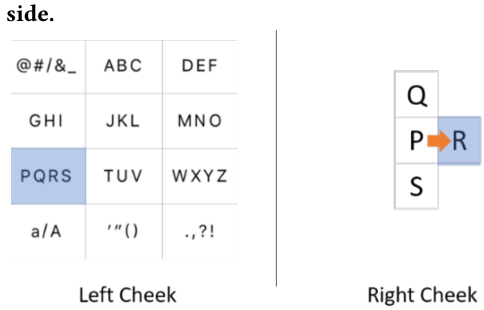

Figure 18: Concept image of character input application.

8 CONCLUSION

In this paper, we proposed CheekInput, a novel interaction method that turns the cheeks into an input interface for HMDs. Photo-reflective sensors were attached onto the frame of an HMD and measure the deformation of the skin that occurs when the cheeks are slightly pulled. This device accepts touches made with one hand on one cheek, touches made with both hands on both cheeks, and touches made with one hand on both cheeks. Our system can recognize gesture inputs made by pulling the cheeks in four directions. We combined these 4 directional gestures for each cheek to extend 16 possible gestures.

The results revealed that CheekInput achieved 80.45% recognition accuracy when gestures were made by touching both cheeks with both hands, and 74.58 % when by touching both cheeks with one hand.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP26700017 and JP16H01741.

REFERENCES

Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 255–258. https://doi.org/10.1145/2702123.2702591

Google. 2013. Google Glass. (2013). https://www.google.com/glass/start/

Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. Facetouch: Enabling touch interaction in display fixed uis for mobile virtual reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 49–60. https://doi.org/10.1145/2984511.2984576

Sean G Gustafson, Bernhard Rabe, and Patrick M Baudisch. 2013. Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 889–898. https://doi.org/10.1145/2470654.2466114

Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–462. https://doi.org/10.1145/1753326.1753394

Akira Ishii, Takuya Adachi, Keigo Shima, Shuta Nakamae, Buntarou Shizuki, and Shin Takahashi. 2017. FistPointer: Target Selection Technique using Mid-air Interaction for Mobile VR Environment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 474–474. https://doi.org/10.1145/3027063.3049795

Kunihiro Kato and Homei Miyashita. 2015. Creating a mobile head-mounted display with proprietary controllers for interactive virtual reality content. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 35–36. https://doi.org/10.1145/2815585.2817776

Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H Thomas. 2017. EarTouch: turning the ear into an input surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 27. https://doi.org/10.1145/3098279.3098538

Naoya Koizumi, Hidekazu Tanaka, Yuji Uema, and Masahiko Inami. 2011. Chewing jockey: augmented food texture by using sound based on the cross-modal effect. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. ACM, 21. https://doi.org/10.1145/2073370.2073387

Shunsuke Koyama, Yuta Sugiura, Masa Ogata, Anusha Withana, Yuji Uema, Makoto Honda, Sayaka Yoshizu, Chihiro Sannomiya, Kazunari Nawa, and Masahiko Inami. 2014. Multi-touch steering wheel for in-car tertiary applications using infrared sensors. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 5. https://doi.org/10.1145/2582051.2582056

LeapMotion. 2012. Leap Motion. (2012). https://www.leapmotion.com/

Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 47.

Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. 2013. Earput: Augmenting behind-the-ear devices for ear-based interaction. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1323–1328. https://doi.org/10.1145/2468356.2468592

Xin Liu, Katia Vega, Pattie Maes, and Joe A Paradiso. 2016. Wearability factors for skin interfaces. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 21. https://doi.org/10.1145/2875194.2875248

Makematics 2012. *Support Vector Machines for Processing*. Makematics. http://makematics.com/code/psvm/.

Yasutoshi Makino, Yuta Sugiura, Masa Ogata, and Masahiko Inami. 2013. Tangential force sensing system on forearm. In *Proceedings of the 4th Augmented Human International Conference*. ACM, 29–34. https://doi.org/10.1145/2459236.2459242

Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial expression recognition in daily life by embedded photo reflective sensors on smart eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 317–326. https://doi.org/10.1145/2856767.2856770

Kohei Matsumura, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2012. Universal earphones: earphones with automatic side and shared use detection. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. ACM, 305–306. https://doi.org/10.1145/2166966.2167025

Christian Metzger, Matt Anderson, and Thad Starner. 2004. Freedigiter: A contact-free device for gesture control. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, Vol. 1. IEEE, 18–21. https://doi.org/10.1109/iswc.2004.23

Microsoft. 2016. HoloLens. (2016). https://www.microsoft.com/microsoft-hololens/

Pranav Mistry and Pattie Maes. 2009. SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*. ACM, 11. https://doi.org/10.1145/1667146.1667160

Hiromi Nakamura and Homei Miyashita. 2010. Control of augmented reality information volume by glabellar fader. In *Proceedings of the 1st Augmented Human international Conference*. ACM, 20. https://doi.org/10.1145/1785455.1785475

Kei Nakatsuma, Hiroyuki Shinoda, Yasutoshi Makino, Katsunari Sato, and Takashi Maeno. 2011. Touch interface on back of the hand. In *ACM SIGGRAPH 2011 Emerging Technologies*. ACM, 19. https://doi.org/10.1145/2048259.2048278

Mark Nicas and Daniel Best. 2008. A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection. *Journal of occupational and environmental hygiene* 5, 6 (2008), 347–352.

Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. SenSkin: adapting skin as a soft interface. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 539–544. https://doi.org/10.1145/2501988.2502039

Julie Rico and Stephen Brewster. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 887–896. https://doi.org/10.1145/1753326.1753458

T Scott Saponas, Daniel Kelly, Babak A Parviz, and Desney S Tan. 2009. Optically sensing tongue gestures for computer input. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 177–180. https://doi.org/10.1145/1622176.1622209

Marcos Serrano, Barrett M Ens, and Pourang P Irani. 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3181–3190. https://doi.org/10.1145/2556288.2556984

Oleg Špakov and Päivi Majaranta. 2012. Enhanced gaze interaction using simple head gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 705–710. https://doi.org/10.1145/2370216.2370369

Katsuhiro Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto. 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 177–185. https://doi.org/10.1109/vr.2017.7892245

Benjamin Tag, Junichi Shimizu, Chi Zhang, Naohisa Ohta, Kai Kunze, and Kazunori Sugiura. 2016. Eye blink as an input modality for a responsive adaptable video system. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 205–208. https://doi.org/10.1145/2968219.2971449

Katia Vega, Marcio Cunha, and Hugo Fuks. 2015. Hairware: the conscious use of unconscious auto-contact behaviors. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 78–86. https://doi.org/10.1145/2678025.2701404

Martin Weigel, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle. 2015. Iskin: flexible, stretchable and visually customizable on-body touch sensors for mobile computing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2991–3000. https://doi.org/10.1145/2702123.2702391

Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More than touch: understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188. https://doi.org/10.1145/2556288.2557239

Anusha Withana, Roshan Peiris, Nipuna Samarasekara, and Suranga Nanayakkara. 2015. zsense: Enabling shallow depth gesture recognition for greater input expressivity on smart wearables. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3661–3670. https://doi.org/10.1145/2702123.2702371

Jacob O Wobbrock, Andrew D Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. ACM, 159–168. https://doi.org/10.1145/1294211.1294238