Digital Full-Face Mask Display with Expression Recognition using Embedded Photo Reflective Sensor Arrays

Yoshinari Takegawa[*,*] Future University Hakodate
Yutaka Tokuda[†,*] Freelance
Akino Umezawa[‡] Future University Hakodate
Katsuhiro Suzuki, Katsutoshi Masai, Yuta Sugiura, Maki Sugimoto[§] Keio University
Diego Martinez Plasencia, Sriram Subramanian[¶] University College London
Keiji Hirata[‖] Future University Hakodate

Figure 1: An oblique view image of a digital full-face mask display and the series of avatar animations reflecting the transition of the facial expression of the wearer

Abstract

This paper presents a thin digital full-face mask display that can reflect an entire facial expression of a user onto an avatar to support augmented face-to-face communication in real environments. Although camera-based facial expression recognition technology has enabled people to augment their faces with avatars, application was limited to face-to-face communication in virtual environments. To enable digital facial augmentation with an avatar in a real space, we propose a digital face mask display system that integrates a lightweight flexible display with a thin facial expression recognition system. The thin wearable facial expression recognition system was implemented with photo reflective sensor arrays which can measure facial expressions at 40 feature points distributed across an entire face. We investigated a ten-class facial expression identification model based on an SVM training algorithm. The trained model achieved an average accuracy of 79% when identifying the facial expressions of multiple users. User experiments indicated that the proposed thin digital full-face mask display allows the wearer to control the facial expression of the avatar with a fast response rate and create a positive sense of self-agency and self-ownership toward the augmented avatar face.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

[*]: yoshi@fun.ac.jp
[†]: yutakamitsue@gmail.com
[‡]: g2217001@fun.ac.jp
[§]: {katsuhirosuzuki, masai, sugiura, sugimoto}@imlab.ics.keio.ac.jp
[¶]: {d.plasencia, s.subramanian}@ucl.ac.uk
[‖]: hirata@fun.ac.jp
[*]: These authors contributed equally to this work

1 Introduction

The face is the most expressive part of our body and is one of the important parts that decides the impression a person makes on others. The psychologist Mehrabian [20] states that the factors in deciding the first impression made by someone at a first meeting are made up of visual information (55%), aural information (38%) and verbal information (7%). Secord [30] asserts that, in face-to-face communication, each person makes assumptions about the other person's personality, based on facial features. For instance, Kawanishi [14] showed the positive correlation between a person having pleasing facial features and other people feeling an affinity with the person. In particular, facial expressions are one of the most important elements of face-to-face communication, and reveal internal emotional states and implicitly create the persona of an individual as perceived by others. Facial expressions are, however, difficult to control without special training and thus not flexible as a self-representation method.

To create a desired persona that exceeds the physical limits of human faces, people have used digital avatars as more effective and flexible media to control the impression made on others in a VR environment. For example, in a remote video communication system, a speaker can digitally augment their real face by altering the appearance of the 3D avatar on a computer screen to look the way they want it to. Then, the speaker can flexibly manipulate the avatar's facial expression through a camera-based facial recognition system. On the other hand, researchers have investigated a digital face augmentation technology to further extend the opportunity for digital self-representation with avatars in a real environment [1, 7, 11, 15, 26, 36]. For instance, Akaike et al. proposed an AR-based face augmentation Head Mounted Display (HMD) that superimposes a still image of an avatar onto the face of an interlocutor seen by the wearer, which can be seen via the video see-through display. Such a camera-based digital face augmentation method, however, requires everyone participating in the communication to wear HMDs to be able to see the virtually transformed face, hindering face-to-face communication with ordinary people in a real environment.

Figure 2: Application of a digital full-face mask: (a) Augmenting the facial expressions of patients with amyotrophic lateral sclerosis (ALS) or mental disorders, (b) Reducing sense of fear in the treatment of child patients and enhancing the friendliness of the mask wearer (e.g. dentist), (c) Changing the face of an interviewer to ease an interviewee's nerves, (d) Changing the default expression of staff working in customer service, to reduce emotional labor, (e) Virtually changing a performer's face to make rehearsal conditions closer to final performance, (f) Changing online gamers' faces to look like their profile pictures even when offline

To enable augmented face-to-face communication in diverse ubiquitous contexts of real environments, we propose e2-MaskZ, a thin digital full-face mask display with facial expression recognition capability. e2-MaskZ allows a user to directly augment their real face with an avatar and intuitively manipulate the avatar's facial expression, by recognizing the wearer's facial expression behind the mask display in real time (Figure1). As the vision of our digital thin mask display to enable direct facial augmentation, we aim to expand new communication opportunities in daily life through the improvement of self-confidence and flexibility of facial expression control (Figure2). For instance, a digital facial mask display could help patients with amyotrophic lateral sclerosis (ALS) to produce a variety of rich facial expressions (Figure2(a)), reduce the sense of anxiety and stress during the dental treatment of child patients (Figure2(b)), and allow an interviewer to conduct a relaxed job interview by controlling the impression the interviewer's face makes on an interviewee (Figure2(c)). As a flexible medium to express the user's optimal persona in various contexts, a digital face mask display could facilitate the reduction of emotional labor in hospitality work (Figure2(d)), extend the freedom of performance and expressiveness in theatre (Figure2(e)), and help users of online games or forums to maintain their online digital personas at offline meetings (Figure2(f)).

In this paper, we demonstrate the proof of concept of a thin digital full face mask display for facial augmentation, which uses a flexible display on the front side for avatar visualization and integrates photo-reflective sensor arrays on the reverse side for millimeter short range facial expression recognition. We describe the SVM-based training algorithm used to build a facial expression identification model and report the robustness of the model regarding its ability to correctly identify ten classes of facial expression for multiple users. Through the subjective evaluation of usability, we demonstrate the feasibility of the proposed digital full-face mask display in terms of the wearer's sense of self-agency and self-ownership. Finally, we discuss the technical limitations of the display and the facial recognition system and describe the solutions to achieve smooth face-to-face communication for future work.

The contributions of this research are as follows:

- Proof-of-concept of a thin digital full face mask display with facial expression recognition to realize digital face augmentation in real environments
- Experimental analysis of the close-proximity-type facial expression recognition method based on photo reflective sensor arrays, and evaluation of the effect on the sense of self-agency and self-ownership toward the controlled avatar face

2 Related Research

2.1 Facial augmentation

Digital mask displays have been investigated to digitally augment the entirety or a part of the user's face in real environments. For instance, digital mask based performance art has been demonstrated with tablet PCs in several media art projects, such as Yamada Taro Project [38] and Toshiba's TabletMan [10]. Yamada Taro Project was a temporary production with an element of anonymity. The performer used an iPad to capture the faces of people on the street and projected those faces onto his own. Toshiba's TabletMan was a digitally augmented super-hero character with a large number of Toshiba tablet PCs attached to his body, for the global promotion of the tablet. Although digital face augmentation display has gained attention as new media technology for media art and advertising, these technologies have not considered application to face-to-face communication and thus lack the capability of facial expression recognition.

On the other hand, in telepresence technology, augmented face-to-face communication systems have been investigated to enhance the presence of a remote user. For example, Misawa et al. proposed ChameleonMask [22]. In order to incarnate a remote user in a human body, another person acting as a proxy for the remote user wears a mask-like tablet display on which the remote user's face is shown. Accordingly, it becomes possible for the proxy to stand in for the remote user in daily conversation, which also improves a sense of both affinity and realism, when compared to a conventional monitor-based remote conversation system. Sakurai et al. also proposed a telepresence system that can alter the appearance of an interlocutor in a remote location and increase creativity in remote group work [23]. In contrast, Suzuki et al. proposed a system that augments a remote interlocutor's non-verbal information to enhance a sense of presence in conversation [34]. While conventional facial augmentation research has focused on a high-presence remote communication system, this study aimed to explore new opportunities of facial communication in a real environment, by augmenting the real face of a speaker with an avatar. Researchers have investigated a variety of wearable display systems to control the facial appearance of a user. For instance, Osawa et al. proposed a glasses-like display device, Agencyglass [26]. Agencyglass resembles a pair of sunglasses made of LCDs which are the same size as the eyes. The pre-recorded movements of the wearer's eyes are displayed on the LCD glasses. To reduce the burden of complex emotional control felt by shop staff who must control their emotions to serve customers with a smile even when they feel depressed, AgencyGlass presents smiling eyes on the LCD glasses displays and enables shop workers to replay their pre-recorded natural eye movements based on situations during customer service. HYPERFACE [5] is a hat-like device

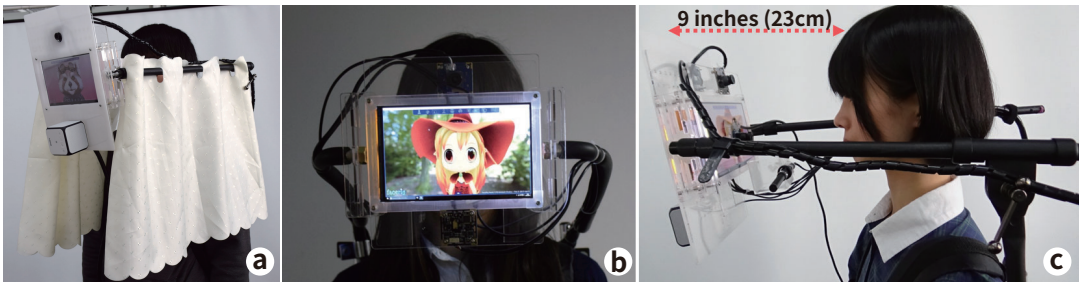

Figure 3: (a) Overview of a webcam-based face-mask display, e2-Mask [37]. (b) Front-view of e2-Mask. (c) Side-view of e2-Mask. The red dotted arrow on the side-view image shows the minimum camera-to-wearer distance needed to capture the entire face of the wearer.

with a projector built into the brim and applies projection mapping onto a person's face to augment the facial expressions of the wearer. For example, virtual make-up was demonstrated by projecting a red tint texture on the cheek of the smiling wearer. On the other hand, ChromoSkin [12, 13] augments conventional cosmetics with a thermochromic powder to computationally control the color of eye-shadow. While Agencyglass, HYPERFACE and ChromoSkin can change a part of the user's face, our proposed digital mask display aims to virtually substitute the whole face with an avatar.

Yamamoto et al. proposed a method of overlaying positive responses on audiences, and implemented a system which overlays an image of a smiling pumpkin on each audience member using see-through HMD [39]. Although this system is effective as a feedback visualization tool during presentation, our research focuses on more interactive scenarios in face-to-face communication, like everyday conversation, which requires greater variety of facial expressions than a simple smiling face.

2.2 Facial expression recognition interface

Researchers have explored wearable interfaces for facial expression recognition [2]. Fukumoto et al. [8] proposed a smile or laughter recognition method by using photo interrupters. Masai et al. [19] proposed smart eye-wear with embedded photo reflective sensors for facial expression recognition. The smart eye-wear can recognize eight types of facial expression. Suzuki et al. [34] proposed a facial expression recognition method using photo reflective sensors embedded in an HMD. As these works focus on identifying facial expression showing different emotions, they have difficulty recognizing the mouth movement of a speaker. Le et al. [16] proposed a facial-performance-sensing HMD utilizing eight built-in strain gauge sensors to recognize upper face expressions, and a depth camera to recognize lower face expressions. Mimicat [31] demonstrated a camera-based method to recognize face movement, in particular the opening / closing motion of the mouth and eyes. Sakashita et al. [29] demonstrated the usability of photo reflective arrays to detect the opening / closing of the mouth. In a similar way to these works, we utilized photo reflective sensors as part of our facial expression recognition method. Since photo reflective sensors are arranged on the whole face of a digital mask display, our proposed system cannot only recognize both facial expression and mouth movement correctly, but also reflect them on the animation of the avatar visualized on the same mask display unit.

2.3 Camera-based Digital Mask Display

We have previously explored a camera-based digital mask display, e2-Mask [37], which used two flat tablet displays and two web-cameras to capture and reflect the face of the wearer on an avatar (Figure 3). Both displays and cameras are mounted on each side of the mask device, i.e. the wearer's side and the interlocutor's side. With the wearer-side video-see-through display, the wearer can view the surrounding environment. Facerig [32] was used to control the facial expression of an avatar based on the expression and orientation of the wearer's face recognized by the wearer-side camera. The avatar was then visualized on the front side of the flat panel display (Figure 3(b)). The user can create a custom avatar to optimize their expressiveness and friendliness in real face-to-face communication. As the critical limitation of a camera-based digital mask display, a minimum camera-to-wearer distance (i.e. 23 cm) is required to recognize a full facial expression, which makes the display unnecessarily bulky (Figure 3(c)). Furthermore, a flat panel display and camera-based face capturing system restrain the movement of the user and limit face augmentation to the front view only. Also, a curtain is required to cover the wearer's real face (Figure 3(a)). To address these issues, in the following sections we explore a thin digital mask display design and implementation method based on a flexible display and a short range facial recognition system based on photo reflective sensor arrays.

3 Digital Full-face Mask Display

3.1 Work Flow

Figure 4 provides an overview of the work flow of the proposed digital full-face mask display, e2-MaskZ. Figure 1 shows an oblique view of e2-MaskZ and demonstrates the operation of reflecting the changing facial expression of the wearer in an avatar. e2-MaskZ is composed of a face-display mask and built-in face-capture mask. The face-capture mask first inputs the facial expression of the wearer, then the face-display mask outputs the encoded face information as the avatar on a curved display screen. To reproduce the convex profile of a human face, the face display mask utilizes an off-the-shelf flexible OLED display, Flexible Top Hat [28]. The flexibility of this display unit also allows us to smoothly overlay the face-display mask onto the face-capture mask.

The photo reflective sensors are laid out on the underside of the face capture mask. The system identifies the wearer's facial expression based on the photo reflective sensor data and produces an avatar presenting the same facial expression. The image of a generated avatar is displayed on the face display mask[1].

In this paper, to secure the system wearer's field of vision, eye-holes are made in the face capture mask and the face display is placed in a position below the eyes. The methods for identifying

[1]As images cannot be sent directly from a PC to the organic EL display used as the face display mask, a tablet or smartphone must also be used. Therefore, Duet Display [6] is used to make the PC recognize an iPhone as an external display and present the created avatar on the iPhone. The organic EL display mounted onto Flexible Top Hat can perform mirroring of the image on the iPhone, by using the iOS application RoStyle, developed by Royole Corporation. This is used to make the image on the iPhone be displayed on the face display mask.

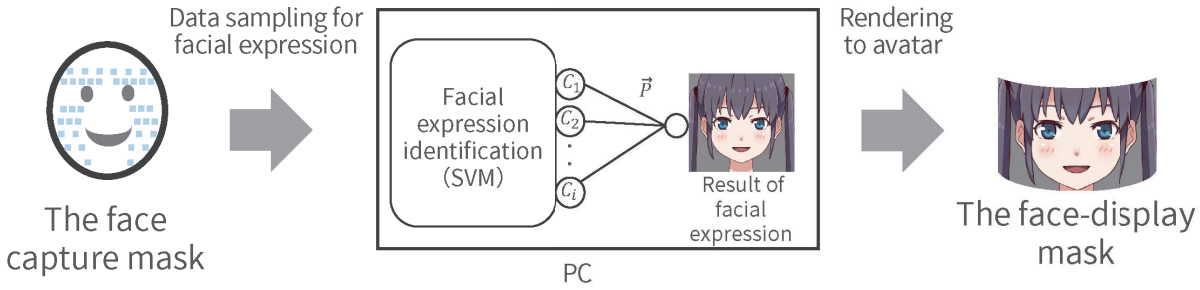

Figure 4: Workflow of a digital full-face mask display: Embedded face capture mask first samples the facial expression of the wearer. Then facial expression identification model is built based on a SVM training algorithm utilizing sample data measured at 40 feature points. Recognized facial expression is rendered to an avatar and displayed on a flexible full-face mask display.

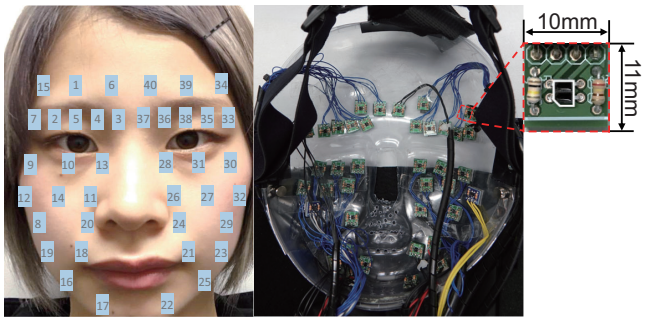

Figure 5: The positions of 40 facial features used to sample facial expressions with photo reflective sensor arrays (left). Face-capture mask with 40 embedded photo reflective sensors (center). The dimensions of a photo reflective sensor unit (top right).

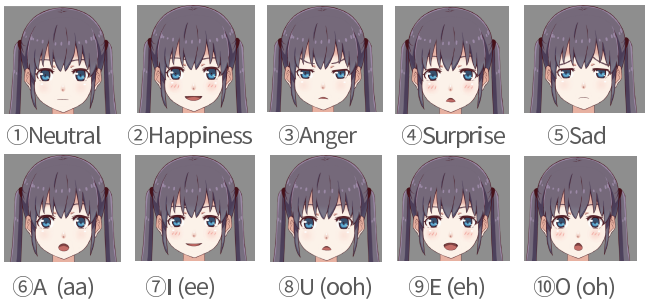

Figure 6: Facial expressions used in the experiment

facial expression and creating an avatar based on identified facial expression are explained hereafter.

3.2 Facial Expression Identification Method

Hardware: As photo reflective sensors, we used Genixtek Corporation's TPR-105 [3] to recognize the facial expression of a mask wearer [19].

As this miniature sensor has small dimensions of 2.7 x 3.2 x 1.4 mm (W x L x H), 40 units are densely arranged at sampling points on the face capture mask, as shown in Figure 5. Following the standard practice of face recognition [24, 25, 33], we picked eyebrows, cheeks and around eyes and mouth as the sampling points of photo reflective sensors. We ensure that the sensors do not come into contact with the eyes, nose and mouth of the mask wearer. Data from each sensor is collected using the Arduino Pro mini, then sent from the Arduino Pro mini to the PC through serial communication.

Software: Every time facial expression changes, the contours of the skin of the face change and simultaneously the distance between the photo reflective sensors and the skin of the face changes with each facial expression. Using this characteristic, we can take the data from each photo reflective sensor as feature values and use machine learning methods to construct a facial expression identification model. Considering a wearable and real-time digital face mask display, we chose Support Vector Machine (SVM) method as it has the advantage of low computational cost, and its feasibility for facial expression recognition using photo reflective sensors had been reported in prior work [19]. The input from each photo reflective sensor was first encoded as 10-bit data through the Arduino then normalized between 0 and 1 on the PC. After this, the SVM (C = 10, gamma = 1.0) algorithm of rbf kernel was applied and the facial expression identification model was constructed.

The 40 pieces of photo reflective sensor data that have been normalized are taken as feature values then the facial expression class is estimated by applying these to the facial expression identification model. To create the avatar's intermediate facial expression, described in section 4.3, the probability distribution of the feature values to belong to each class of facial expression was constructed in a database.

3.3 Output of Avatar

The user's facial expression is reflected onto the avatar based on the facial expression identification result explained in section 4.2. This method refers to the work of Suzuki et al. [34, 35]. When the wearer's facial expression can be expressed by n dimension facial expression parameters, the avatar's facial expression parameters $\vec{P}$, such as the degree to which the left and right eyes are open and the width of the mouth, can be composed by the following formula:

$$\vec{P} = \vec{P_B} + \sum_{k=1}^{n} c_i \times (\vec{P_i} - \vec{P_B})$$

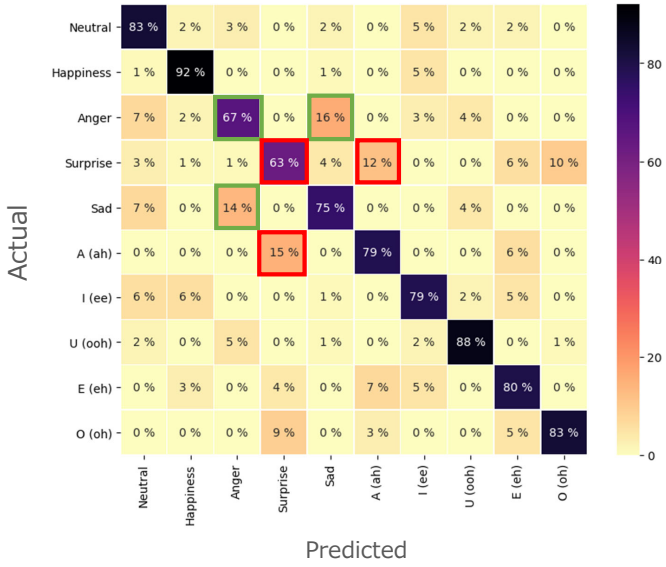

Figure 7: Results of confusion matrix

where $\overrightarrow{P_B}$ indicates straight-face facial expression parameters, $\overrightarrow{P_i}$ (i = 1, 2, ..., m) expresses m number of facial expression parameters besides the straight-face facial expression parameters and $\overrightarrow{C}$ = (c_1, c_2,..., c_m) represents the probability value of multiple-class classification of facial expressions.

Changing the avatar's expression by the above method makes it possible to express not only basic facial expressions, but also subtle ones, and also to output intermediate facial expressions that occur as facial expression is changing. We created an avatar 3D model by Live2D [17] and rendered it with Unity3D Game Engine. Communication between the programs on the PC uses UDP and comprises transmission of facial expression identification results and a keyboard command input.

4 Facial Expression Identification Accuracy

An experiment was conducted to evaluate the accuracy of the facial expression identification device. 10 types of facial expression were used in the experiment, as shown in Figure 6. We selected no. 1 to no. 5 based on the constant expressions defined by Ekman [27]. No. 6 to no. 10 are the facial expressions formed when pronouncing vowel sounds. Because the user has a conversation while wearing e2-MaskZ, it is essential to express not only facial expressions displaying emotion, but also the movement of the mouth when speaking. For this reason, 10 types of expression were used that include both facial expressions showing emotion and facial expressions when pronouncing vowels.

We recruited seven university students (four female, three male). In the process of the experiment, each subject first put on the prototype mask, then was instructed to imitate facial expressions no.1 to no.10, shown in Figure 6. Six sets of this experiment were conducted for each subject. For each facial expression, 10 data samples were collected. Since we repeated this experiment six times for 10 different facial expressions, we acquired 600 samples in total per subject.

4.1 Results

Six-fold cross validation (leave-one-set-out) was applied to the data of 600 samples (100 samples of test data, 500 samples of training data). As a result, the accuracy was 79%. It took about 3 minutes to collect 600 samples of data.

Table 1: Evaluation index for each facial expression

Facial expression	Precision	Recall	F value
Neutral	0.76	0.83	0.79
Happiness	0.87	0.92	0.90
Anger	0.75	0.67	0.71
Surprise	0.68	0.63	0.66
Sad	0.74	0.75	0.74
A (ah)	0.78	0.79	0.78
I (ee)	0.79	0.79	0.79
U (ooh)	0.88	0.88	0.88
E (eh)	0.77	0.80	0.79
O (oh)	0.87	0.83	0.85

4.2 Discussion

4.2.1 Accuracy of Facial Expression Identification Device

Figure 7 shows the accuracy of the confusion matrix of all the subjects combined. Under the condition of having learned the wearer's facial expressions, e2-MaskZ displayed 79% accuracy, and the time for calibration was only about 3 minutes per subject. The facial expression identification result was high overall despite the short calibration duration.

The 'oh' facial expression had the highest accuracy, while the 'surprise' facial expression had the lowest value. Let us analyze what caused surprise to have the lowest identification accuracy. Among the correct data for surprise, 63 % were estimated to be surprise, while 12% were estimated to be 'ah'. In contrast, among the estimation data for surprise, 15% were correct data for 'ah'. From these results, it can be said that the surprise facial expression is being mis-identified as the 'ah' facial expression.

The facial expression with the second lowest accuracy was anger. Among the correct data for anger, 67% were estimated to be anger, while 16% were estimated to be sadness. In contrast, among the estimation data for anger, 14% were correct data for sadness. In this experiment, the subjects commented that 'distinguishing between sadness and anger was difficult'. The high rate of mis-identification in anger and sadness could be due to the fact adult subjects don't express anger or sadness daily, leading to the difficulty of discerning between the two expressions.

From these results, we consider increasing the number of photo reflective sensor arrays on the face capture mask so that we can obtain more characteristics of the wearer's facial expressions and improve the accuracy of facial expression identification to reduce the rate of mis-identification.

4.2.2 Variation of Photo reflective Sensors

As shown in Figure 5, e2-MaskZ has 40 photo reflective sensors laid out all across the face of the mask. This means that, even if face size or shape differs, there are sensors that contribute to facial expression identification, which increases the versatility of the mask.

The average value of the photo reflective sensor data for each facial expression was calculated for 600 sample data, then, taking the straight face as the standard, the difference between each facial expression and the straight face was calculated. The value of accumulating differences with the same sign (+/-) was calculated. Figure 8 illustrates the validations of photo reflective sensors for three subjects, as examples. The horizontal axis of the graph represents the photo reflective sensor ID, while the vertical axis represents the normalized total of variation for each facial expression. The greater the normalized total variation of a photo reflective sensor, the more sensitively it can be said to be reacting to transformation of facial

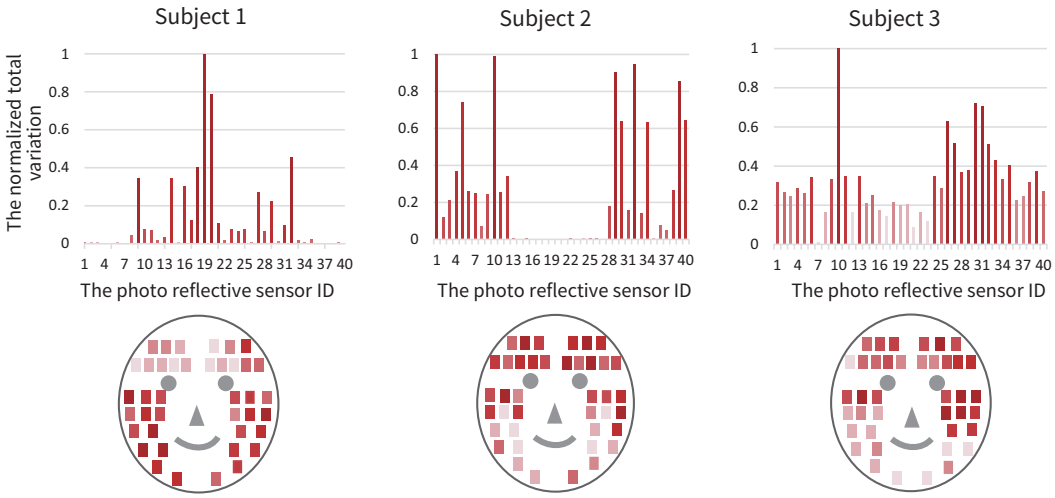

Figure 8: Variation of photo reflective sensor data of three subjects

expression. The three face diagrams of each subject visualize the degree of accumulated variation, with a heat map at each sensor position; a darker red color indicates more significant variation at that point.

Figure 8 shows that the facial expression variation in the photo reflective sensor data changes greatly for sensor numbers 13 to 26 (lower half of the face) for Subject 1, and numbers 1 to 13 and 28 to 40 (upper half of the face) for Subject 2. In the case of Subject 3, the variation in the photo reflective sensor data is approximately similar overall, except for sensor number 10. Despite the small detected variation, the facial expression identification of all three subjects shows sufficiently high accuracy, over 73 %, to demonstrate that successful identification is possible. Since e2-MaskZ assumes use by all kinds of people, it must be able to identify facial expressions with high accuracy, regardless of the size and shape of the wearer's face. As e2-MaskZ has photo reflective sensor arrays laid across the full face of the mask, we achieved the average accuracy of 79% among all our subjects, indicating the robustness of our full-face mask system.

5 The Sense of Self-agency and Self-ownership

To evaluate the feasibility of a digital face mask display based on photo-reflector sensor arrays (e2-MaskZ), we conducted user experiments to compare the sense of self-agency and self-ownership [9] toward the controlled avatar, between our mask and the camera-based digital face mask display (e2-Mask) [37].

5.1 User Experiment of Digital Face Mask Displays

We recruited four female and four male university students as the subjects in this experiment. The task of the experiment was to reproduce five emotional states of facial expressions (i.e., neutrality, happiness, anger, surprise, and sadness in Figure.) as shown in Figure 6 with the augmented avatar face on the two types of digital mask displays (i.e., e2-Mask and e2-MaskZ). We calibrated the proposed facial recognition model for each subject by collecting 1400 samples of each emotional expression from the subject before the experiment. As the benchmark method of a camera-based digital face mask display, we used e2-Mask with off-the-shelf facial expression recognition software, Facerig [32]. The subjects were instructed to practice facial expression control of the augmented avatar face for up to 20 minutes for both digital face mask displays (i.e. e2-Mask and e2-MaskZ). After the practice, we asked each subject to wear one of the digital face mask displays and guided them to a seat facing a wide mirror (2000 mm x 830 mm), positioned 30cm away from the mirror, so that they could look at their entire body with the augmented avatar face reflected in the mirror. To make a counter balance, we asked four of the eight subjects to start the experiment with the e2-MaskZ device and the other subject group to start with the e2-Mask device. To reduce the effect of fatigue on changing of facial expression, we also set a 10 minute break after the first half of the experiment, after which the two groups swapped over the digital face mask displays. All the subjects were asked to memorize the top five emotional expressions in the Figure 6 as the reference for each emotional expression, then reproduce the randomly instructed facial expressions without the reference image, during the experiment. We asked the subjects to keep facing forward, to avoid the effect of facial movement on the sense of self-agency and self-ownership. This process was repeated three times. After the end of the experiment, we asked subjects to answer three questions about the usability of the two digital face mask displays on a scale from 1 to 5, giving reasons for their answers, as shown in Table 2; Question 1 evaluates the sense of self-agency for recreating the intended facial expression with the augmented avatar face; Question 2 evaluates the sense of self-agency based on the response rate of changing facial expression; Question 3 evaluates the sense of self-ownership toward the augmented avatar face.

5.2 Results

Table 2 presents the mean value (M), the standard deviation (SD), the significant probability (P) of both the proposed method and benchmark method for each item in the questionnaire. Regarding the average value, a smaller value indicates a more positive response to the question. The two-sided significant probability (P) was calculated by performing the Wilcoxon signed-rank test on the results of each question for both the proposed method and benchmark method. Question 1 showed 2.8 for the mean value of the proposed method (e2-MaskZ), while the mean value of the benchmark method (e2-Mask) was 3.0, which is the neutral value. Question 2 showed 2.0 for the mean value of the proposed method, while the average value of the benchmark method was 2.6. Both values were lower than the neutral value. Question 3 showed 2.6 for the average value of the proposed method, while the average value of the benchmark method was 3.3. The average value of the proposed method was lower than the value of the benchmark method and the neutral value. Subjects who showed positive response for both Question 1 and Question 2

Table 2: Comparison of proposed method and benchmark method

Question number	Question content	Answer	Proposed method M	Proposed method SD	Benchmark method M	Benchmark method SD	P
1	How well did the avatar recreate your intended facial expression?	1: Very well - 5: Not at all well	2.8	1.1	3.0	1.0	0.50
2	How smoothly did the avatar's facial expression change?	1: Very smoothly - 5: Not at all smoothly	2.0	0.5	2.6	1.1	0.28
3	How strongly did you feel a sense of self-ownership toward your augmented avatar face?	1: Felt strongly - 5: Did not feel at all	2.6	1.4	3.3	0.8	0.35

also gave positive feedback for Question 3. Likewise, subjects who showed negative response for Question 1 and Question 2 also gave negative feedback for Question 3.

5.3 Discussion

Overall, the subjects showed positive responses regarding the sense of self-agency for reproducing the expected augmented facial expressions with both digital face mask displays, and there was little difference between the proposed thin digital face mask display and the camera-based benchmark device. In both methods, there were cases in which it was difficult to reproduce the expected facial expression due to failure in tracking facial features. For example, the benchmark method failed to track the closed mouth in the neutral face and instead showed a semi-open mouth on the avatar. The lack of eye tracking in the proposed method failed to reproduce the correct gaze direction on the augmented avatar face. While both display methods showed positive feedback for the response rate of facial expression control, the subjects found faster response rate with the proposed digital face mask display than the benchmark method. Although the result indicates a positive correlation between the sense of agency and the sense of self-ownership, the sense of agency in smooth facial control showed a more significant influence on the sense of self-ownership toward the avatar. One of the subjects commented "If there is some delay in facial expression change, I cannot feel self-ownership of the augmented avatar." Another subject also said "I was able to feel a sense of self-ownership because I found my facial expression and the avatar's expression were smoothly linked when I was controlling the mask." These results indicate that the proposed thin wearable facial recognition system based on photo reflective sensor arrays allow users to control the augmented avatar face with little stress and has the advantage of a faster response rate compared to the conventional camera-based method, which helps to create a better sense of self-ownership toward the digital avatar face.

6 Limitations and Future Work

The proposed method can currently only recognize ten discrete facial expressions and cannot identify direction of gaze, or blinking. To address this limitation, in our future work we plan to place more photo reflective sensor arrays around the eyes to estimate the state of the eyes with a machine learning method [18]. With more high-density micro photo reflective sensor arrays, we plan to construct a 3D physical facial model [16] and track micro-expressions [21] to enable smoother facial expression transformation on the augmented avatar face. To improve the accuracy of facial expression recognition, machine learning algorithms other than SVM algorithm can be considered. We plan to use Random Forest as a classifier [4] to consider the probability of facial state transition between sample expressions, as an extra cue to help narrow down the options of potential future facial expressions that might be transitioned to from the current state. The potential mismatch of the sizes of the digital face mask display and the face of the user is another practical problem to address. When the mask was much smaller than the user's face, some photo reflective sensors touched the face and failed to obtain the corresponding facial features correctly. As a simple solution, an extra nose pad or padding around the edge of the mask can be added to overcome a small range of mismatch between the facial feature points and sensor array positions. To overcome significant size differences, we must make new masks of different sizes to allow us to cover a user face-size range that includes particularly small or large faces.

7 Conclusion

This paper aimed to explore a thin digital face mask display which can enable digital self-representation by directly replacing a human face with an avatar, to extend the control of facial expression and freedom of face-to-face communication in real environments. Digital face mask displays can be the new media interface for an individual to flexibly create their ideal persona, as perceived by others in real life, expanding face-to-face communication opportunities beyond the constraints of physical appearance, situational bias, Lookism, and established personal relationships. This paper addressed the technical challenge of creating a thin digital face mask display with conventional camera-based facial expression recognition techniques. We successfully demonstrated the proof-of-concept prototype which integrated a flexible OLED display with a millimeter-proximity sensing facial expression recognition system that consisted of 40 photo reflective sensor arrays and a 10-class facial expression identification model based on SVM. Our experimental result showed the proposed close-proximity facial expression recognition system can classify the facial expressions of the user wearing the mask display with an average accuracy of 79%. The user study of the sense of self-agency and self-ownership toward the controlled avatar facial expression suggested the photo-reflective-sensor based facial expression recognition system has high feasibility, in terms of recreating the expected facial expression, and can have some advantage over the conventional camera-based method, regarding the response rate. The study results indicated that the subjects cared more about the sense of self-ownership than the precision of facial expression control.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19H04157.

References

[1] Y. Akaike, J. Komeda, Y. Kume, S. Kanamaru, and Y. Arakawa. Ar go-kon: A system for facilitating a smooth communication in the first meeting. In *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, pp. 120–126, 2014.

[2] G. Bernal, T. Yang, A. Jain, and P. Maes. Physiohmd: A conformable, modular toolkit for collecting physiological data from head-mounted displays. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, ISWC '18, p. 160–167. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3267242.3267268

[3] G. Corporation. Photo reflector (reflective) type tpr-105, http://akizukidenshi.com/catalog/g/gi-03812/, 2020. Date last accessed March 24, 2018.

[4] A. Dapogny, K. Bailly, and S. Dubuisson. Pairwise conditional random forests for facial expression recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3783–3791, 2015.

[5] designboom. eun kyung shin's social mask is an ai gadget that displays human emotion in real time, https://www.designboom.com/technology/eun-kyung-shin-hyperface-artificial-intelligence-social-mask-08-04-2017/, 2017. Date last accessed July 20, 2018.

[6] duet. Duet display, https://ja.duetdisplay.com, 2020. Date last accessed Octorber 1, 2019.

[7] C. Frueh, A. Sud, and V. Kwatra. Headset removal for virtual and mixed reality. In *ACM SIGGRAPH 2017 Talks*, SIGGRAPH '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3084363.3085083

[8] K. Fukumoto, T. Terada, and M. Tsukamoto. A smile/laughter recognition mechanism for smile-based life logging. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, p. 213–220. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2459236.2459273

[9] S. Gallagher. Gallagher, s. 2000. philosophical conceptions of the self: Implications for cognitive science. *Trends in cognitive sciences*, 4:14–21, 02 2000. doi: 10.1016/S1364-6613(99)01417-5

[10] GREATWORKS. Tablet man, http://www.greatworks.co.jp/works/tablet-man.html, 2013. Date last accessed April 24, 2017.

[11] S. Hagiwara and K. Kurihara. Development and evaluation of a "gaze phobic komyusho" support system using see-through hmd based on social welfare approach. *Computer Software*, 33(1):52–62, Jan. 2016. doi: 10.11309/jsst.33.1_52

[12] C. H.-L. Kao, M. Mohan, C. Schmandt, J. Paradiso, and K. Vega. Chromoskin: Towards interactive cosmetics using thermochromic pigments. pp. 3703–3706, 05 2016. doi: 10.1145/2851581.2890270

[13] C. H.-L. Kao, B. Nguyen, A. Roseway, and M. Dickey. Earthtones: Chemical sensing powders to detect and display environmental hazards through color variation. pp. 872–883, 05 2017. doi: 10.1145/3027063.3052754

[14] C. Kawanishi. The effect of accuracy motivation on face function in person perception. *The Japanese Journal of Psychology*, 68(6):465–470, Feb 1998. doi: 10.4992/jjpsy.68.465

[15] G. Koulieris, K. Akşit, M. Stengel, R. Mantiuk, K. Mania, and C. Richardt. Near-eye display and tracking technologies for virtual and augmented reality. *Computer Graphics Forum*, 38:493–519, 05 2019. doi: 10.1111/cgf.13654

[16] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P.-L. Hsieh, A. Nicholls, and C. Ma. Facial performance sensing head-mounted display. *ACM Trans. Graph.*, 34(4), July 2015. doi: 10.1145/2766939

[17] Live2D. Live2d cubism, https://www.live2d.com/en/, 2020. Date last accessed February 24, 2020.

[18] K. Masai, K. Kunze, and M. Sugimoto. Eye-based interaction using embedded optical sensors on an eyewear device for facial expression recognition. In *Proceedings of the Augmented Humans International Conference*, AHs '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3384657.3384787

[19] K. Masai, Y. Sugiura, M. Ogata, K. Kunze, M. Inami, and M. Sugimoto. 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*, IUI '16, p. 317–326. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2856767.2856770

[20] A. Mehrabian et al. *Silent messages*, vol. 8. Wadsworth Belmont, CA, 1971.

[21] W. Merghani, A. K. Davison, and M. H. Yap. A review on facial micro-expressions analysis: Datasets, features and metrics, 2018.

[22] K. Misawa and J. Rekimoto. Chameleonmask: Embodied physical and social telepresence using human surrogates. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 14(2):115–128, Apr 2017.

[23] N. Nakazato, S. Yoshida, S. Sakurai, T. Narumi, T. Tanikawa, and M. Hirose. Smart face: Enhancing creativity during video conferences using real-time facial deformation. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work Social Computing*, CSCW '14, p. 75–83. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2531602.2531637

[24] T. Nishime, S. Endo, K. Yamada, N. Toma, and Y. Akamine. Feature acquisition from facial expression image using convolutional neural networks. *Journal of Robotics, Networking and Artificial Life*, 3:9, 05 2016. doi: 10.2991/jrnal.2016.3.1.3

[25] H. Nomiya, S. Sakaue, and T. Hochin. Recognition and intensity estimation of facial expression using ensemble classifiers. *International Journal of Networked and Distributed Computing*, 4:203–211, 2016. doi: 10.2991/ijndc.2016.4.4.1

[26] H. Osawa. Emotional cyborg: Complementing emotional labor with human-agent interaction technology. In *Proceedings of the Second International Conference on Human-Agent Interaction*, HAI '14, pp. 51–57. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2658861.2658880

[27] F. W. V. Paul Ekman. *Facial action coding system*. Stanford University, Palo Alto, 1977.

[28] ROYOLE. Flexible top hat, http://www.royole.com/jp/flexible-top-hat, 2019. Date last accessed October 30, 2019.

[29] M. Sakashita, T. Minagawa, A. Koike, I. Suzuki, K. Kawahara, and Y. Ochiai. You as a puppet: Evaluation of telepresence user interface for puppetry. *In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*, pp. 217–228, 2017.

[30] P. Secord. Facial features and inference processes in interpersonal perception. *Person Perception and Interpersonal Behavior*, pp. 300–315, 1958.

[31] R. Shoji, T. Yoshiike, Y. Kikukawa, T. Nishikawa, T. Saori, S. Ayaka, T. Baba, and K. Kushiyama. Mimicat: Face input interface supporting animatronics costume performer's facial expression. In *ACM SIGGRAPH 2012 Posters*, SIGGRAPH '12. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2342896.2342983

[32] Steam. Facerig, http://store.steampowered.com/app/274920/facerig/, 2017. Date last accessed April 24, 2017.

[33] K. Suzuki, Y. Kionshita, S. Sakurai, T. Narumi, T. Tanikawa, and M. Hirose. Gender-impression modification enhances the effect of mediated social touch between persons of the same gender. *Journal of Augmented Human Research*, 1(2):379–389, Oct 2016.

[34] K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. Affectivehmd: Facial expression recognition and mappingto virtual avatar using embedded photo sensors. *Transactions of the Virtual Reality Society of Japan*, 22(3):379–389, Sep 2017. doi: 10.18974/tvrsj.22.3_379

[35] K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*, pp. 177–185. IEEE, 2017.

[36] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. Facevr: Real-time gaze-aware facial reenactment in virtual reality. *ACM Trans. Graph.*, 37(2), June 2018. doi: 10.1145/3182644

[37] A. Umezawa, Y. Takegawa, and K. Hirtata. e2-mask: Design and implementation of a mask-type display to support face-to-face communication. In *International Conference on Entertainment Computing*, pp. 88–93. Springer, 2017.

[38] vimeo. Yamada taro project, https://vimeo.com/82250584, 2017. Date last accessed April 24, 2017.

[39] K. Yamamoto, K. Kassai, I. Kuramoto, and Y. Tsujino. Presenter supporting system with visual-overlapped positive response on audiences. In *Advances in Affective and Pleasurable Design*, pp. 87–93. Springer, 2017.