Evaluation of Facial Expression Recognition by a Smart Eyewear for Facial Direction Changes, Repeatability, and Positional Drift

15

KATSUTOSHI MASAI, KAI KUNZE, and YUTA SUGIURA, Keio University
MASA OGATA, National Institute of Advanced Industrial Science Technology
MASAHIKO INAMI, The University of Tokyo
MAKI SUGIMOTO, Keio University

This article presents a novel smart eyewear that recognizes the wearer's facial expressions in daily scenarios. Our device uses embedded photo-reflective sensors and machine learning to recognize the wearer's facial expressions. Our approach focuses on skin deformations around the eyes that occur when the wearer changes his or her facial expressions. With small photo-reflective sensors, we measure the distances between the skin surface on the face and the 17 sensors embedded in the eyewear frame. A Support Vector Machine (SVM) algorithm is then applied to the information collected by the sensors. The sensors can cover various facial muscle movements. In addition, they are small and light enough to be integrated into daily-use glasses. Our evaluation of the device shows the robustness to the noises from the wearer's facial direction changes and the slight changes in the glasses' position, as well as the reliability of the device's recognition capacity.

The main contributions of our work are as follows: (1) We evaluated the recognition accuracy in daily scenes, showing 92.8% accuracy regardless of facial direction and removal/remount. Our device can recognize facial expressions with 78.1% accuracy for repeatability and 87.7% accuracy in case of its positional drift. (2) We designed and implemented the device by taking usability and social acceptability into account. The device looks like a conventional eyewear so that users can wear it anytime, anywhere. (3) Initial field trials in a daily life setting were undertaken to test the usability of the device.

Our work is one of the first attempts to recognize and evaluate a variety of facial expressions with an unobtrusive wearable device.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**;

Additional Key Words and Phrases: Wearable, facial expression, affective computing, eyewear computing

ACM Reference format:
Katsutoshi Masai, Kai Kunze, Yuta Sugiura, Masa Ogata, Masahiko Inami, and Maki Sugimoto. 2017. Evaluation of Facial Expression Recognition by a Smart Eyewear for Facial Direction Changes, Repeatability, and Positional Drift. *ACM Trans. Interact. Intell. Syst.* 7, 4, Article 15 (December 2017), 23 pages.
https://doi.org/10.1145/3012941

This work is supported by JST CREST grant number JPMJCR14E1.
Authors' addresses: K. Masai, Y. Sugiura, and M. Sugimoto, Interactive Media Lab, School of Science for Open and Enviromental Systems, Graduate School of Science and Technology, Keio University, 223-8522, Japan, emails: {masai, sugiura, sugimoto}@imlab.ics.keio.ac.jp; K. Kunze, Graduate School of Media Design, Keio University, 223-0061, Japan, email: kai@kmd.keio.ac.jp; M. Ogata, Media Interaction Group, Information Technology Research Institute (ITRI), National Institute of Advanced Industrial Science and Technology, 305-8560, Japan, email: masa@masaoagta.com; M. Inami, Research Center for Advanced Science and Technology, The University of Tokyo, 153-8904, Japan, email: inami@inami.info.

https://doi.org/10.1145/3012941

1 INTRODUCTION

Sometimes we find ourselves among others who speak a language that we do not understand. Even in that situation, however, it is still possible to guess some contextual information by observing nonverbal clues: for example, if they are fighting or in love, or if it is a business or private meeting. Even when we communicate in the same language, we rely heavily on nonverbal clues such as pauses, gestures, and tone of voice. Some estimate that 60% to 90% of our daily communication is nonverbal [14, 20]. While these nonverbal clues vary in kind (some auditory and some visual), facial expressions are perhaps among the most accessible and important.

As nonverbal clues play an essential role in our everyday interpersonal interactions, it seems natural to incorporate them in the field of Human-Computer Interaction. As computing systems become increasingly ubiquitous and support us in everyday situations, they need to be able to process more contextual information to improve the quality of human-computer interactions. Recognizing facial expressions would be an important step toward improving user experience.

Facial expressions are vital in communicating a person's intentions and emotional states. Affective computing explores the possibility of incorporating human affection in computing [2, 25]. A device that recognizes facial expressions may open up new opportunities for a more naturalistic user experience in human-computer interactions since facial expressions provide rich information about our emotional states [12]. A common approach to recognizing facial expressions is computer vision based, which analyzes recorded images to identify certain expressions. However, this method is only reliable in an experimental setting, and there is still relatively little research that tries to detect facial expressions in a daily life scenario.

Our goal is to create an unobtrusive and truly wearable device that recognizes people's facial expressions in everyday life. To this end, we have developed a device that can recognize facial expressions, specifically in the form of a smart eyewear (see Figure 1). We believe that "wearability" is important for tracking users' emotional states for the long term. In this project, we focus on skin deformations around the eyes caused by the movement of facial muscles in order to detect facial expressions in an efficient and minimally obtrusive way. We use 17 photo-reflective sensors that are integrated into the front frame of the glasses to detect the skin deformation around the eyes. The sensors we used for our prototype are small enough that they can be potentially integrated into glasses for everyday usage. Our approach improves the usability in a daily life setting, compared to the camera-based systems. Our device is also "wearable" in terms of social acceptability as the design follows that of conventional eyewear. The high wearability can, in turn, translate into higher tracking ability. It is the kind of wearable device that can provide a good vehicle for understanding the user's affective patterns in day-to-day scenarios, as discussed by Picard et al. [24]. Our method can detect various facial movements robustly by applying SVM to the data collected from the 17 photo-reflective sensors that are embedded in the front frame of the glasses.

In this article, we focus on recognizing eight universal facial expressions (see Figure 2). In addition to the universal six facial expressions (happiness, disgust, anger, surprise, fear, and sadness) defined by Ekman [4], we include contempt, which is sometimes considered as a universal facial expression [19] and "neutral" as a baseline for detection. We will discuss the possible usage scenarios and offer some insights into our affective states acquired from our system. More complex facial expressions such as a fake smile or a "happy surprise" are out of the scope of our current project, though they will be interesting signals to tackle in the future.

The aims of our work are as follows: (1) Recognizing the user's expressions in the eight universal categories (neutral, happiness, disgust, anger, surprise, fear, sadness, and contempt) with a wearable device in a reliable and robust manner. We used 17 photo-reflective sensors that cover most movements of the facial muscles. The acquired data are applied to SVM for robust recognition. We evaluated the robustness toward usage in a daily life setting: when the user changes the

Fig. 1. User wears our smart eyewear. It can recognize the wearer's facial expressions.

Fig. 2. Our smart eyewear recognizes eight universal facial expressions.

head direction, when the user uses the device on different days, when the user walks, and when the device slips down the nose. (2) Designing a truly wearable device that is socially acceptable and can be used anytime, anywhere. We designed and implemented our system in the form of a fully packaged, conventional-looking eyewear, so the device comfortably fits the context of daily usage. To this end, we used small photo-reflective sensors that are small and light enough to be integrated into everyday glasses. (3) Observing and analyzing the long-term distribution of the user's facial expressions in a daily life setting. This work is an extended version of the previous submission [17].

2 RELATED WORK

2.1 Facial Expression Recognition with a Camera

There are some notable works related to facial expression recognition in the field of computer vision. The two main streams are the recognition of affect and the recognition of facial action units [30]. According to Tian et al., the general approach to automatic facial expression analysis (AFEA) consists of three steps: face acquisition, facial data extraction and representation, and facial expression recognition [29]. One AFEA approach achieved an accuracy of 91.5% for recognizing basic expressions [16] using the Cohn-Kanade Database [11]. However, there are three major problems with camera-based systems regarding usability in a daily life setting. First is the limited field of view for tracking. Most of the previous works used a single fixed camera installed in the environment. As the position of the camera is fixed, it has difficulty in tracking the users continuously over a wide area for a long period. Tracking the users becomes especially difficult when they move constantly or there is an obstacle between them and the camera. Second, the AFEA camera systems, for the most part, work best with a clear frontal image of the face. However, in a real-life setting, a frontal view of the face is not always available. Since a facial image acquired from a nonfrontal view reduces the accuracy of facial expression recognition, this type of system is not optimal for real-life tracking of facial expressions. Finally, a camera-based system does not fit social contexts. Even though the aforementioned problems may be solved with a wearable camera, such a device will likely stir public concerns for privacy. Also, since a camera needs to be at a certain distance from the people to capture their facial expressions, this might interfere with a natural flow of personal interactions. Kimura et al. proposed an eyeglass-based, hands-free

videophone to overcome these problems [13]. It can yield a frontal face image with fish-eye cameras and capture facial expressions. However, the system is wired to laptop computers, which poses a major problem in wearability.

2.2 Facial Expression Recognition with Wearable Devices

One of the first attempts to detect facial expressions with a wearable device was Expression Glasses [27], which can recognize specific facial expressions (confusion/interest) by measuring facial muscle movement with piezoelectric sensors. Gruebler and Suzuki designed a wearable device that can read positive facial expressions using facial EMG signals [10]. Their device has to be attached to the side of a face, but it can record the user's affective state for more than 4 hours with high accuracy. It can be used during therapeutic interventions and to support medical professionals. Li et al. used a depth camera to capture expressions on the lower half of the face and eight strain gauges to capture expressions on the upper half of the face inside a head-mounted display [15]. They also mapped the input signals to a 3D face model. These prior works used contact-based sensors. While they performed well under experimental conditions, the measurement processes require continuous physical contact, meaning that the sensors/electrodes need to be attached to the user the whole time. This need for physical contact can make the user experience rather uncomfortable, especially over a longer period of time.

2.3 Sensing with Photo-Reflective Sensors

Fukumoto et al. used photo-reflective sensors attached to the glasses to capture skin deformations at the corners of eyes and cheeks that occur with happy facial expressions [6]. They then used threshold-based clustering to distinguish smiles from laughs. While efficient, this method does not scale well for multiple users because of the individual variations in determining the appropriate threshold (i.e., how much the skin around the eyes moves while smiling or laughing varies from person to person). Besides, due to the limited number of sensors, it can miscategorize other facial expressions as the target ones. There are some other interesting applications involving a limited number of photo-reflective sensors. For instance, Nakamura et al. proposed a device with one photo-reflective sensor, which intuitively and seamlessly controls augmented reality information using the natural movement of eyebrows when users try to focus and stare at something [21].

Photo-reflective sensors have also been used to capture skin deformations on various parts of the human body aside from the face. For example, Ogata et al. leveraged skin deformations on the forearms to use skin as an interface (SenSkin) [22]. This work used two arrays of six photo-reflective sensors to detect gestures (pinch, touch, pull-up, pull-down, pull-right, and pull-left) on the skin surface. Their previous work (iRing) also leveraged skin deformations on the finger to capture finger gestures and external input [23].

3 APPROACH TO FACIAL EXPRESSION RECOGNITION

In this research, we develop a device in the form of eyewear. Eyewear computing is a promising technology for facial expression and affect recognition in real life. Since the head is the primary location for most of the human senses, we can gain access to a variety of physiological signals by placing sensors in the head area. There remains a design problem as anything worn on the head is quite noticeable, yet with increasingly smaller Printed Circuit Boards (PCB), sensors, and actuators, we can now build smart glasses that are similar in appearance to normal eyewear, making them socially acceptable in terms of both appearance and comfort.

We use a large number of photo-reflective sensors and apply a machine-learning method for measuring facial expressions. This approach has the following advantages: (1) an adaptive and robust facial expression recognition that works with a variety of users, enabled by the richness

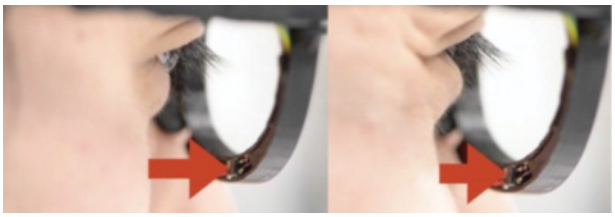

Fig. 3. Skin deformations change the distances between the sensors and the skin surface. The deformations occur when the facial expression changes.

of the sensor information and the machine-learning framework; (2) noncontact measurement: the sensors are unobtrusive and do not require physical contact, which improves the wearability of the device; (3) smart appearance: the sensors are small enough to be integrated into everyday glasses (e.g., NJL5908AR by New Japan Radio Co.: $1.06 \times 1.46 \times 0.5$ mm), making the device suitable for everyday usage; (4) simplicity: the processing required to interpret the sensor readings is minimal, so no elaborate feature extraction is necessary, and we can therefore keep the processing cost and energy consumption little, which is crucial for practical real-time recognition; and (5) affordability: the device only uses photo-reflective sensors with a microprocessing unit and can be manufactured at a low cost.

To capture facial expressions, we leverage skin deformations caused by the movement of facial muscles (see Figure 3). When users move their facial muscles, three-dimensional deformations occur on the skin surface. Each facial expression involves different movements of facial muscles. The movements of the eyelids, the eyebrows, the nose, and the cheeks all cause three-dimensional skin deformations around the eyes. The movement of the mouth also causes skin deformation under the eyes because the muscle movement around the mouth causes a cheek deformation that extends to the area below the eyes. According to [5], these movements are the greater parts of Action Units (AUs) with which the Facial Action Coding System codes human facial expressions. Therefore, placing sensors to capture the skin deformations around the eyes makes it possible to detect most muscle movements related to the target facial expressions [28]. Skin deformations are captured by measuring the distances between the skin surface and the photo-reflective sensors embedded in the various spots of the eyewear device.

3.1 Photo-Reflective Sensor

As discussed in Related Work, photo-reflective sensors are sometimes used in the field of Human-Computer Interaction to measure human skin deformations [6, 21–23]. We use infrared (IR) reflective sensors for this project. The sensors are composed of an IR LED and IR phototransistor. We used a 62K-ohm resistor for the phototransistor of the sensors and an 180-ohm resistor for the LED of the sensors.

To establish the basic characteristics of skin surface reflection captured by an IR photo-reflective sensor, we measured the voltage from the sensor (see Figure 4). It changes by the distance between the sensor and the skin surface. We collected 30 samples at each position. Figure 5 shows the average and standard deviation at each distance. The standard deviation is quite small (at most 0.014V). The correspondence is not linear. We can obtain the proximity to the skin. For closer distances, the photo-reflective sensors have a higher resolution.

4 IMPLEMENTATION

Figure 6 shows the data processing pipeline of our system. In this section, we describe hardware and software implementation details.

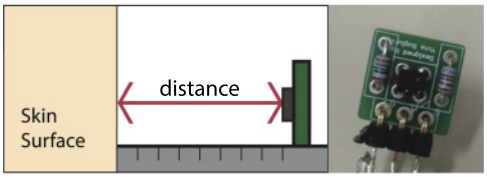

Fig. 4. (a) We measured the distance between a photo-reflective sensor and skin surface. (b) PCB used for this experiment.

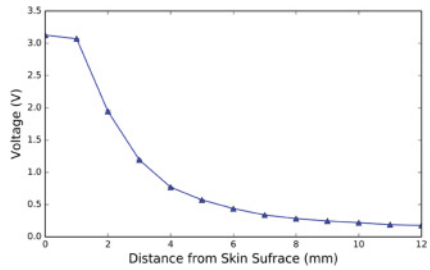

Fig. 5. The voltage change of the photo-reflective sensors related to skin surface distance.

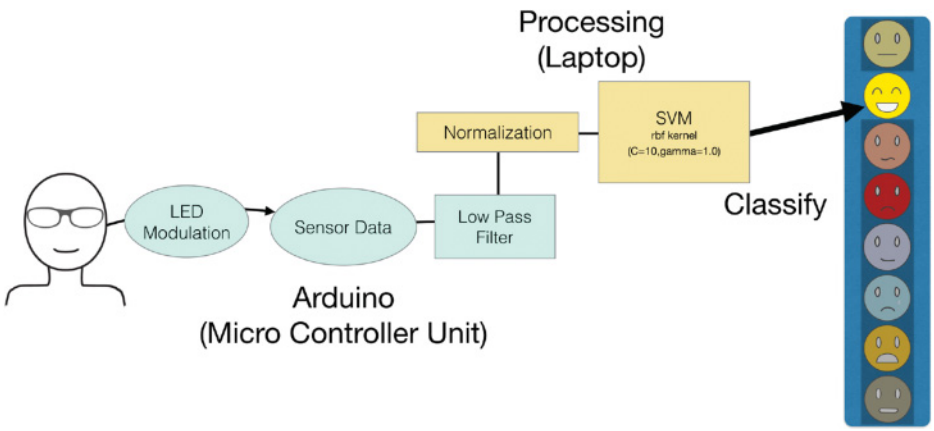

Fig. 6. Data processing pipeline.

4.1 Hardware

Figure 7 shows the components of our prototype. The prototype incorporates 17 photo-reflective sensors (SG-105 by Kodenshi, the placement can be seen in Figure 8), a 16-channel multiplexer (CD74HC4067 by Sparkfun), a transistor (IRLU3410PBF by International Rectifier), Arduino Fio, Xbee, and lithium polymer battery. The weight of the prototype is around 60g. The front frame is 3D printed, and the temple tips are taken from a regular commercial eyewear. An eyewear band is added to stabilize the position of the eyewear. The transistor is used to modulate the LED of the photo-reflective sensors because the sensors are easily influenced by ambient light such as the fluorescent lighting in the environment. We measure the difference between the values with LED on and off. The switching frequency is around 80Hz. With this method, it is possible to reduce the influence of ambient light. Xbee enables serial communication via ZigBee at 57,600 bits per second.

4.2 Software

An Arduino microcomputer convert the input voltage from each sensor into a 10-bit digital value. In order to take the difference of sensor values between LED is on and off, we multiply minus one to the value when LED is off. The value is smoothed out by applying a moving average to five pairs (LED is on/off) of each sensor value in a row to reduce noises. One data sample is a collection of 17 elements from the sensors. Each element of the data sample is the average of 10 sensor values. The

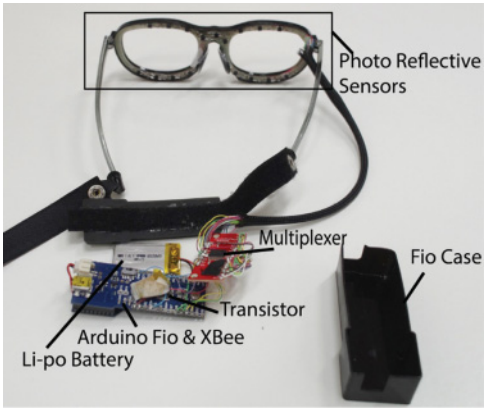

Fig. 7. System components.

Fig. 8. The placement of photo-reflective sensors.

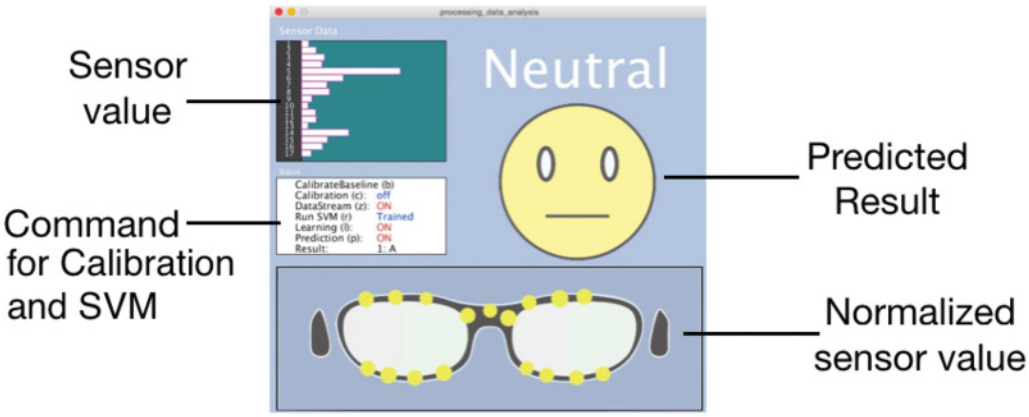

Fig. 9. User interface.

data sample is then sent to Java/Processing. In the Processing environment, we normalize the data sample and record the normalized data sample with a desired facial expression label as a training set. With the training set, we apply an SVM algorithm with a radial basis function (rbf) kernel (C = 10, gamma = 1.0) to classify facial expressions in real time. Figure 9 shows the visualization of our system. The emoticon shows the classification result. The bar graph represents the raw data. The yellow dots correspond to the positions of the sensors, and the size of the dot is correlated to the normalized value of each sensor. We use only one data sample for classification. For later experiments, we record normalized data samples described in this section.

4.2.1 *Machine Learning.*

(1) When the user makes a neutral expression with the device on, we set the baseline of the elements in the data sample (*BLV*) as 0.5.
(2) In this stage, the user dynamically moves his or her facial muscles. For each sensor, the range of each sensor value (*Range*) in the data sample is determined during this calibration process. Based on these values, the normalized sensor value (*NSV*) for each sensor is calculated as follows. An element of the data sample, which is the moving average on

each time frame, is defined as $SensorInput$. $Tolerance$ is chosen as 40 experimentally:

$$Range = (Max - Min) + Tolerance \tag{1}$$

$$\begin{cases} NSV = 0.5 + (SensorInput - BLV)/Range \\ if\ NSV > 1,\ then\ NSV = 1 \\ if\ NSV < 0,\ then\ NSV = 0. \end{cases} \tag{2}$$

$Range$ includes $Tolerance$ since there is a tradeoff with relying only on the acquired data.

- Advantage: Normalization can set the appropriate range for each sensor by measuring the range of $SensorInput$ that varies depending on the geometry of each user's face and the position of the sensor. The normalization can improve robustness because it can accommodate the weight of each $SensorInput$.
- Disadvantage: The position of the device during the calibration phase is not always stable. The user may move his or her facial muscles too dynamically and cause the eyewear to dislocate, resulting in inaccurate measurement of the Max and Min values. On the other hand, if the facial movement during calibration is not dynamic enough, it may reduce the amount of information that can be obtained. Therefore, it is not always possible to normalize the data sample in an optimal manner.

(3) During the learning phase, normalized data samples are stored with facial expression labels that are selected as the desired outputs. A label is attached to each data sample.
(4) For real-time classification, the normalized data sample is applied to SVM that is trained with the normalized data samples with the labels.

4.2.2 Algorithm. In addition to the training set that includes the output labels and the normalized data samples, the calculated values (CV) from two different sensors are also used for SVM. The calculation formula for $(S_i, S_j \mid 1 \leq i, j \leq 17)$ is shown as follows:

$$\begin{cases} CV = (S_i - S_j)/2 + 0.5 \\ \\ if\ CV > 1,\ then\ CV = 1 \\ if\ CV < 0,\ then\ CV = 0. \end{cases} \tag{3}$$

Sensor placement is shown in Figure 8. We considered the calculated values from adjacent sensors located in the bottom part of the front frame as well as the values from the sensors that have a vertical relationship. The calculated values from adjacent sensors in the bottom part are important because the surface is smooth and the skin is considered as an elastic model, and thus they correlate with each other as discussed by Ogata et al. [22]. The data from the sensors that have a vertical relationship can partially inform the position of the eyewear and the face. In total, we used 33 dimensions (normalized input: 17 + adjacent data: 7 [(10,11), (11,12), (12,13), (13,14), (14,15), (15,16), (16,17)] + vertical data: 9 [(1,10), (2,11), (3,12), (4,13), (5,14), (6,15), (7,16), (8,17), (9,17)]).

5 SYSTEM EVALUATION

For the system evaluation, we conducted four experiments. In the first experiment, we recorded the test dataset immediately after user-dependent training. The results suggested the user dependency of our system. We then evaluated the tradeoff between the number of sensors and accuracy, and the robustness to changes in head direction as well as to the removal and remount of the device. In the second experiment, we had a trial for multiple days to assess the possibility of long-term usage. In the third experiment, we tested the robustness of the system by taking measurements while walking. In the final experiment, we evaluated the influence of the vertical displacement

Fig. 10. Evaluation with different face directions.

Fig. 11. Experimental setting.

of our device on recognition accuracy. We also describe observations from a demonstration at SIGGRAPH Emerging Technologies 2015 [18].

5.1 Evaluation 1: Basic Setup

Eight users (four Japanese, one French, one Chinese, one Taiwanese, and one Sri Lankan; two of them female; average age: 27.3) participated in our evaluation. They were asked to sit in a chair and mimic the pictures of an American male (retrieved from the images of a man from a TV show "Lie to Me") making the universal facial expressions based on Ekman et al. [4]. First, the user looked straight ahead with a neutral facial expression, setting the baseline for the sensors. Then, the user moved the facial muscles for the calibration. We collected the data samples of eight facial expressions in different poses: looking straight ahead (three times), looking up (three times), looking down (three times), looking left (two times), looking right (two times), and taking off the device and putting it back on (two times) (see Figure 10). We collected data samples with different head directions because the movement of the head alone causes skin deformations as a result of the effects of gravity and joint coupling of muscles. The experimenter manually recorded 10 data samples with facial expression labels at regular 50-millisecond intervals in the middle of the time the user kept his or her maximum pose of each facial expression. In other words, in one recording of each facial expression, it takes 10 data samples for 1/2 second. Overall, each user's dataset includes 1,200 data samples (10 samples per expression per time * 8 facial expressions * 15 times in different poses). All recordings were conducted indoors (see Figure 11).

5.1.1 Accuracy with User-Dependent Training. First, we tested the accuracy in measuring each user's performance with user-dependent training. For each dataset, one recording (10 data samples) of each facial expression was divided into two sets, with the former half as a training set (600 data samples: 5 data samples per expression per time * 8 facial expressions * 15 times in different poses) and the latter half as a test set (600 data samples). We applied SVM with an rbf kernel (C = 10,

Table 1. Confusion Matrix (Within Subjects)

		Classified Results							
		N	H	D	A	Su	F	Sa	C
Actual Value	Neutral (N)	96.8%	0.7%	0.2%	0%	1.2%	1.0%	0.2%	0%
	Happiness (H)	0.3%	98.3%	0.5%	0%	0%	0%	0%	0.8%
	Disgust (D)	1.7%	0.2%	84.3%	3.2%	0%	5.5%	3.7%	1.5%
	Anger (A)	0.3%	0%	0.7%	96.5%	0.8%	0.3%	1.3%	0%
	Surprise (Su)	2.0%	0%	1.2%	0.2%	91.2%	4.7%	0.8%	0%
	Fear (F)	1.5%	0%	4.2%	0.8%	4.7%	87.5%	0%	1.3%
	Sadness (Sa)	4.5%	0%	2.5%	1.8%	0.8%	4.7%	85.7%	0%
	Contempt (C)	1.5%	0.3%	0%	0%	0%	0.3%	0%	97.8%

Table 2. User Dependency Matrix

		Test Data							
		A	B	C	D	E	F	G	H
Training Data	User A	99.2%	36.8%	31.8%	34.5%	28.8%	40.5%	48.0%	28.3%
	User B	39.2%	98.7%	23.3%	13.5%	30.0%	33.8%	32.8%	26.7%
	User C	42.5%	22.3%	84.8%	37.7%	21.0%	37.3%	40.7%	22.2%
	User D	19.8%	29.8%	33.5%	85.0%	32.0%	25.3%	30.5%	18.0%
	User E	23.3%	27.3%	18.7%	20.8%	89.3%	17.3%	39.8%	39.5%
	User F	44.3%	34.2%	28.2%	27.8%	27.7%	97.3%	35.0%	34.2%
	User G	43.8%	31.0%	24.7%	20.0%	27.7%	28.8%	88.7%	28.5%
	User H	12.5%	25.2%	15.7%	14.8%	26.2%	17.8%	29.8%	95.2%

gamma = 1.0) to the training set. In other words, we trained with user A's training set and tested with user A's test set. We achieved 92.8% accuracy on average (facial-expression-based result was 84.3%–97.8%; user-based result: 84.8%–99.2%). By learning from the data samples obtained with different head directions, our device was able to classify the facial expressions correctly, regardless of where the head was directed at the time of the measurement. Table 1 shows the confusion matrix of the results. As shown in the matrix, disgust can be similar to anger or fear. Besides, surprise and fear are close facial expressions with a 4.7% error to each other.

5.1.2 User Dependency. Second, we evaluated user dependency by training with each user's training set and testing with all other users' test sets. For example, we trained with user A's training set and tested with user B, C, . . . , F's test sets, respectively. The results are shown in Table 2. When the training set and the test set are taken from different users, the accuracy score drops significantly: accuracy is 48.0% at best (using the training set from user A and the test set from user G). This result indicates that, with the current system, users need to calibrate the device and train individually for accurate classification of their facial expressions.

Table 3 shows the confusion matrix of the result using the training set and test set that are merged from all users' datasets. Though there is user dependency, happy expressions can be recognized with relatively high accuracy (71.9%). The other facial expressions were harder to recognize (e.g., sadness: 17.2%, contempt: 22.7%).

Table 3. Confusion Matrix (Between Subjects)

		Classified Results							
		N	H	D	A	Su	F	Sa	C
Actual Value	Neutral (N)	74.9%	1.1%	6.3%	1.3%	5.2%	3.2%	2.6%	5.3%
	Happy (H)	5.3%	71.9%	2.6%	5.1%	0.4%	7.7%	2.8%	4.2%
	Disgust (D)	15.8%	7.5%	23.5%	12.2%	14.0%	17.5%	4.1%	5.2%
	Angry (A)	14.9%	7.8%	13.4%	26.6%	13.8%	15.5%	4.9%	3.1%
	Surprise (Su)	27.1%	3.2%	10.1%	4.1%	30.9%	11.7%	8.9%	4.2%
	Fear (F)	14.5%	6.5%	11.4%	8.8%	17.3%	27.7%	5.5%	8.2%
	Sad (Sa)	27.8%	8.6%	7.4%	6.4%	15.9%	13.6%	17.2%	3.1%
	Contempt (C)	29.5%	7.8%	8.6%	5.0%	11.7%	11.2%	3.7%	22.7%

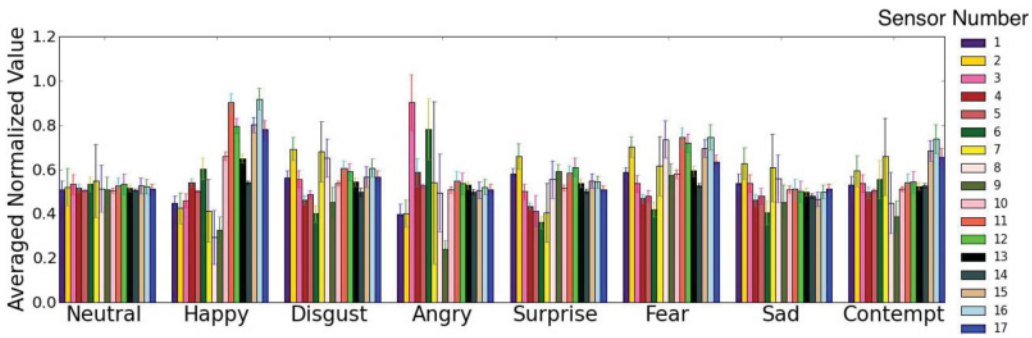

Fig. 12. Distribution of sensor value changes for each facial expression.

Figure 12 shows the distribution of data samples of each facial expression from user A. The sensor numbers (1 through 17) correspond with the numbers shown in Figure 8.

5.1.3 Number of Sensors. Next, we evaluated the tradeoff between accuracy and the number of sensors. We used the datasets from all eight users for this purpose. We merged eight users' training sets into the training set of 4,800 data samples and the test set of 4,800 data samples, respectively. We applied SVM in the same way as we did in Section 5.1.1. Using the training sets, we began by choosing the values from only one sensor that had the best accuracy based on the result of SVM. Next, we added the values from another sensor with the second-best accuracy. We repeated the process until all values of 17 sensors were included. For this analysis, we only applied the normalized 17-dimensional data to SVM.

As shown in Figure 13, the experiment yielded 84.1% accuracy with 17 sensor values. The accuracy improved with the addition of values from more sensors. With the values from 13 sensors, our system achieved more than 80% (81.0%) accuracy. Sensors 1, 9, 12, and 15 were left out. Sensors 1 and 9, as well as 12 and 15, are symmetrically located at the opposite ends of the frame. The farther the sensors are from the center of the face, the greater the distance between the eyewear frame and the skin surface on the face becomes. We think the greater distance results in the sensors having less information because they are more vulnerable to the noise of ambient light. Photo-reflective sensors work well when the distance between the sensor and the target is less than 10.00mm.

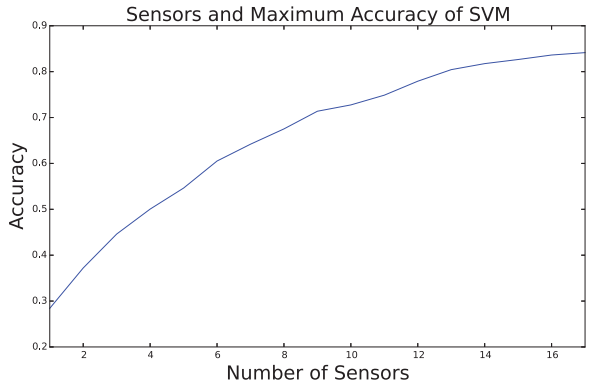

Fig. 13. Tradeoff between recognition accuracy and number of sensors.

Table 4. Accuracy by Different Conditions Compared to Looking Straight Ahead

	Condition					
	Upward	Down	Left	Right	Take Off & On	Average
User A	86.7%	72.9%	84.4%	93.8%	79.4%	83.4%
User B	67.5%	50.4%	48.1%	56.9%	45.6%	53.7%
User C	35.4%	66.7%	56.9%	46.2%	53.1%	51.7%
User D	24.6%	32.1%	65.6%	28.1%	64.4%	43.0%
User E	44.2%	47.9%	63.1%	58.1%	68.1%	56.3%
User F	50.0%	58.8%	59.4%	66.9%	40.0%	55.0%
User G	40.4%	45.8%	61.9%	43.8%	53.8%	49.1%
User H	81.7%	37.5%	46.9%	29.4%	55.0%	50.1%
Average	53.8%	51.5%	60.8%	52.9%	57.4%	

5.1.4 Accuracy When User Changes Head Direction and Removes/Remounts the Device. We have already shown that we can recognize facial expressions even when there are changes in head direction and when the user removes and remounts the device by obtaining data samples at those conditions in the training phase. We also evaluated how those conditions influence accuracy using the same dataset as before. We trained with the data samples obtained when the user looked straight ahead and tested with the data samples obtained in other conditions. The result is shown in Table 4. Accuracy varies among the users, but mostly it is between 50% and 60%, indicating the relative robustness of the recognition system.

5.2 Evaluation 2: Reliability over Time

In our second experiment, we collected data samples from three of the participants in the first experiment on different days (the data samples were obtained in the looking-straight position only). Users were asked to sit in a chair and put on the device. After the calibration, we collected data samples of the eight facial expressions three times each (240 data samples: 10 samples per expression per time * 8 expressions * 3 times) on each day. Like Evaluation 1, the experimenter recorded manually while the user kept the maximum pose of each facial expression at regular 50-millisecond intervals. We conducted the procedure on three different days, and so we acquired

Table 5. Accuracy on Different Days

	Day 1	Day 2	Day 3	Average
User A	83.8%	91.7%	90.4%	88.6%
User E	72.9%	70.8%	69.2%	71.0%
User F	86.3%	69.6%	68.8%	74.9%

720 data samples with facial expression labels from each user. We used the data obtained on two of the three days as a training set (480 data samples) and the data samples from the remaining day as a test set (240 data samples). We applied SVM in the same way as we did in Section 5.1.1. The results are shown in Table 5. The averaged accuracy for the three users was 78.1%. By making bigger the size of the training set, the repeatability can be ensured. It suggests the possibility of long-term usage. In the confusion matrix, the most dominant error was classifying 33.3% of the anger cases and 23.3% of fear cases as disgust.

5.3 Evaluation 3: Usage During Walking

For this experiment, we evaluated the effect of walking on the recognition of facial expressions because the activity may cause changes in the position of the eyewear, which influences sensor values slightly. We collected 480 data samples with facial expression labels from each user. Four users participated (two males and two females, two of them Japanese, one Chinese, and one German; age range: 25–59). We collected 10 data samples for the eight facial expressions in the looking-straight position. We repeated the process three times (dataset A: 240 data samples). This process follows that in Section 5.2. After the data samples had been collected in the stable position, users were asked to walk along a corridor at a natural speed. We manually collected 10 data samples at regular 50-millisecond intervals for the eight facial expressions three times each while they were walking and holding their maximum pose for each facial expression (dataset B: 240 data samples). Carrying a laptop, the experimenter walked along with each participant. The experimenter asked him or her to make all facial expressions one by one. Soon after recognizing his or her maximum pose of each facial expression, the experimenter recorded data samples with a laptop. We used dataset A as a training set and dataset B as a test set. We applied SVM in the same way as we did in Section 5.1.1. The result was an average accuracy of 73.2%, which is slightly worse than the result found in Evaluation 2. We assume this is because walking caused the device to shift its position.

5.4 Evaluation 4: Robustness to Positional Drift

To make our system robust, the noise by the positional drift of the eyewear should be considered as shown in the last evaluation. In our fourth experiment, we evaluated the robustness to the slipping of the device down the nose. We do not consider the slip to the side because it should not be a major issue if the eyewear is properly fitted to the user. On the other hand, the downward slip of the glasses is a common occurrence.

In this evaluation, sensor value distribution is different from other experiments as we collect data samples on various levels of the positional drift (levels). Hence, we normalized the dataset based on the average and standard deviation of training datasets.

First, in order to examine the relationship between the distance and the sensor values, we measured the distance d between the yellow mark and the base of the wearer's ear shown in Figure 14 in seven different positions. The distance corresponds to the degree of the positional drift. At the same time, we measured the sensor values of neutral expression in each position. We applied

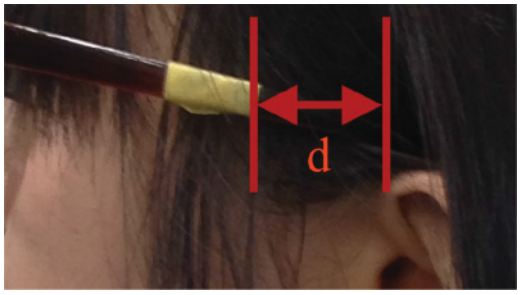

Fig. 14. The distance between the end of the yellow part and the base of the ear were measured.

principal component analysis (PCA) to reduce the 17-dimensional sensor data to the 1-dimensional data. The result can be seen in Figure 15 (top: raw data, bottom: PCA value). There is a linear relationship between the distance and the sensor values.

Next, we evaluate the effect of the slippage to the robustness of recognizing eight facial expressions. We collected 960 data samples with facial expression labels in total (10 data samples per expression per time * 8 facial expressions * 3 trials * 4 positional drift levels) from five participants (four male and one female). They are all Japanese graduate students aged 22 to 27. We followed the same procedure as Evaluation 1 except that we collect data samples at different levels instead of different poses. Level 1 is the base state where there is no slippage. The bigger the number of the level is, the greater the degree of positional drift. Figure 16 shows the snapshots of user B with different levels.

Figure 17 and Figure 18 show the averaged sensor values of all facial expressions at different levels of users A and B, respectively. The sensor numbers correspond to the ones shown in Figure 8. The levels are marked by colors. As the degree of the slip is user dependent, the levels are defined relatively. Sensor value distributions for users C, D, and E are shown in Figures 19 to 22. These figures focus on particular expressions at different levels (neutral, happy, angry, surprise) from different users. Two expressions (anger and surprise) from user E are included to show how the sensor value distributions differ by expressions.

We evaluated the possibility of predicting facial expressions at one positional drift level using the data samples taken at another level. The training set includes data samples from two trials at one certain level (160 data samples: 10 data samples per expression per trial * 8 expressions * 2 trials), while the test set includes data samples from another trial (80 data samples: 10 data samples * 8 expressions * 1 trial) at one level. We applied the cross-validation method: we evaluated three test sets from three trials at each level. We trained with the training set of level 1 and tested with the test set of levels 1 to 4, respectively. The process is repeated for the training sets from each level and each user. The SVM applied was different from the one previously used (linear kernel, C = 500) because the SVM with rbf kernel did not perform well. The result was then averaged for all users. The matrix can be seen in Table 6. At all levels, the accuracy of facial expression recognition was best when we used the training set and the test set on the same level (78.0%–87.8%). The farther the distance between the levels we used for the training set and the test set becomes, the worse the accuracy. We conclude it is hard to predict facial expressions when the slip of the glasses happens without the dataset that includes the data samples at the level of the slip.

We also evaluated the accuracy of facial expression recognition with the data samples of all levels. For each participant's dataset, the data samples taken in two of the three trials (640 data samples: 10 data samples per expression *8 facial expressions *2 trials *4 levels) were merged as

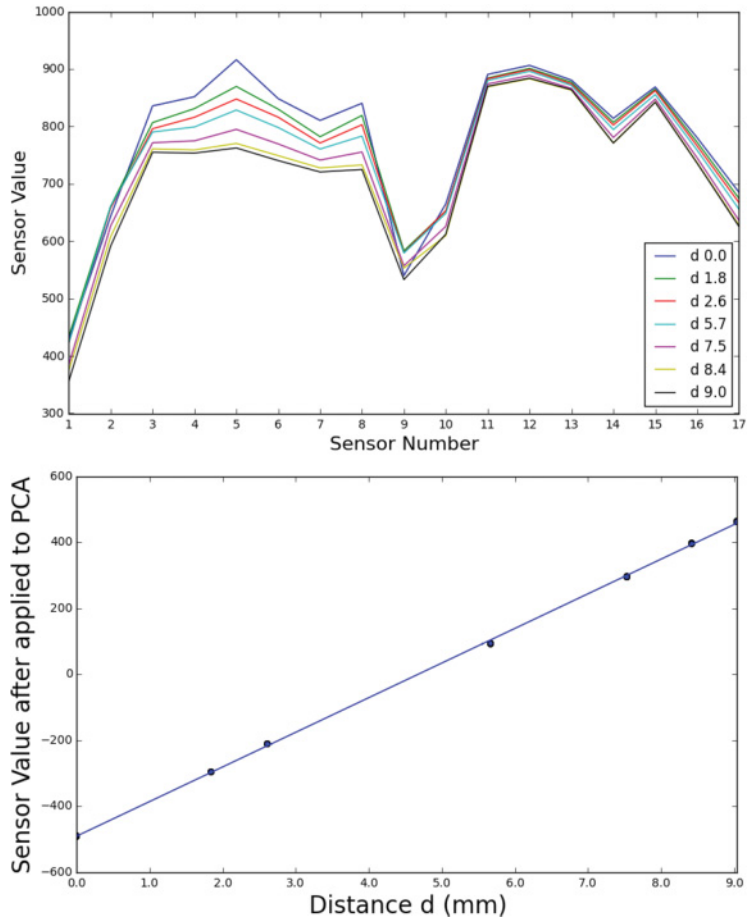

Fig. 15. The distance and sensor value (neutral) change depending on the level of positional drift. Top: raw data. Bottom: PCA.

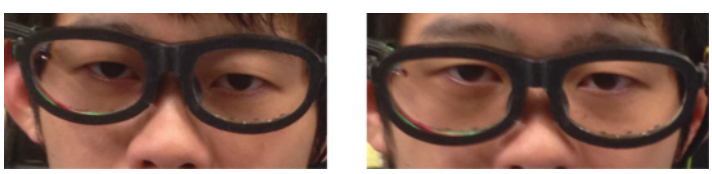

Fig. 16. The positional drift level of the eyewear (left: level 1, right: level 4).

a training set and the data samples from the remaining trial were used (320 data samples) as a test set. We applied the cross-validation method with SVM (linear kernel, C = 500) to the datasets. The results are shown in Table 7. The average recognition rate was 86.8%. We also applied PCA to the training sets. The average accuracy of facial expression recognition slightly improved to 87.7%. The best results are shown with (user A: 12, user B: 13, user C: 12, user D: 15, user E: 16) principal components, respectively. Even when the positional drift of the glasses happens, our system can recognize facial expressions with robustness by learning the data samples taken at different positional drift levels.

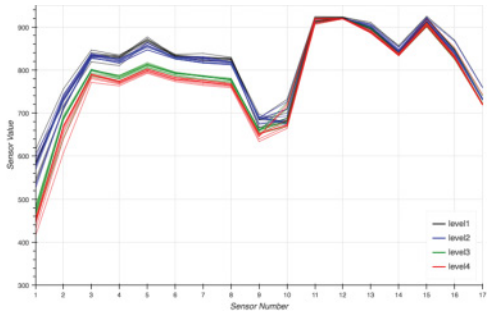

Fig. 17. Facial expression distribution on each level (user A).

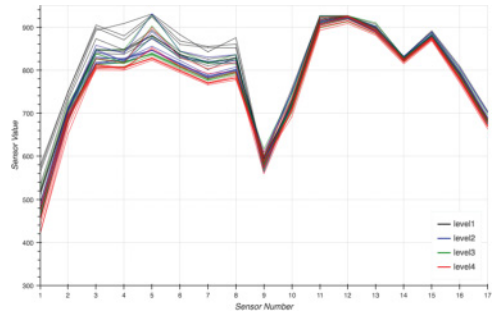

Fig. 18. Facial expression distribution on each Level (user B).

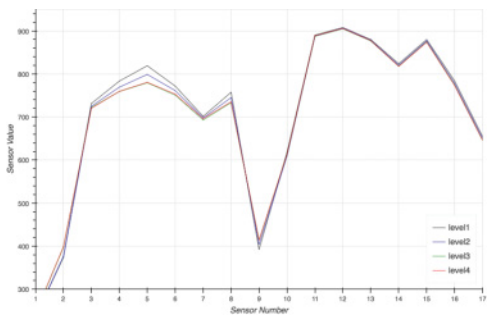

Fig. 19. Sensor value distribution on each level (user C, neutral).

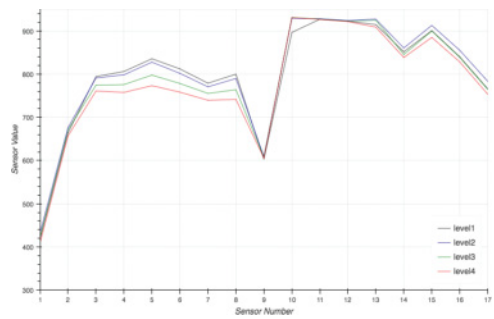

Fig. 20. Sensor value distribution on each level (user D, happy).

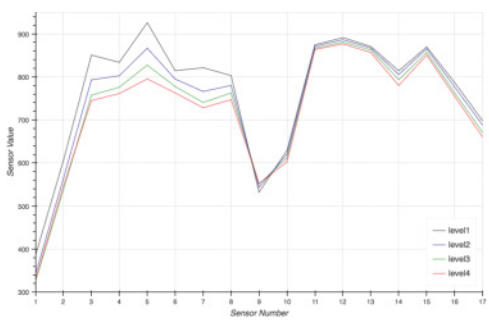

Fig. 21. Sensor value distribution on each level (user E, angry).

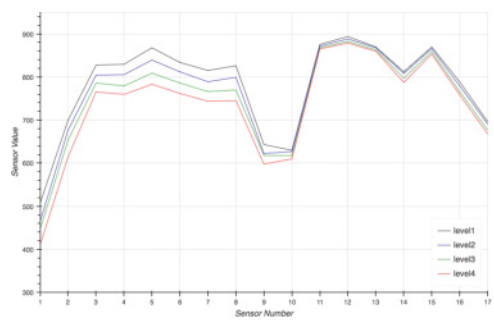

Fig. 22. Sensor value distribution on each level (user E, surprise).

Table 6. Accuracy of Facial Expression Recognition Using Training Set and Test Set on Different Levels

	Level 1	Level 2	Level 3	Level 4
Level 1	83.7%	39.6%	23.7%	21.8%
Level 2	46.5%	79.6%	45.9%	22.0%
Level 3	25.2%	42.9%	87.8%	38.7%
Level 4	23.0%	26.0%	44.6%	78.0%

Table 7. Accuracy of Facial Expression Recognition Using the Datasets That Include All Levels

	User A	User B	User C	User D	User E	Average
Result with 17 sensors	89.3%	85.5%	84.2%	82.1%	92.8%	86.8%
Best result (PCA)	91.0%	86.7%	82.3%	83.5%	94.9%	87.7%

5.5 Demonstration at SIGGRAPH Emerging Technologies 2015

We demonstrated our eyewear device at SIGGRAPH Emerging Technologies 2015. During the demonstration, we had more than 200 users from various international backgrounds try on the device. As the demonstration proceeded, we came to realize that the size and shape of the eyewear had to be adjusted to each user for accurate recognition. Three major issues were present: (1) The device sometimes slipped out of place when some users changed facial expressions. The slippage changed the accuracy of the recognition because the device relies on the proximity sensing between the front frame of the eyewear and the skin surface on the face. (2) For those who had high nose bridges, some of the sensors positioned between users' eyebrows saturated and did not work well. This problem could occur even with the neutral expression. (3) The sensors placed on the top can measure the distance to the eyelid or the eyebrows depending on the shape of the users' face or the position of the eyewear. Therefore, the sensor data can be different for each user, requiring individual training and calibration. For these reasons, the eyewear should be customized to each user. We also observed that the eyewear seemed to work better for Caucasians compared to Asians as Caucasians tend to be more expressive.

6 INITIAL FIELD TRIALS

We conducted initial field trials for daily life by recording facial expressions of a user in a daily living or home scenario. The recording was a 90-minute time series consisting of four activities: (1) playing Go game with a computer for 32 minutes (a board game involving two players originated in China), (2) playing with a dog for 16 minutes, (3) watching an episode of *Friends* for 22 minutes, and (4) programming on a computer for 20 minutes. Additionally, playing a shooting game with friends (won and lost: 10 minutes) and watching an episode of the crime drama *Crime Scene Investigation (CSI)* for 45 minutes were recorded on the following day to compare (1) the activities including social interactions (human-human and human-animal) as well as individual activities and (2) the same activities in a different context (e.g., winning vs. losing games, watching comedy vs. drama series). Figure 23 shows the frequency distribution of recognized facial expressions during the recording period. The figure presents a normalized distribution for every 2 minutes with a logarithm contrast enhancement. The distribution of facial expressions varied depending on the activities. For instance, happy expressions were mostly observed while interacting with the dog. The sitcom also produced happy expressions between intervals of neutral ones. While playing the Go game, negative expressions sometimes occurred, but they were eventually replaced with a happy expression reflecting the progression of the game (from facing challenges to winning the game). During individual activities, the user tended to show fewer facial expressions. For example, while programming induced some angry expressions, the dominant expression was neutral. On the other hand, the user showed more various facial expressions (other than neutral) while interacting with the dog. The user also showed various facial expressions while playing the shooting game with a friend. The results are shown in Figure 24. During the game, the user displayed more various facial expressions such as happy and angry while playing with another person, compared to when the user played alone with the computer (the Go game). It suggests

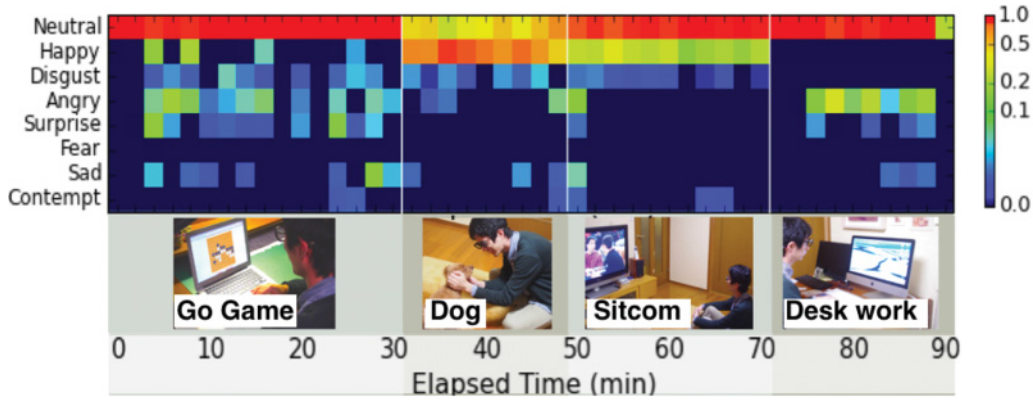

Fig. 23. Distribution of predicted result of facial expressions based on recorded sensor values for a long time.

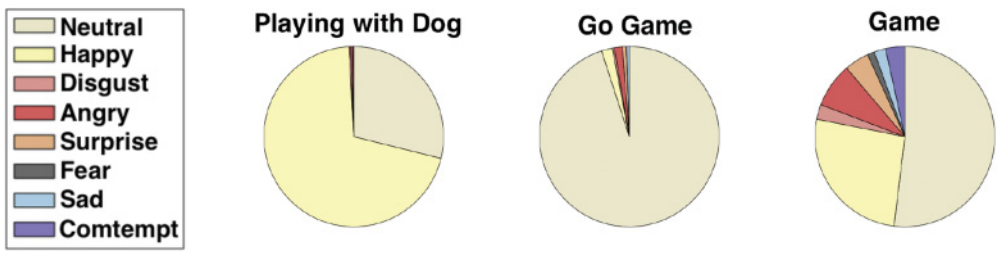

Fig. 24. Facial expression ratio during activities.

that social interactions induce more different facial expressions. Although this is only a preliminary field trial, it is tempting to speculate that facial expressions can be used as an indication of communication between people. This is consistent with the work by [1].

The distribution of facial expressions may also be influenced by the nature of the interactions as seen in the case of gaming in the trial (see Figure 25). When the user won, he showed more happy expressions, especially in the latter half of playing. When the user lost, negative expressions were displayed, the dominant expression being anger. However, the user frequently displayed happy expressions during the gameplay even when he eventually lost.

While watching the episode of the sitcom *Friends*, the user sometimes showed happy expressions. When the user watched an episode of the drama *CSI* that included some graphic scenes such as murder, dissection of a human body, and bleeding, disgust and surprise were detected. While this is an unsurprising result, it is once again tempting to speculate whether analyzing the distribution of facial expressions may be able to provide some feedback on the user experience.

7 DISCUSSION

In this article, we presented a smart eyewear prototype that can recognize the wearer's facial expressions with 92.8% accuracy using user-dependent training in the experimental setting. This result shows the potential of our proposed approach. Though many of the users remarked that they did not see clear differences between surprise and fear, fear and disgust, and disgust and anger when they looked at the pictures presented for instruction, it was possible to classify those expressions. The subtlety of differences between some expressions such as anger and disgust may lead to miscategorization. However, we believe this owes more to the ambiguous nature of

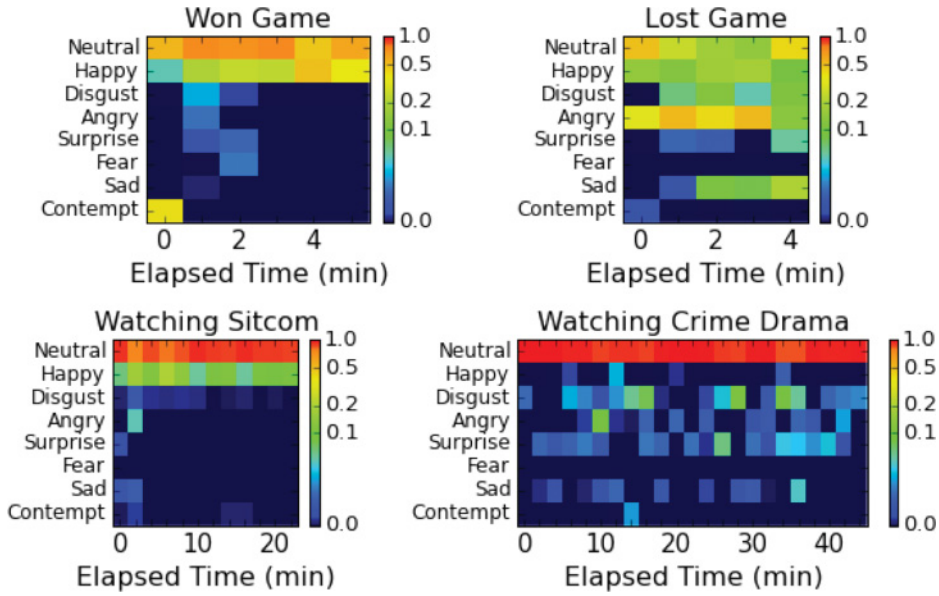

Fig. 25. Comparison between activities.

human facial expressions than to the design of our recognition system. We should also note that there is some empirical evidence that challenges the universality of basic emotions (See [26] for a discussion on cross-cultural recognition of facial expressions.)

Our analysis using the user dependency matrix showed the need for user-dependent training. This does not seem to pose a major problem with our device since eyewear is a personalized item as pointed out by Scheirer et al. [27]. Our device can be designed as personalized eyewear, intended to be used by a single user. However, having external datasets may help reduce the cost of the learning phase and improve the accuracy of recognition. Ideally, such external datasets should consist of data samples from a large number of users rather than a large volume of samples from a few users as suggested by Girard et al. [7].

The evaluation of our system found that accuracy was reduced in the case of long-term usage and walking. This reduction is due to the physical condition of the device such as the effects of ambient light and positional drift. Another factor for the reduction in accuracy was the difficulty of reproducing the same facial expressions across different times and conditions. It was quite challenging for the users to try to repeat the exact same facial expression. They made slightly different facial expressions with varying intensities. This is an inherent problem in developing a facial recognition system. However, our findings suggest that this may be made less of a problem by increasing the size of training sets with more trials. In our experiment that measured the effects of positional drift, we were able to achieve an accuracy of 87.7% by learning the data samples at different levels of positional drift. Requiring the users to conduct extensive and repetitive training would probably be unrealistic for real-life usage, but creating a new system that utilizes external training datasets may mitigate this issue.

We classified facial expressions by basic emotion categories. However, our facial expressions may not represent our inner emotions. Our face may reflect mental effort or convey a communication signal. Information other than basic emotions was not considered in this article, but some of the results may have been better understood if we had ways to recognize such nonemotional

signals. For example, in the field trials, angry facial expressions were recognized during computer programming or the Go game. It is not difficult to imagine that this was more a reflection of mental effort, confusion, or frustration than an expression of anger. Du et al. suggested the existence of 21 expressions [3]. The information provided by complex facial expressions would be useful for understanding the user experience in depth. We regard our work as a first step toward recognizing more complex facial expressions in daily life.

8 LIMITATIONS

In this study, we assumed that skin deformations around the eyes indicate changes in facial expression. However, other behaviors such as yawning, rubbing one's eyes, and resting one's cheek in one's hand can also cause skin deformation. These normal behaviors can affect our system of facial expression recognition. Moreover, readings from the photo-reflective sensors are affected by the condition of the facial skin. Factors such as tanning, sweat, makeup, and facial swelling may require users to do calibration and training again.

The eyewear device only collects data from the sensors around the eyes. Although movements of the mouth are partly detectable, there are several mouth movements our system cannot detect because mouth movements and cheek deformations do not have a one-to-one correspondence. For instance, the action of opening the mouth lowers the cheeks slightly, which makes the distance between the eyewear frame and the skin surface greater. Therefore, when the sensor values on the lower part changed, our system could not determine whether movement of mouth or cheek deformation caused the changes.

During the experiments, the participants made intentional facial expressions, although the goal of our work is to capture natural facial expressions. The posed and natural expressions are similar to a certain extent, but there is no doubt some differences exist between them. Such differences can have an adverse influence on the accuracy of natural facial expression recognition, especially because our natural facial expressions are more subtle than posed ones. In the field trial, our device recognized the expressions that were high in intensity, which suggests that our system was unable to pick up less intense expressions. That means our system can only provide an approximate picture of the pattern of facial expressions in a daily-life setting at this time.

9 USAGE SCENARIOS

In this section, we would like to present a few scenarios of an eyewear device that recognizes a wearer's facial expression for future research. The first scenario is "Collaborative Media Tagging" that uses facial expressions or emotions to evaluate contents. The idea is to record and compile facial expressions of the users on a large scale as they read books or watch movies/videos and index the contents so they can be searchable. The second is "Care System for Older Adults." A wearable device that can provide an overview of users' facial expression changes in their daily lives could be used to increase awareness of the user's emotional state. We can imagine a system that would notify children, nudging them to give their parents a call and talk to them when the system detects they are feeling sad and smiling less. The third is "Supporting System for People with Autism Spectrum Disorders." People with autism have difficulty in creating facial expressions of emotion. Our system can help them create facial expressions by giving them motivation to intentionally express their emotion with feedback if they could successfully make the expressions or not. The fourth is "Happiness Map." If we get to the point where we can reliably recognize facial expressions in our daily lives, we could combine facial expressions with location and demography. It may be possible to then search for places where the users frequently smile and laugh when comparing places to visit, live, or work. Finally, the notification to make users aware of their facial expressions in daily settings may have the potential to apply for emotion regulation [8, 9].

10 CONCLUSION AND FUTURE WORK

We presented a novel smart eyewear that recognizes a wearer's facial expressions. With the eyewear device, we conducted an evaluation to recognize eight universal facial expressions. The experimental results showed recognition rates of 92.8% for one-time use regardless of facial direction or removal/remount of the device; 78.1% for repeatability and multiple-day usage after a training process; and 87.7% if we take the positional drift of the glasses into account. The robustness in daily scenes can be achieved by learning more data. Moreover, we designed our prototype to be socially acceptable, following the looks of existing eyewear made possible by small photo-reflective sensors. Our system still has room for improvement regarding calibration and accuracy, yet we believe it is a major step in quantifying the flow of facial expressions in daily life.

We regard the following four issues as our future work.

First, we would like to design a natural learning process to capture natural facial expressions. To this end, we plan to apply an unsupervised learning method to the sensor data collected in daily-life settings. With this approach, we would not need any labeling for sensor data, and categorizing natural facial expressions would be possible.

Second, we plan to design an optical filter to reduce the influence of ambient light. The prototype used photo-reflective sensors that function on the basis of IR reflection. Because sunlight contains enormous amounts of IR light, sensor data saturates when the sensors are exposed to direct sunlight. With the current system, the sensors are not covered by anything and are easily influenced by an intense ambient light even though we applied light modulation to the LED of the photo-reflective sensors.

Third, we are looking to improve the calibration process. For robust recognition, the prototype requires each user to calibrate under various situations. We want to reduce this process by generating an additional dataset based on already trained data. We will also consider transferring the learning method so that we can use other users' datasets to calibrate another's effectively.

Finally, we would, of course, like to try to detect more subtle facial expressions. We used a classification algorithm for the prototype, and the intensity of facial expressions was not measured. This meant the current system could not measure complicated facial expressions such as ones that contain both happiness and surprise. Since our real-life facial expressions are not limited to the eight categories we considered in the present study, we need to work on recognizing more complicated or subtle facial expressions. Also, we think physiological information such as skin temperature, skin conductance, and eye gaze may be able to provide additional information that can contextualize some of the more ambiguous facial expressions. These sensory inputs could be integrated into the current form of the system. We also believe that photo-reflective sensors can detect skin deformations caused by eye movements. By applying a time-series data processing algorithm, we should be able to detect the four directions of eye movements and blinks. Ultimately, our distant goal is to detect facial expressions related to cognitive loads such as attention, interest, fatigue, and concentration by combining these different types of information.

REFERENCES

[1] U. Hess, R. Banse, and A. Kappas. 1995. The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology* 69, 1 (1995), 280–288.

[2] R. A. Calvo and S. D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 1, 1 (Jan. 2010), 18–37. DOI: http://dx.doi.org/10.1109/T-AFFC.2010.1

[3] S. Du, Y. Tao, and A. M. Martinez. 2014. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences* 111, 15 (2014), E1454–E1462. DOI: http://dx.doi.org/10.1073/pnas.1322355111. arXiv:http://www.pnas.org/content/111/15/E1454.full.pdf.

[4] P. Ekman. 1989. The argument and evidence about universals in facial expressions. In *Handbook of Social Psychophysiology*. 143–164.

[5] P. Ekman and W. V. Friesen. 1977. Facial action coding system. Consulting Psychologists Press, Stanford University, Palo Alto.

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

[7] J. M. Girard, J. F. Cohn, L. A. Jeni, S. Lucey, and F. D. Torre. 2015. How much training data for facial action unit detection? In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG'15)*, Vol. 1. 1–8. DOI: http://dx.doi.org/10.1109/FG.2015.7163106

[8] J. J. Gross. 1988. The emerging field of emotion regulation: An integrative review. *Review of General Psychology* 2, 3 (1988), 271–299.

[9] J. J. Gross. 2013. *Handbook of Emotion Regulation*. Guilford Publications.

[10] A. Gruebler and K. Suzuki. 2014. Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Transactions on Affective Computing* 5, 3 (July 2014), 227–237. DOI: http://dx.doi.org/10.1109/TAFFC.2014.2313557

[11] T. Kanade, J. F. Cohn, and YingLi T. 2000. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 46–53. DOI: http://dx.doi.org/10.1109/AFGR.2000.840611

[12] D. Keltner, P. Ekman, G. C. Gonzaga, and J. Beer. 2003. Expression of emotion. In *Handbook of Affective Sciences*. 411–414.

[13] S. Kimura, M. Fukuomoto, and T. Horikoshi. 2013. Eyeglass-based hands-free videophone. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC'13)*. ACM, New York, 117–124. DOI: http://dx.doi.org/10.1145/2493988.2494330

[14] M. L. Knapp, J. A. Hall, and T. G. Horgan. 2013. *Nonverbal Communication in Human Interaction*. Cengage Learning.

[15] H. Li, L. Trutoiu, Olszewski, L. K. Wei, T. Trutna, P. Hsieh, A. Nicholls, and C. Ma. 2015. Facial performance sensing head-mounted display. *ACM Trans. Graph.* 34, 4, Article 47 (July 2015), 9 pages. https://doi.org/10.1145/2766939

[16] G. Littlewort, I. Fasel, M. S. Bartlett, and J. R. Movellan. 2002. Fully Automatic Coding of Basic Expressions from Video. University of California, San Diego, CA 92093.

[17] K. Masai, Y. Sugiura, M. Ogata, K. Kunze, M. Inami, and M. 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 (IUI'16)*. ACM, New York, 317–326. DOI: http://dx.doi.org/10.1145/2856767.2856770

[18] K. Masai, Y. Sugiura, M. Ogata, K. Suzuki, F. Nakamura, S. Shimamura, K. Kunze, M. Inami, and M. Sugimoto. 2015. AffectiveWear: Toward recognizing facial expression. In *ACM SIGGRAPH 2015 Emerging Technologies (SIGGRAPH'15)*. ACM, New York, Article 4, 1 pages. DOI: http://dx.doi.org/10.1145/2782782.2792495

[19] D. Matsumoto. 1992. More evidence for the universality of a contempt expression. *Motivation and Emotion* 16, 4 (1992), 363–368. DOI: http://dx.doi.org/10.1007/BF00992972

[20] A. Mehrabian. 1972. *Nonverbal Communication*. Transaction Publishers.

[21] H. Nakamura and H. Miyashita. 2010. Control of augmented reality information volume by glabellar fader. In *Proceedings of the 1st Augmented Human International Conference (AH'10)*. ACM, New York, Article 20, 3 pages. DOI: http://dx.doi.org/10.1145/1785455.1785475

[22] M. Ogata, Y. Sugiura, Y. Makino, M. Inami, and M. Imai. 2013. SenSkin: Adapting skin as a soft interface. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST'13)*. ACM, New York, 539–544. DOI: http://dx.doi.org/10.1145/2501988.2502039

[23] M. Ogata, Y. Sugiura, H. Osawa, and M. Imai. 2012. iRing: Intelligent ring using infrared reflection. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST'12)*. ACM, New York, 131–136. DOI: http://dx.doi.org/10.1145/2380116.2380135

[24] R. W. Picard and J. Healey. 1997. Affective wearables. *Personal Technologies* 1, 4 (1997), 231–240. DOI: http://dx.doi.org/10.1007/BF01682026

[25] R. W. Picard. 1995. Affective computing. Perceptual Computing Section Technical Report No.321, MIT Press, Cambridge.

[26] J. A. Russell. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin* 115 (1994), 102–141.

[27] J. Scheirer, R. Fernandez, and R. W. Picard. 1999. Expression glasses: A wearable device for facial expression recognition. In *CHI'99 Extended Abstracts on Human Factors in Computing Systems (CHI EA'99)*. ACM, New York, 262–263. DOI: http://dx.doi.org/10.1145/632716.632878

[28] Y. Tian, T. Kanade, and J. F. Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 2 (Feb. 2001), 97–115. DOI: http://dx.doi.org/10.1109/34.908962

[29] Y. Tian, T. Kanade, and J. F. Cohn. 2011. Facial expression recognition. In *Handbook of Face Recognition*, S. Z. Liand A. K. Jain (Eds.). Springer, London, 487–519. DOI: http://dx.doi.org/10.1007/978-0-85729-932-1_19

[30] Z. Zeng, Roisman G. I. Pantic, M., and T. S. Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (Jan. 2009), 39–58. DOI: http://dx.doi.org/10.1109/TPAMI.2008.52

Received July 2016; revised May 2017; accepted June 2017