Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear

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

Yuta Sugiura
Advanced Institute of Science and Technology (AIST)
Koto, Tokyo, Japan
yuta.sugiura@aist.go.jp

Masa Ogata
Keio University
Yokohama, Japan
ogata@ayu.ics.keio.ac.jp

Kai Kunze
Keio University
Yokohama, Japan
kai@kmd.keio.ac.jp

Masahiko Inami
The University of Tokyo
Bunkyo, Tokyo, Japan
inami@inami.info

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

ABSTRACT
This paper presents a novel smart eyewear that uses embedded photo reflective sensors and machine learning to recognize a wearer's facial expressions in daily life. We leverage the skin deformation when wearers change their facial expressions. With small photo reflective sensors, we measure the proximity between the skin surface on a face and the eyewear frame where 17 sensors are integrated. A Support Vector Machine (SVM) algorithm was applied for the sensor information. The sensors can cover various facial muscle movements and can be integrated into everyday glasses.
The main contributions of our work are as follows. (1) The eyewear recognizes eight facial expressions (92.8% accuracy for one time use and 78.1% for use on 3 different days). (2) It is designed and implemented considering social acceptability. The device looks like normal eyewear, so users can wear it anytime, anywhere. (3) Initial field trials in daily life were undertaken.
Our work is one of the first attempts to recognize and evaluate a variety of facial expressions in the form of an unobtrusive wearable device.

Author Keywords
Wearable; Smart Eye Glasses; Eyewear Computing; Facial Expression Recognition

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

DOI: http://dx.doi.org/10.1145/2856767.2856770

Figure 1. User Wears Our Smart Eyewear

INTRODUCTION
The concept of Affective Computing has been around recently [17] and a number of researchers are gaining more and more insights into affect detection using various approaches [1]. One of the approaches is recognizing facial expressions from video sequences. However, this approach is only reliable in an experimental setting.

Our goal is to recognize people's facial expressions in daily life. Keltner et al. described that our facial expressions provide information about our emotional states [8]. Facial expressions can let us intuit affective states including state of mind, wellbeing, social interaction, etc. because they change spontaneously over the course of a day depending on our mental and physical condition and when we interact with the environment. Although we can tell our own facial expressions, we are not usually aware of our facial expressions or their flow in daily life as it requires cognitive effort. Therefore, we believe quantifying the state of facial expressions in daily life will help us better understand our affective states and lifestyle.

For recognizing people's facial expressions in daily life, we have developed a wearable device that can recognize facial expressions, specifically a smart eyewear. We believe that "wearability" is important for tracking users over the long term. Moreover, as discussed by Picard et al., a wearable can provide a good opportunity to learn about user's affective patterns in daily scenarios [16]. In this approach, we leverage the skin deformation caused by the movement of facial muscles. We use 17 photo reflective sensors that are integrated into the front frame of glasses to detect the skin deformation around the eyes. The sensors we used for our prototype are small and are potentially integrated into everyday glasses. Thus, the usability in daily life is higher than that of the camera-based system in terms of tracking and suitability in social contexts. Our method can consider various facial movements robustly by applying SVM to data acquired from 17 photo reflective sensors that are distributed into the front frame of the glasses.

In this paper, we focus on recognizing eight facial expressions (Figure 2): Neutral, the universal six facial expressions (Happiness, Disgust, Anger, Surprise, Fear, and Sadness) defined by Ekman [3], and contempt, which is sometimes considered as a universal facial expression [12]. We only consider facial expressions since our system only detects skin deformation caused by facial expression changes. We discuss the insights into our affective states acquired from our system. We do not consider complex facial expressions such as a fake smile though they are an interesting area to understand our minds and lifestyle more deeply.

The aims of our work are as follows. (1) Enabling the recognition of eight of a wearer's expressions that are defined as universal (neutral, happiness, disgust, anger, surprise, fear, sadness, and contempt) with a wearable device. We used 17 photo reflective sensors that cover most movements of facial muscles and applied SVM to make the system robust. (2) Social acceptance. We designed and implemented our system as a form of eyewear (fully packaged) because eyewear fits the context of daily usage (Figure 1). Hence, users can wear it anytime, anywhere. To achieve this, we use small photo reflective sensors to capture facial expressions. Those sensors can be integrated into everyday glasses. (3) The observation and analysis of the distribution of facial expressions with long-term use of our system.

Figure 2. Variation of Recognizable Facial Expressions

RELATED WORK

Facial Expression Recognition

There are a number of notable works related to facial expression recognition in computer vision. According to Tien 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 [19]. One AFEA approach achieved the high accuracy of 91.5% for recognizing basic expressions [11] using the Cohn-Kanade Database [7]. However, there are three major problems with camera-based systems in terms of usability in daily life. First, one of the biggest problems is a limited field of view tracking. Most works use a single fixed camera installed in the environment. The camera has difficulty tracking users continuously over a wide area for a long time. Also, users are difficult to track when they move constantly or there is an obstacle between them and the camera. Second, the AFEA camera systems mostly work best with a clear image of a face from a frontal view. However, in real environments, a frontal-view of a face is not always available. A facial image acquired from a non-frontal view reduces accuracy of facial expression recognition. Finally, a camera-based system does not fit social contexts: Even if we solve the problems above with a wearable camera, people still have mental resistance to being captured on camera in their daily lives due to privacy issues. In addition, the camera needs a certain distance from users to capture their facial expressions and so designing the device acceptable to society is difficult. Thus, Kimura et al. proposed a eyeglass-based hands-free videophone [9]. It can yield a frontal face image with fish-eye cameras and capture facial expressions. However, the system is wired to laptop computers, and its size and weight were not comfortable for practical use.

Facial Expression Recognition With Wearable Devices

The one of the first approaches that tried to detect facial expressions with a wearable device was Expression Glasses [18], which can recognize specific expressions (confusion/interest) by measuring facial muscle movement using piezoelectric sensors. Gruebler and Suzuki designed a wearable device that can read positive facial expressions using facial EMG signals [6]. It has to be attached to the side of a face, but it can record the user's affective state for more than four hours with high accuracy. It can be used during therapeutic interventions and to support medical professionals. Li et al. used a depth camera and strain gauge to capture displacement of strain signals from upper facial expressions inside a head-mounted display [10]. They also mapped the input signals to a 3D face model. All of them use contact-base sensors. These prior works showed good performances under experimental conditions. However, the measurement processes are

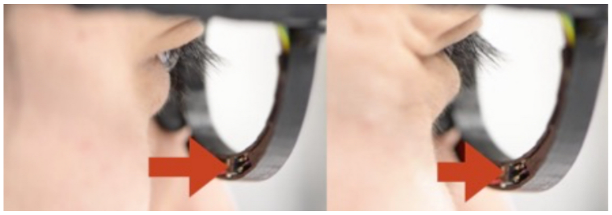

Figure 3. Skin Deformation due to Facial Expression Change

sensitive to losing physical contact with displacement sensors or electrodes. Also, the comfort of wearing the device can be an issue.

Sensing with Photo Reflective Sensors

Fukumoto et al. used photo reflective sensors to capture skin deformation at the corners of eyes and cheeks [5]. The two photo reflective sensors were attached to the outside of glasses to detect happy facial expressions (smiles/laughter) with a practical threshold determination. This method does not scale well for multiple users because of the diversity of the appropriate threshold. In addition, due to the limited number of sensors, it can miscategorize other facial expressions as the target ones. 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 [13].

Other than detecting facial expressions, photo reflective sensors are used for capturing skin deformation on various parts of the human body. For example, Ogata et al. leverage skin deformation of forearms to use skin as an interface (SenSkin) [14]. This work uses 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 leverages skin deformation of the finger to capture finger gestures and external input [15].

FACIAL EXPRESSION RECOGNITION WITH EMBEDDED PHOTO REFLECTIVE SENSORS ON SMART EYEWEAR

One of the enabling technologies for affect recognition in real life is eyewear computing. The emerging field of eyewear computing is perfect for affect recognition, as the head (the location of most of our senses as well as the brain) is a natural position for sensing due to human physiology. The remaining problem is social acceptance, yet with smaller and smaller Printed Circuit Boards(PCB), sensors, and actuators, we can now build smart glasses that are not much different in appearance from normal eyewear, making them acceptable from fashion and comfort perspectives to the general public.

In this study, we use a large number of photo reflective sensors and apply a machine learning method in consideration of skin deformation caused by facial expressions. This approach has three advantages on facial expression recognition. (1) An adaptive and robust recognition on variety of users and their facial expressions due to the rich sensor information and

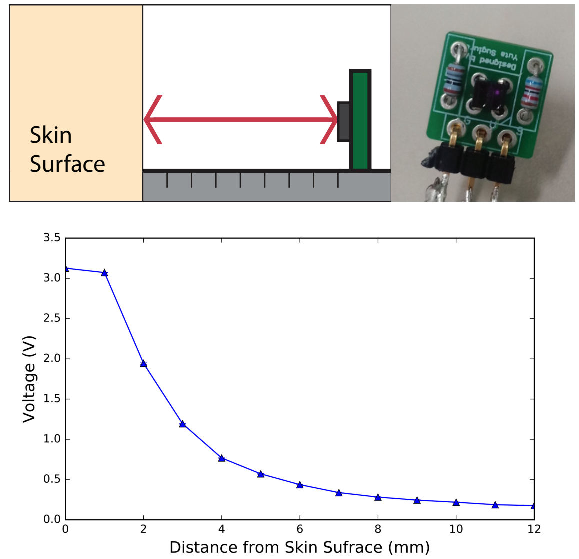

Figure 4. Characteristics of Photo Reflective Sensor to Skin Surface

learning framework. (2) Non-contact measurement: The sensors are noninvasive and they don't need physical contact that is difficult to keep reproducibility and worsen wearing comfort. (3) Smart appearance: The sensors can be integrated into everyday glasses because of the small form factor (e.g SG-105: 3.2 x 2.7 x 1.4 mm). Thus, the users are able to wear the device in various daily scenes.

To capture facial expressions, we leverage skin deformation caused by the movement of facial muscles. When users move their facial muscles, three-dimensional skin deformation occurs. Each facial expression includes different movements of facial muscles. The movements of eyelids, eyebrows, nose, and cheeks cause the 3 three-dimensional skin deformation around eyes. The movement of a mouth also causes the skin deformation under eyes because the muscle movement around the mouth influences cheek shape. According to [4], those movements are the greater part of action units with that the Facial Action Coding System codes human facial expressions. Therefore, placing sensors to capture skin deformation around eyes makes it possible to detect most facial movements related to facial expressions that should be recognized [20]. Skin deformation is captured by measuring the distances between various spots of the eyewear and skin surface on a face. These distances change with different facial expressions. 17 photo reflective sensors integrated into the eyewear frame measure these distances.

In addition, there are two more advantages of our approach. First, it is simple: The processing needed to interpret the readings is minimal, so no elaborate feature extraction needs to be done. Thus, we can keep the processing cost and energy consumption small, which is important for real-time recognition. Second, it is low cost in price: We just use photo reflective sensors with a micro processing unit.

Photo Reflective Sensor

As discussed in Related Work, photo reflective sensors are sometimes used in HCI field to measure human skin movement [5, 13, 14, 15]. We use IR reflective sensors. The sensor is composed of an IR LED and IR photo transistor.

Figure 4 shows a basic characteristics of skin surface reflection captured by an IR photo reflective sensor. We measured the voltage from the sensor that changes in accordance with the distance between the sensor and the skin surface. 30 samples were collected at each position. The figure indicates the average and standard deviation at each distance. The standard deviation is quite low (at most 0.014 V).

The correspondence is not linear, but we can obtain proximity. In the closer distance, the photo reflective sensor shows a higher resolution.

Usage Scenarios

We can assume various scenarios using our device. First usage scenario is "Happiness Map". Given we can achieve facial expression recognition in daily life, one simple way to apply it is to combine facial expressions with location and demography. People can search for the "happiest" place to live or work by looking at others' facial expressions. They can also obtain an overview of how events and policies change the affect state of a region or country.

The second is a "Care System for Older Adults". Facial expressions can provide insights into how people live their lives. Our system can be used as a care system for older adults. They can get an overview of their situation (and potential deterioration of the mood). Since more and more elderly people live away from their children, they tend to feel lonely more easily than before. Our system can help them and their children to take appropriate and productive care of both (e.g., our system notifies the children of the appropriate times to give their parents a call and talk to them when the system detects they are feeling sad and smiling less).

The third is "Collaborative Media Tagging". Facial expressions or emotions can be used to enhance contents. If a lot of people read a book or watch a video with their facial expressions recorded, the content can be indexed and made searchable.Also, content creators could obtain a better grasp on their techniques, making their storytelling more effective (e.g. what surprises a reader/watcher and what does not).

Finally, our device can be used for mental management. Emotional impact is a very delicate problem, as people should feel empowered by technology not influenced or even oppressed. Still given the rise of depression and other mental illnesses related to emotional imbalances, we believe people can benefit from devices that can help them understand and motivate to alter their emotional state. However, these influences should be aligned with the user's long-term goals (not influencing them in "unwanted ways"). Tsujita and Rekimoto outlined a basic example: We can increase people's happiness by making them smile more [21]. Yoshida et al. proposed an emotion evoking system [22]. This mirror system changes facial expressions artificially and revealed that the artificial facial expression can change the user's emotion. Their work shows the

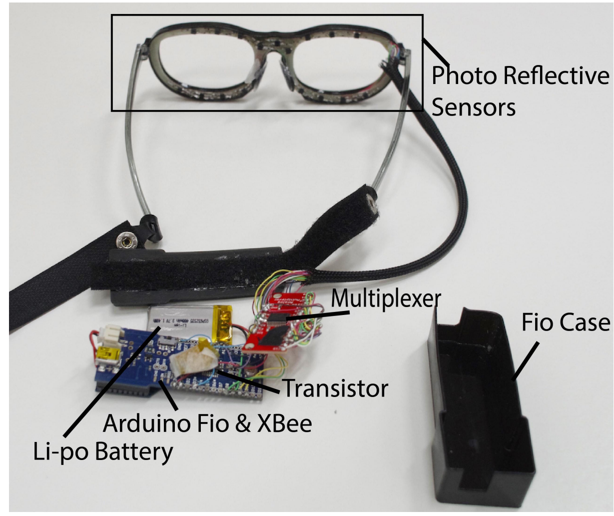

Figure 5. System Components

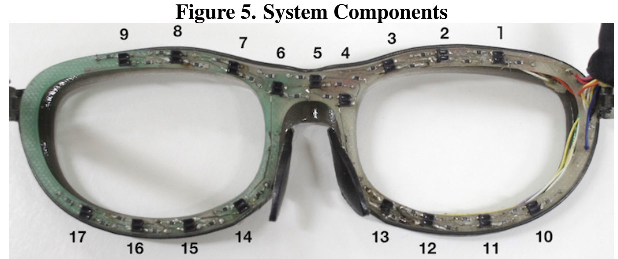

Figure 6. The Placement of Photo Reflective Sensors

interrelation between the cognition of facial expression and emotion. Along those lines, we wonder what would happen if we applied technologies to change some of the symptoms of mental disease. For example, a depressed person usually talks little with others and stays in bed longer (i.e., is less physically active). If a system successfully encourages this person to be more social and active with the assistance of caring people who got his/her information, our work might help deal with his/her mental illnesses. Of course, the proposed solution may not work in every case (e.g. when personalization is needed) and depend on the severeness. However, it may be a promising first step.

IMPLEMENTATION

Hardware

Our prototype includes 17 photo reflective sensors (SG-105), a 16-channel multiplexer (CD74HC4067), a transistor (IRLU3410PBF), Arduino Fio, Xbee, and li-po battery. The weight of the prototype is around 60g. The front frame is 3D printed, and the temple is taken from commercial eyewear. We applied an eyewear band to stabilize the position of the eyewear. The transistor is used to modulate LED of the photo reflective sensors because the sensors are easily influenced by ambient light such as fluorescent light. We measure the difference between the values with LED on and off by switching it on and off frequently at around 80 Hz. With this method, it is possible to reduce the influence of ambient light. Xbee enables serial communication via zigbee at 57600 bits per second.

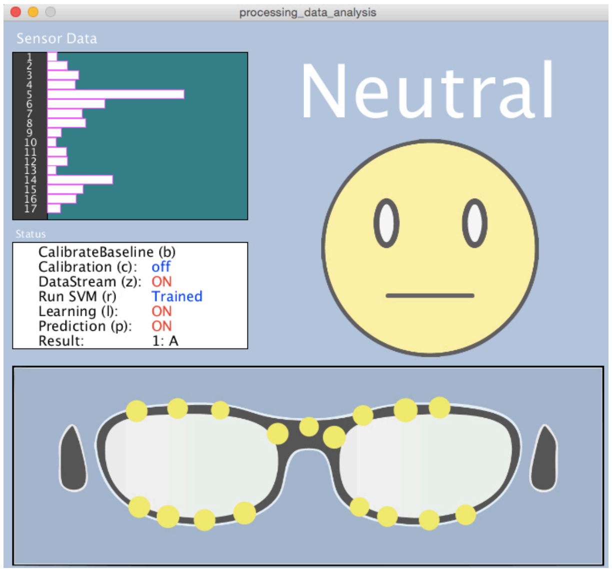

Figure 7. User Interface

Software

In an Arduino environment, input from each sensor is converted into 10-bit data. The data is smoothed out by applying five pieces of data in a row to the moving average in order to reduce the noise and is then sent to Java/Processing. In Processing environment, we normalize sensor data and apply an SVM algorithm with rbf kernel (C = 10, gamma = 1.0) to classify facial expressions in real-time. Figure 7 shows the visualization of our system. The emoticon shows the result of classification. The bar graph represents the raw data. The yellow circles correspond to the positions of each sensor, and their sizes correlate to the normalized values of each sensor. We use only one piece of frame data because we aim at classifying facial expression states.

Machine Learning

1. When a user wears our eyewear device and makes a neutral expression, we set the baseline of each data(BLV) as 0.5.

2. The range of each sensor value ($Range$) is determined during the calibration process. In this stage, a user dynamically moves his/her facial muscles, and the maximum (Max) and minimum (Min) values of each data are determined. With these values, the normalized sensor value (NSV) is calculated as follows. The raw data collected from 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 of each sensor by measuring the range of sensor data that is different among them according to the geometry of each user's face and the position in which each sensor located. The normalization can improve the robustness because it can accommodate the distance changes between the sensors and the skin surface.
   - Disadvantage: The position of the device during calibration phase is not always stable. If a user moves his/her facial muscles too dynamically, this movement changes the position of the eyewear. It changes Max and Min values in an unwanted way. Also if the facial movement during calibration is not enough, it can reduce the information. Therefore, we cannot always normalize the sensor values in the right way.

3. During the learning phase, normalized data is stored with a facial expression label where users choose a desired output for the normalized input. For each label, input from one frame is used.

4. For real-time classification, the normalized value is applied to SVM that is trained with the dataset collected at the previous step.

Algorithm

In addition to the trained data that includes the output label and normalized input, the calculated value (CV) from two different sensor is also used for SVM. The calculation formula for ($S_i, S_j \mid 1 \leqq i, j \leqq 17$) is shown below

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

Sensor placement is shown in Figure 6. We considered the calculated value from adjacent sensors integrated into the lower part of the front frame and sensors that have a vertical relationship. The adjacent sensors of the lower part are important because the surface is smooth and skin is considered as an elastic model, thus they correlate with each other as discussed by Ogata et al. [14]. The data from the sensors that have a vertical relationship also helps because the relationship between the position of the eyewear and the face can be partly considered. All told, we used 33 dimension (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)]).

SYSTEM EVALUATION

For the evaluation, we conducted three experiments. In the first experiment, we obtained a dataset to evaluate the accuracy immediately after user-dependent training with a confusion matrix to prove user dependency. Also, the trade-off between the number of sensors and accuracy was evaluated in this dataset. In the second experiment, we had a trial for multiple day use to check a possibility of long term use. In the third experiment, we proved the robustness while walking. We also described the observations from a demonstration at SIGGRAPH Emerging Technologies 2015.

Evaluation 1 in Basic Setup

We asked eight users (six males and two females / four Japanese, one French, one Chinese, one Taiwanese, and one Sri Lankan / average age: 27.3) to participate in our evaluation. They were asked to sit in a chair and mimic the pictures of an American male making facial expressions described by Ekman [3]. Each time a user looks straight and makes a neutral facial expression, we set the baseline for each sensor. We collected data of eight facial expressions in different poses: Look straight (three times), look up (three times), look down (three times), look left (two times), look right (two times), and take off and put on again (two times). In one recording of each facial expression, 10 pieces of data were collected at regular 50-millisecond intervals. Overall, 1200 pieces of data (8 facial expressions * 10 datasets * 15 times) were collected from all users. All recordings were performed indoors (Figure 8).

Accuracy with User-Dependent Training

First, we tested the accuracy of each user's performance with user dependent training. For each dataset, one recording (10 data) of each facial expression are divided, with the former half as training set (5*8*15 = 600 data) and the latter half as test set (600 data). The training set was applied to SVM with rbf kernel (C = 10, gamma = 1.0). 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 the data of each face direction, our device can classify correctly regardless of the facial expression. Table 1 shows the confusion matrix of the results. As seen in the matrix, disgust can be similar to anger or fear. In addition, surprise and fear are close facial expressions with a 4.7% error to each other.

Table 1. Confusion Matrix (with User-Dependent Training)

		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%

User Dependency

Second, we evaluated the user dependency. We trained with each user's training data and tested with all users' test data. The results can be seen in Table 2. The scores are quite low: Accuracy is 48.0% at most (using training data from User A and test data from User G). This result suggests that, with the current system, a user requires individual training for classification.

Figure 8. Experimental Setting

Figure 9. Evaluation with Different Face Directions

Table 2. User Dependency Matrix

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

Table 3 shows the confusion matrix of all the results. Though there is user dependency, happy expressions can be recognized with relatively high accuracy (71.9%). The other facial expressions did not show good performances (e.g. sadness: 17.2%, contempt: 22.7%).

Table 3. Confusion Matrix(User-Dependency)

		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%

Figure10 shows the distribution of sensor value on each facial expression from User A. The sensor number was associated with Figure 6.

Number of Sensors

Next, we evaluated the trade-off between accuracy and the number of sensors. We used all data from eight users. We separated data into training set (4800 data) and test set (4800 data) and applied SVM in the same way as in Basic Setup Analysis. For the selection of sensors, first we chose one sensor with which the result of SVM showed the best accuracy. Next, we added to SVM one more sensor different from the sensors we had already used to obtain the best accuracy. We repeated the process until all sensors were used. For this analysis, we only applied the normalized 17 dimensional data to SVM.

As Figure 11 depicts, the experimental result shows 84.1% accuracy with 17 sensors. The accuracy improves with more sensors. With 13 sensors, our system achieves more than 80% (81.0%) accuracy. Sensors 1, 9, 12, and 15 are left out. The sensor numbers correspond to those in Figure 6. Sensors 1 and 9 and also 12 and 15 are located at symmetric positions

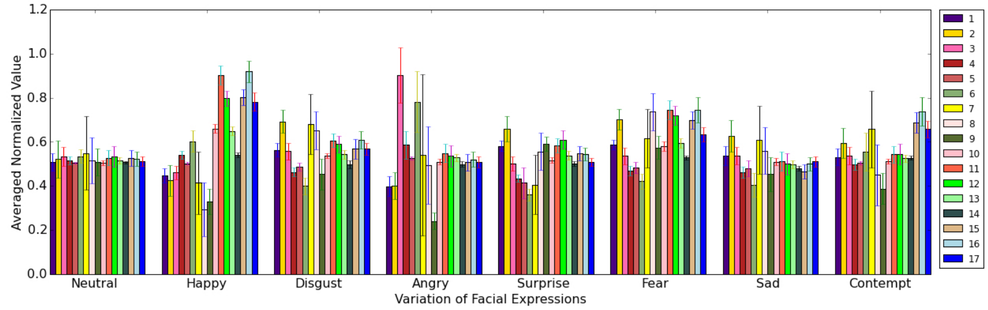

Figure 10. Distribution of Sensor Value on Each Facial Expression

at the opposite ends of the frame. The further the sensor from the center of the face, the longer the distance between eyewear frame and skin surface on face. We think it makes the sensors have less information because they can be influenced more by ambient light and photo reflective sensors works well with closer distance than 10.00 mm.

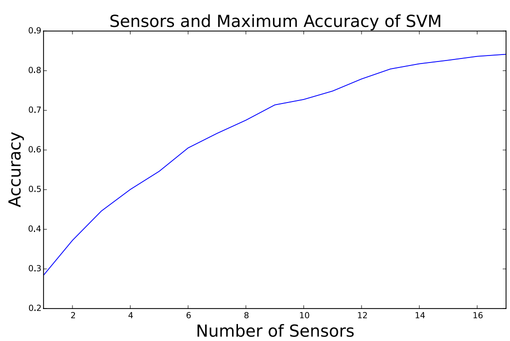

Figure 11. Trade-off of Number of Sensors

Evaluation 2: Robustness for Multiple Use

We also collected data from three of the participants on different days with the same procedure (only when they look straight). For each day, we collected data about eight facial expressions three times (8*10*3 = 240 data). We used data collected from three different days. We used data from two days as training data and the data from the other day as test data. The results are shown in Table 4. On the whole, the averaged accuracy for three users was 78.1%. It suggests the possibility of long-term use. In the confusion matrix, the most dominant error was classifying 33.3% of the anger cases and 23.3% of fear cases as disgust.

Table 4. Accuracy on a different day

	Day1	Day2	Day3	Average
A	83.8%	91.7%	90.4%	88.6%
E	72.9%	70.8%	69.2%	71.0%
F	86.3%	69.6%	68.8%	74.9%

Evaluation3: Usage during Walking

For this experiment, we collected 480 pieces of data from each user. Four users participated (two males and two females / two Japanese, one Chinese, and one German / Age: 25-59). Users were asked to sit in a chair and wear the device. After the calibration, we collected 10 pieces of data of eight facial expressions when they looked straight. We repeated the process three times (Data set A: 240 data). After that, users were asked to walk in a corridor at natural speed. We collected 10 pieces of data of eight facial expressions three times while they were walking. (Data set B: 240 data) We used Data set A as training set and Data set B as test set. The result was an average of 73.2%, which is slight worse than in Evaluation 2. We assume this is because walking caused to change the eyewear position.

Demonstration at SIGGRAPH Emerging Technologies 2015

We have demonstrated our eyewear device at SIGGRAPH Emerging Technologies 2015. During the demonstration, we had more than 200 users from international background. We found that the size and shape of eyewear needs to be adjusted to each user for accurate recognition. There are three major issues we had. 1) The eyewear device sometimes slipped out of place when some user changed facial expressions. It changed the accuracy of the recognition because the device based on the proximity sensing between front frame of the eyewear and skin surface on face. 2) Specific sensors located between users' eyebrows saturated and did not work well for users who have well - defined features despite they did not show any facial expression. 3) The sensors placed on the top measures the distance to eyelid or eyebrows depending on the shape of users' face or the position of the eyewear. Therefore, the meaning of sensor data can be different with each user, requiring individual training and calibration. Due to these 3 reasons, the eyewear should be customized to each user. We observed that the eyewear seems to work better for Caucasian compared to Asian people as they tend to have more expressive faces.

FIELD TRIALS

As initial field trials in daily life, we recorded facial expressions of a user in a daily living/home scenario: A 90 minutes time series consists of playing Go game with a computer (a board game involving two players originated in China, 32 minutes), playing with a dog (16 minutes), watching a sitcom (an episode of "Friends", 22 minutes), and programming on a desk (20 minutes). Also, playing a shooting game with friends(won and lost: 10 minutes) and watching an episode of

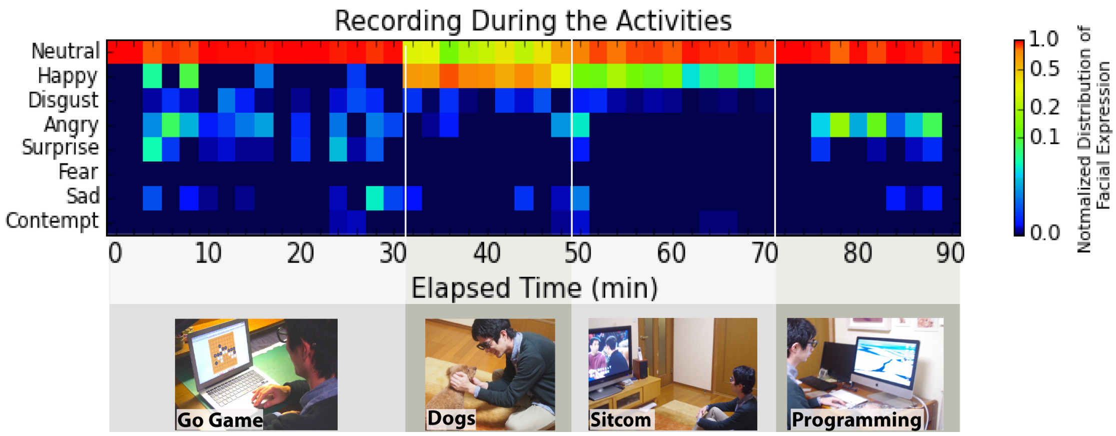

Figure 12. Recording of Facial Expressions for long time

crime drama "Crime Scene Investigation(CSI)" (45 minutes) were recorded on the other day so that we can compare 1) the activities include social interaction (human-human, human-animal: "game", "playing with dog") and the activities include several activities performed alone by the user and 2) the same activity in different contexts (e.g. gaming win-lose, tv comedy-drama). Figure 12 shows the frequency distribution of recognized facial expressions during the recording period. The figure presents a normalized distribution in each 2min with a logarithm contrast enhancement.

The distribution of facial expressions varied on the activities. For instance, many happy expressions were observed during interacting with the dog. The sitcom also caused happy expressions at intervals of neutral ones. While playing Go Game, negative expression was sometimes shown, but at the end, happy expression was dominant probably because he must have won finally.

When the user did the activities by himself, the user tended to show fewer facial expressions. For example, the programming activities induced the angry expression; yet, the dominant expression is neutral. On the other hand, the user shows other more diverse facial expressions (not neutral) while interacting with the dog. The user also showed diverse facial expressions while playing the shooting game with the user's friend. The result is shown in Figure 13. During the game, the user presented more facial expressions such as happy and angry expressions while playing with another person comparing to the time when the user played alone with the computer (the go game). Social interaction seems to cause more facial expressions. In other words, facial expressions can be an indication for communication between people.

However, the distribution of facial expressions also varies according to the results of the interactions (in our case the game)(Figure 14). When the user won a game, the user showed more happy expression, especially in the latter half. During the lost game, the user showed negative expressions and the dominant expression is anger. However, we can still recognize that the user enjoyed the lost game, because the user frequently showed a happy expression.

When watching the one episode of the sitcom "Friends", the user sometimes showed happy expression. On the other day, the user watched the episode of the drama "CSI" that includes the scene such as murder, dissecting the human body and bleeding. The user showed the disgust expression and surprise expression. The user showed the different expressions that correspond to the contents of TV. The distribution of facial expressions can be a good feedback for the contents creators.

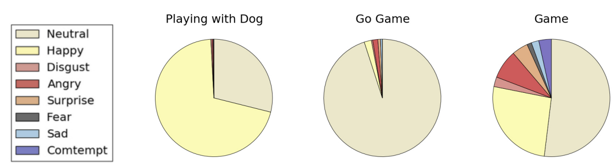

Figure 13. Facial Expression Ratio during Activities

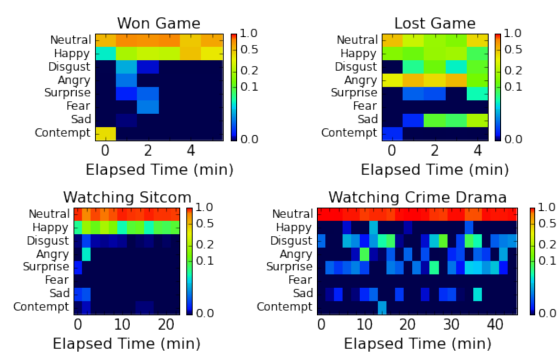

Figure 14. Comparison between Activities

DISCUSSION

We achieved high accuracy 92.8% in the basic setup. It shows a potential of our proposed approach even though many users mentioned that they did not identify the clear difference between surprise and fear, fear and disgust, and disgust and anger when they looked at the pictures for the instruction. If a user makes a very similar facial expression for anger and disgust, most disgust cases may be miscategorized as anger or vice versa.

The user dependency matrix shows the need for user dependent training. In the current conditions, it is necessary to have a personalized training. We do not think this is a big problem as discussed by Scheirer et al. [18], eyewear is a personalized item. Thus, we can design personalized eyewear, and it can be used for a long time by one user. However, we believe external datasets help to reduce the cost of the learning phase and improve the accuracy of recognition. We might try to classify a group of people that share the same culture on facial expressions.

In the long term use and walking situations, the accuracy was reduced. It can be changed due to the physical condition of the device such as the position of the device, artifacts from ambient light. Another factor to reduce accuracy was the difficulty of reproducing the same facial expressions in different conditions and trials. For example, it is difficult for users to make exactly the same specific facial expression many times. They make slightly different facial expressions with different intensities. It makes similar facial expressions difficult to classify. This is controversial because our facial expressions in daily life have the same characteristics. However, it can be solved by increasing the number of training dataset with many trials.

In field trials, "angry" facial expressions were recognized during computer programming or Go game. However, it can be related to mental effort, confusion or frustration instead of "angry" emotion as facial expressions contain the information more than basic emotions. For example, Shichuan et al. suggested the existence of 21 expressions [2]. The information of complex facial expressions is useful in terms of understanding the user in depth, we may consider other sensory inputs to predict them. We regard our work as a first step to recognize complex facial expressions.

LIMITATIONS

We assume only facial expression changes cause skin deformation. However, physiological behavior such as yawning, rubbing one's eyes, and resting one's cheek in one's hand can also cause skin deformation. These can be artifacts of facial expression recognition. Moreover, reflection to skin or sensor value from photo reflective sensors changes in accordance with the condition of a user's face. Factors such as tanned skin, sweat, makeup, and a swollen face in the morning may require users to do external calibration and training.

The eyewear only collects data from the sensors around eyes. Although movement of the mouth is partly detectable, there are several mouth movements our system cannot detect because mouth movement and cheek deformation do not have one-to-one correspondence. For instance, the movement of opening a mouth lowers cheeks slightly, which makes the distance between the eyewear frame and the skin surface longer. Therefore, our system cannot recognize which movement of mouth or cheek makes the distance longer.

CONCLUSION AND FUTURE WORK

We presented a novel smart eyewear that recognizes a wearer's facial expressions. With the eyewear device, we conducted an experiment to recognize eight universal facial expressions. The experimental result shows recognition rates: 92.8% for one-time use and 78.1 % for continuous use in a different day after a training process. Also, we designed our prototype looks like everyday eyewear by using small photo reflective sensors so as to be socially acceptable. This enables facial expression recognition by our smart eyewear in daily life. Our system still has room for improvement in terms of calibration and accuracy, yet we believe it is an important step for quantifying the flow of facial expressions toward assessing affect in daily life.

We regard following three issues as our future work.
First, during the experiments, the participants made intentional facial expressions, although the goal of our work is to capture natural facial expressions. Of course, these are similar, but it is true there is certain gap between intentional and natural expressions. To capture unintentional facial expressions, we need to design a natural learning process.

Second, we use 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 strong ambient light though we apply light modulation to LED of a photo reflective sensor. Therefore, we will design an optical filter to reduce the influence of ambient light.

Finally, we used a classification algorithm, thus the intensity of facial expressions is not measured. This means the current system cannot measure complicated facial expressions such as one that contains a smile and surprise. Since our facial expressions are not always simple in daily life, and the subtle differences may have different meanings in our affect, we need to consider those complicated or subtle facial expressions to detect affect deeply as a future work.

ACKNOWLEDGMENTS

This research was supported by CREST, JST

REFERENCES

1. R.A. Calvo and S. D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *Affective Computing, IEEE Transactions on* 1, 1 (Jan 2010), 18–37. DOI: http://dx.doi.org/10.1109/T-AFFC.2010.1
2. Shichuan Du, Yong Tao, and Aleix 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
3. Paul Ekman. 1989. The argument and evidence about universals in facial expres-sions. *Handbook of social psychophysiology* (1989), 143–164.
4. Paul Ekman and Wallace V Friesen. 1977. Facial action coding system. (1977).
5. Kurara Fukumoto, Tsutomu Terada, and Masahiko Tsukamoto. 2013. A Smile/Laughter Recognition Mechanism for Smile-based Life Logging. In *AH*. ACM,

New York, NY, USA, 8. DOI: http://dx.doi.org/10.1145/2459236.2459273

6. A. Gruebler and K. Suzuki. 2014. Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals. *Affective Computing, IEEE Transactions on* 5, 3 (July 2014), 227–237. DOI: http://dx.doi.org/10.1109/TAFFC.2014.2313557
7. T. Kanade, J.F. Cohn, and YingLi Tian. 2000. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. 46–53. DOI: http://dx.doi.org/10.1109/AFGR.2000.840611
8. Dacher et al. Keltner. 2003. Facial expression of emotion. In *Handbook of affective sciences*. Oxford University, New York, NY, USA, 415–432.
9. Shinji Kimura, Masaaki Fukuomoto, and Tsutomu Horikoshi. 2013. Eyeglass-based Hands-free Videophone. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13)*. ACM, New York, NY, USA, 117–124. DOI: http://dx.doi.org/10.1145/2493988.2494330
10. Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial Performance Sensing Head-mounted Display. *ACM Trans. Graph.* 34, 4, Article 47 (July 2015), 9 pages. DOI: http://dx.doi.org/10.1145/2766939
11. Gwen Littlewort, Ian Fasel, M Stewart Bartlett, and Javier R Movellan. 2002. Fully automatic coding of basic expressions from video. *University of California, San Diego, San Diego, CA* 92093 (2002).
12. David 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
13. Hiromi Nakamura and Homei 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, NY, USA, Article 20, 3 pages. DOI: http://dx.doi.org/10.1145/1785455.1785475
14. Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. SenSkin: Adapting Skin As a Soft Interface. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 539–544. DOI: http://dx.doi.org/10.1145/2501988.2502039
15. Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita 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, NY, USA, 131–136. DOI: http://dx.doi.org/10.1145/2380116.2380135
16. R.W. Picard and J. Healey. 1997. Affective wearables. *Personal Technologies* 1, 4 (1997), 231–240. DOI: http://dx.doi.org/10.1007/BF01682026
17. Rosalind W Picard. 1995. Affective computing. (1995).
18. Jocelyn Scheirer, Raul Fernandez, and Rosalind 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, NY, USA, 262–263. DOI:http://dx.doi.org/10.1145/632716.632878
19. Yingli Tian, Takeo Kanade, and JeffreyF. Cohn. 2011. Facial Expression Recognition. In *Handbook of Face Recognition*, Stan Z. Li and Anil K. Jain (Eds.). Springer London, 487–519. DOI: http://dx.doi.org/10.1007/978-0-85729-932-1_19
20. Ying-li Tian, Takeo Kanade, and Jeffrey F. Cohn. 2001. Recognizing Action Units for Facial Expression Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 2 (Feb. 2001), 97–115. DOI: http://dx.doi.org/10.1109/34.908962
21. Hitomi Tsujita and Jun Rekimoto. 2011. Smiling Makes Us Happier: Enhancing Positive Mood and Communication with Smile-encouraging Digital Appliances. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/2030112.2030114
22. Shigeo Yoshida, Tomohiro Tanikawa, Sho Sakurai, Michitaka Hirose, and Takuji Narumi. 2013. Manipulation of an Emotional Experience by Real-time Deformed Facial Feedback. In *Proceedings of the 4th Augmented Human International Conference (AH '13)*. ACM, New York, NY, USA, 35–42. DOI: http://dx.doi.org/10.1145/2459236.2459243