Poster: Mapping Natural Facial Expressions Using Unsupervised Learning and Optical Sensors on Smart Eyewear

Katsutoshi Masai
Keio University
3-14-1, Yokohama, Kanagawa,
Japan
masai@imlab.ics.keio.ac.jp

Yuta Sugiura
Keio University
3-14-1,Yokohama, Kanagawa,
Japan
sugiura@keio.jp

Kai Kunze
Keio University
4-1-1, Yokohama, Kanagawa,
Japan
kai@kmd.keio.ac.jp

Maki Sugimoto
Keio University
3-14-1, Yokohama, Kanagawa,
Japan
sugimoto@imlab.ics.keio.ac.jp

UbiComp/ISWC'18 Adjunct, October 8–12, 2018, Singapore, Singapore
ACM 978-1-4503-5966-5/18/10.
https://doi.org/10.1145/3267305.3267562

Abstract

Our communication highly depends on nonverbal clues, especially on facial expressions. This paper presents the mapping of spontaneous facial expressions in daily conversation using the optical sensors on smart eyewear and unsupervised learning method(Self-Organizing Map) to see the potentially detectable expressions. We had the case study of five to ten minutes of the unscripted communications with five users. It showed that our system could map the various facial expressions of the users such as social smile and the smile of enjoyment. The study also demonstrated that the map trained with the datasets of five users could categorize the similar expressions of each user into the shared clusters among the users.

Author Keywords

Wearable Computing; Facial Expression; Unsupervised Learning

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction

Humans are social animals that exchange information with others in various contexts [4]. We share various information through languages or nonverbal channels. According to [3],

Figure 1: The smart eyewear we developed

Figure 2: The examples of spontaneous facial expressions in the experiments

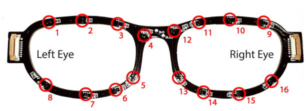

Figure 3: The sensors on the device

among nonverbal information, facial expression plays a vital role in our daily interaction. We can recognize and understand intention and emotion of others unconsciously from facial expressions.

A conventional technique for recognizing facial expressions is Automatic Facial Expression Analysis using camera images [2]. It showed a high accuracy of recognizing basic emotions if a frontal face of the user is available. However, in everyday conversation scenarios, the view of the frontal face is not always available due to the body movement, head-pose direction changes, hand gestures and so on.

In the previous work [6], we have developed the smart eyewear that can measure facial behavior(Figure 1). The optical sensors measure the depth change between the sensor and the skin surface of the wearer. The depth change corresponds to the movement of facial muscles since it causes the skin deformation around eyes. The advantages of using the device are as follows. First, the device is wearable. It can measure regardless of the optical occlusion by the hand gesture, directional change of the faces, and the user's body movement. Also, it is comfortable to wear. However, in the previous work, we only classified basic emotions from the posed facial expressions.

This paper explores the potentially detectable facial expressions using the smart eyewear. We focus on the natural expressions shown in daily conversations that are different from each(Figure 2). We map the recorded facial expressions by the camera based on the sensor data structure using an unsupervised learning method. We recorded the five user's facial expressions in a daily conversation scenario. In such an unscripted communication, the users showed their natural expressions. We used Self-Organizing Map(SOM) [5] for unsupervised classification.

Methods

Our device incorporates 16 optical sensors in the front frame 3. The optical sensors are photo reflective sensors: NJL 5901-AR (manufactured by New Japan Radio Co., Ltd.). We used the sensors located at both ends of the upper part for measuring the ambient light. The information of these two sensors reflects on the movement of head direction since it does not get influenced by the change of facial expressions.

We acquired the sensor data that consists of a 16-dimensional 10 bits values from the device. We also recorded a sequence of pictures from the built-in camera. We synchronized those two data using timestamps. The sampling frequency is about 30 Hz for both. We applied a five sequences simple moving average filter to each dimension of sensor data in order to reduce the noise. Then we eliminated outliers. We normalized the dataset so that the time-series data points in each sensor dimension have zero mean and unit variance.

We used SOM [5] for the unsupervised classification. SOM is an artificial neural network that can summarize the nonlinear data by preserving the topological properties of the inputs. We used MiniSom [7] for the implementation. We mapped the sensor data into a 2-dimensional space with 7 X 7 neurons. For the hyperparameters, we set sigma that determines the range to update the weights of the neurons to 1.0, and a learning rate to 0.4. The initial value of the weight for each neuron was set randomly from one of the sensor data in the dataset. We used all data in the dataset for training. The order of the data for the training was randomized, and we trained with 1500 iterations.

There were many data in one cluster. To determine the facial expression regarded as representative of the cluster (representative facial expression), we chose the picture that

Figure 4: Sensor Data Median Map

Figure 5: Mapped with another dataset

corresponds to the sensor data that is closest to the medians of sensor data in one cluster. This method is fast since there is no need to apply face recognition to all the pictures. Also, we can consider all of the data even when the frontal face is not available in the pictures.

Case Studies of Daily Conversations

We used the approach in the daily conversations to see the potential of classifying spontaneous or subtle facial expressions using the device. We recorded five users' facial behaviors in daily conversations with our eyewear device and a built-in camera on MacBook Pro 2016. They sat down in the same place and talked about random topics such as the holiday experiences, their favorite sites and so on. They talked with two or three friends including the observer. We did not limit the topics and any head movement of the user. We have done all of the recordings indoors.

Reliability

We recorded the 5 minutes conversation of the first user two times. For each recording, we asked the user to start with a neutral face. We trained the SOM using a dataset from the first conversation.

Figures 4 show the visualized maps of the first user's dataset of the former conversation. It was able to map similar facial expressions to the same or close clusters using the sensor data. The method picked out the representative facial expression corresponding to the median of the sensor data in the cluster. For the first user, there were two main expression clusters. The first cluster is a laugh that appeared on the top three rows. The other cluster is a neutral face. Also, The clusters were made by left-right face direction. It is interesting to see that our device has the potential to detect face direction. However, this depends on the place of ambient light sources.

We mapped the dataset of the second conversation to the trained SOM. We normalized each dataset to zero mean and unit variance previously as we can assume the facial expression appeared on the two conversations are not so different. Figure 5 shows the result of the mapping. From the map, almost every cluster has a similar representative expression, yet there were different expressions in the same place (for example, the first row and second column). This result suggests the potential to use our method for a semi-supervised approach. For example, measuring the expressions first with the camera and the device, then the user measure using only the eyewear device. It is useful for long-time recording to see how the frequency of the user's facial expression changed based on the user's facial expressions that had already appeared.

Various Smiles

The second user made a conversation with her friends and the experimenter while all sat down. We recorded the 5 minutes conversation two times, yet the data of the second conversation was lost. We show the result of Sensor Median map in Figure 6.

The characteristics of user 2 are that several kinds of smiles have the different clusters. For example, the right middle areas show the smile of enjoyment where the corner of the eyes wrinkle while the bottom right areas show social smiles that the eyes are neutral [1]. Besides, the map showed the intensity of the smiles as it can show the transition.

The Map Trained with Multiple Users

We trained the map using all five participants' datasets trimmed to five minutes conversation per each. Each dataset was normalized to mean 0 and unit variance respectively and was merged into one dataset. Figure 7 shows the map trained with all users. Each cluster demonstrates the dom-

Figure 6: The Sensor Median map of the user 2

Figure 7: The map trained with multiple users' data

inant user's expression. We show only when the number of one of the user's data occupies the half in one cluster. Overall, the map has consistently shown that the similar expressions are close to each other. It means we could make the clusters across the users. From the left bottom areas of the map, two users' smiles are smaller than the others' smiles of adjacent areas. It suggests that our method can compare how different smiles the users make in the same situation.

Limitation

Only five students participated in the study. We plan to examine how our method can generalize with diverse nationalities and a large number of people in different ambient light conditions. Also, if the facial expressions in the test dataset are not similar to the expressions in the training dataset, the expressions cannot be classified correctly. Moreover, we applied to SOM the static sensor data. The data corresponds to the geometric change. We did not consider time-series information although the dynamics of facial expressions have abundant information about how the facial expressions of the user change.

Conclusion

We explored the potential of the smart eyewear that measures facial behavior using photo-reflective sensors. Especially, we focused on the spontaneous expressions in daily conversations. In the case studies, we mapped the facial images based on the sensor data structure using Self-Organizing Map. We realized that our device can map based on the face direction, the intensity of smiles, various types of smiles. We also showed that the method could visualize the main expressions the users showed. Besides, we were able to map the expressions across the users. In the future work, we evaluate how accurate the device can classify the intensity or various type of facial expressions using the ground truth of human coding.

Acknowledgements

This research is supported by JSPS KAKENHI (YYK8F20) and JST CREST (JPMJCR14E1).

REFERENCES

1. P. Ekman and W. V. Friesen. 1982. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior* 6, 4 (jun 1982), 238–252.
2. B. Fasel and J. Luettin. 2003. Automatic facial expression analysis: A survey. *Pattern Recognition* 36, 1 (2003), 259–275.
3. C. Frith. 2009. Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364, 1535 (2009), 3453–3458.
4. C. D. Frith and U. Frith. 2012. Mechanisms of Social Cognition. *Annual Review of Psychology* 63, 1 (2012), 287–313.
5. T. Kohonen. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, 1 (01 Jan 1982), 59–69.
6. K. Masai, Y. Sugiura, M. Ogata, K. Kunze, M. Inami, and M. Sugimoto. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. New York, NY, USA.
7. G. Vettigli. 2018. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. (2018). `https://github.com/JustGlowing/minisom`