Capturing User-defined Facial Features for Scientific Evidence of Elderly Care

Kosuke Hirayama [1], Sachio Saiki [1], Masahide Nakamura [1,2], Kiyoshi Yasuda[3]
[1]Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada, Kobe, 657-0011, Japan
[2]RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan
[3]Faculty of Informatics, Osaka Institute of Technology, 5-16-1 Omiya, Asahi-ku, Osaka, 535-8585, Japan
Email: hirayama@ws.cs.kobe-u.ac.jp, sachio@carp.kobe-u.ac.jp, masa-n@cs.kobe-u.ac.jp, fwkk5911@mb.infoweb.ne.jp

***Abstract*—To achieve the scientific long-term care supported by interpretable evidence, we propose the "facial expression sensing service" in this paper. The proposed service allows a user to define custom facial features, so as to capture subtle changes of facial expression. Once the features are defined, the service automatically measures and records the values from real-time media stream obtained from a camera. In the operation, the service recognizes a face of a target person within a media stream, measures the features with timestamp, and records the data in a database. The data can be used to understand how the target person changes the emotion before, during, and after the care treatments. To show the practical feasibility, we conduct an experiment that investigates emotion of elderly people talking to a virtual agent.**

***Index Terms*—Stream mining, Facial expression analysis, Long-term care, Care effect**

I. INTRODUCTION

Japan is facing a super-aged society. The number of elderly people who need care is increasing, which leads to a chronic lack of care resources. Under this circumstance, the Japan Government has declared the practice of "scientific long-term care" as a national policy [1]. It aims at optimal use of care resources by corroborating the effect of the care by scientific evidence. In order to achieve scientific long-term care, it is essential to evaluate the effect of the care objectively and quantitatively. As an objective metric to quantify the care effect, we focus on "facial expressions". The facial expressions give useful hints to interpret the emotions of a person under care. As a practical example, there is a systematic methodology called facial expression analysis [2].

To analyze one's face and emotion automatically, we can use emerging image processing technologies. Especially in recent years, owing to powerful machine learning technologies, various companies provide cognitive APIs for face, which automatically recognize and analyze faces within a given picture. These APIs produces values of typical emotions such as happiness and sadness, which would be helpful for the evaluation of care effect. In actual medical and care scenes, we have to detect subtle changes of a face that may not be covered by the existing APIs. This is because the facial expressions of elderly people under care tend to be weakened due to functional and/or cognitive impairment. Also, the emotion analysis conducted by the existing API is a black box. Therefore, it is difficult for caregivers to understand why and how the value of the emotion is measured.

To solve such problems, we propose the "*facial expression sensing service*". The proposed service allows a user to define custom facial features, so as to capture subtle changes of facial expression. Once the features are defined, the service automatically measures and records the values from real-time media stream obtained from a camera or a video file. To define the custom facial features, an user first selects some points on facial landmarks, such as eyes and a nose. The user then defines a feature by the distance between the points. In the operation, the service recognizes the face of a target person within a media stream, and measures the distance between every pair of the selected points. Simultaneously, the service attaches a timestamp to each data, and records the data in a database. The user can retrieve the time-series data, and use the data for various analyses. Thus, the proposed service enables to obtain custom facial features efficiently. Moreover, the analysis of subtle facial changes and the explainable emotion analysis becomes possible, and the user's expertise of emotion analysis also can be cooperated.

In this paper, we have specified functions of facial expression sensing service and implemented the service. As a preliminary experiment, we have measured changes in facial expressions from videos of elderly people undergoing care by using the proposed service. Also, we have compared the result from the proposed service and that from the existing API.

II. PRELIMINARIES

A. The current situation around scientific long-term care

As we described in Chapter I, to evaluate the effect of care quantitatively and objectively is important for scientific long-term care. However, the evaluation of care tends to be conducted with the methods which depend on a subjective scale, such as observations and questionnaires. Therefore, we face a difficulty to use the obtained data as scientific evidence. Moreover, these assessment methods compel a heavy burden on care evaluators, subject people, or both.

In addition, since scientific care needs using large-scale data, efficient data collection is required. In the past, to measure care effects objectively, the experiment that measured the changes in the facial expressions of a subject undergoing care was conducted [3]. In this experiment, facial features such as the height of eyebrows and the opening of eyes are measured from images based on methods of P. Ekman et al. However, they measured the features manually from the recorded video. In addition, in the report, they have been concluded that "...Currently, we have not generalized this study successfully because of the limited number of cases. ... To establish an objective evaluation method, We have to collect and analyze both the data of the objective people and the caregiver more."

From the above, to realize practicing scientific long-term care, we may have to develop a computer system that enables us to automatically collect data that contributes to the quantitative assessment of care.

B. Cognitive API

With the rapid technological growth of cloud computing, now we can use cloud services to process various tasks which are previously executed on edge sides. Under such a kind of background, various services utilizing artificial intelligence, which becomes explosively popular in recent years, are now getting available as a cloud service. One of the actual cases of cloud service using artificial intelligence is cognitive services. Cognitive services extract useful information from visual/sound/linguistic/knowledge data by analyzing them and enable computers to recognize these types of information. Using cognitive services can realize functions about cognition which are previously difficult for computers. Many public clouds provide these kinds of services as "cognitive API". For example, cognitive APIs for vision can extract various information from an input image.

C. Emotion analysis with a cognitive API

Cognitive APIs can automatically quantify emotions, so we might utilize them to scientific long-term care. In the past, our research group has developed a real-time emotion analysis service "Face Emotion Tracker" using a cognitive API and has attempted a quantitative evaluation of care [4]. In this previous study, they conducted care using a virtual agent [5] for 5 elderly people living in an elderly facility. During the care, FET captures pictures of a subject person and send them to a cognitive API (Microsoft Face API [6]) to analyze the emotion. A part of the result of the FET experiment is shown in figure 1 as time-series graphs (Some outliers has been eliminated). FET outputs 8 kinds of emotions ("anger", "contempt", "disgust", "fear", "happiness", "neutral", "sadness", "surprise") from 0 to 1 as probabilities, and this time, we pay attention to the value of "neutral" and "happiness". Subject A was relatively expressive, and the values of "neutral" and "happiness" appear alternately during the care. Subject B was 99 years old, and the change in facial expression was not so apparent because of the decline of facial muscles. As a result, in contrast to subject A, only neutral was observed at almost all timings. However, the accompanying person who observed the subject people evaluated that subject B felt better and better as the care progressed. So we can say that there was a gap between the obtained results and the actual. Thus,

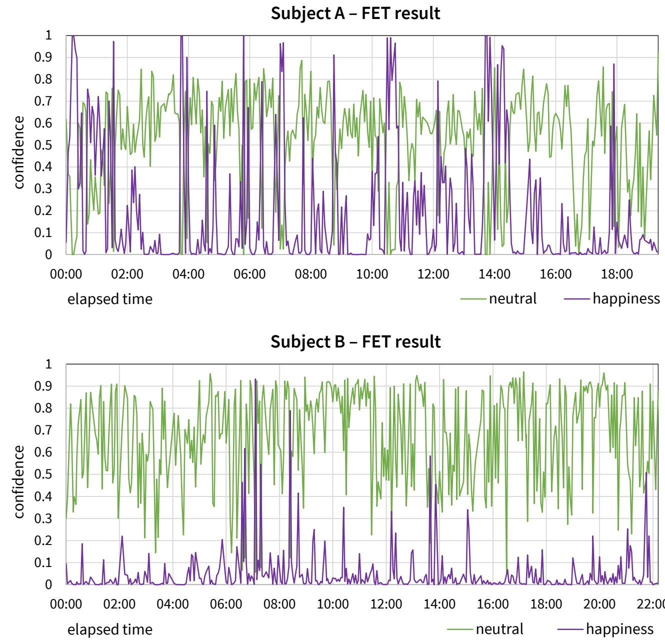

Fig. 1. A part of the result of the FET experiment.

the reliability of the result of emotion analysis with cognitive API may be degraded depending on the characteristics of the subject person.

When observing the facial expressions of care targets, we will face many cases toward people with poor facial expressions for their actual emotions. The factors are aging, dementia, the sequela of stroke, ALS (Amyotrophic Lateral Sclerosis), and Parkinson's disease, and so on. Therefore, we will need to take some measures.

Besides, the cognitive API which FET used is built with machine learning technology. Therefore, we cannot follow the process to a result outputted from an input. As a result, we may face difficulty when we are going to use the obtained data as evidence for the effect of care.

III. Facial expression sensing service

In this chapter, regarding facial expression sensing service stated in I, we consider and explain functions that will be necessary with the service.

A. Summary

We aim to provide two functions to users with this service.

1) *Measure facial features*
   When estimating emotions from facial expressions, what facial features should be measured may vary widely depending on the purpose and subject person. For this reason, we enable users to freely define the required features and make the service can be applied in any analysis case. Besides, users may need to measure the person whose facial expression weakened. For this, the service should be able to record even the most subtle facial changes. To deal with such cases, we make the service continuously record facial features as time-series

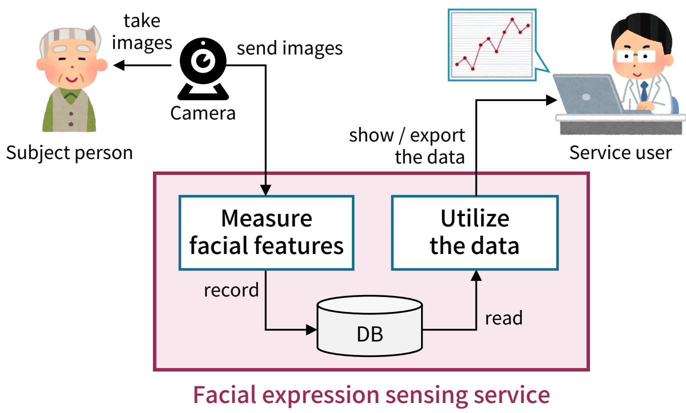

Fig. 2. Conceptual diagram of facial expression sensing service.

data. By providing these functions, the service will be able to contribute to the subtle and explainable emotion analysis.

2) *Utilize the data*
It is necessary that users can utilize the obtained record of features. The service visualizes the time series data of features as a graph to assist the user's facial expression analysis task. Also, the service can export the data so that users can utilize them for various purposes.

We show a conceptual diagram of the proposed service in Figure 2. Firstly, the service gets the face image of the subject person from the camera. The image capture continues at a specific time intervals. Next, the service extracts the positions (= *facial feature points*) corresponding to each part of the face such as eyebrows, eyes, nose, and mouth from the image, and measures facial features according to the profile defined by users in advance. Finally, the service displays the measured features on the screen and saves them in the database sequentially. The user can view the saved data later with respect to each measurement.

In the next section, we will explain each function in detail.

B. Explanation of each function

1) Measure facial features: In this service, we define the distance connecting two feature points as "*facial features*". The user can define features to be measured by selecting arbitrary feature points. The user also can define the average of the other two features as a new feature. The user can create groups from multiple features as a "feature set".

When measuring features, the user has to create "measurement session". To create this, the user selects a predefined feature set. The service will measure features that are included in the feature set selected by the user. All features in the set will be measured simultaneously. In addition to the feature set, the user can input the session name, the name of the target person, and a memo.

When the measurement starts, the service receives images from the camera connected to the computer at specific time intervals. The user can adjust this interval. The service detects the face from the obtained image and then extracts the facial

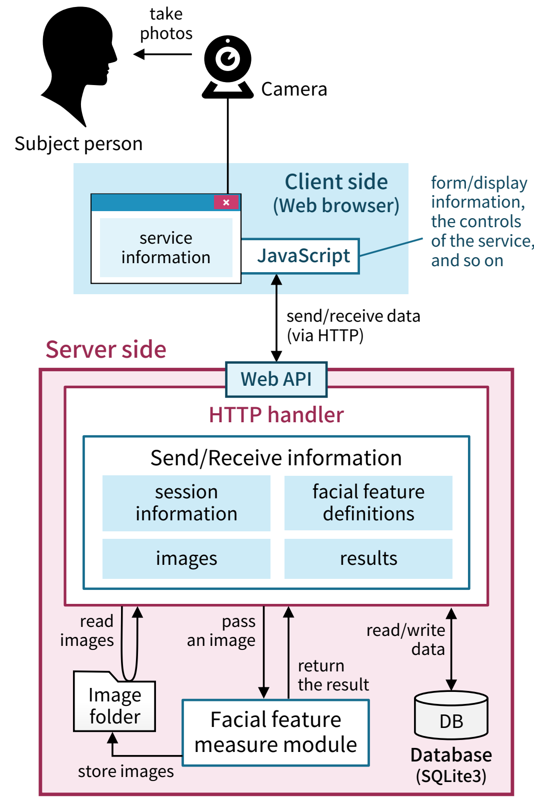

Fig. 3. Overall implementation of facial expression sensing service.

feature points. Then the service measures the value of features according to the user definition. After that, the service displays the measured features to the user and saves them in the database. At the same time, the service saves the face image too. This enables the user to check the features and the face at the same time when viewing the measurement results later.

In addition to the above-mentioned real-time measurement functions, the service accepts video files to automatically measure the features. As a result, the user can conduct analyses using a pre-captured video.

2) Utilize the data: Users can view the records of features saved during a measurement session by selecting the session name. The service reads the records specified by the user from the database and displays them as time-series graphs. The values of features are plotted as points on graphs. When the point is clicked, the service displays the corresponding face image. In addition to display graphs, the service can write out measurement results to a file and users can download it.

IV. Implementing the Prototype

A. Implementation

We implemented the prototype of facial expression sensing service as explained in the previous chapter. We built this as a Web application, so users can use it on a Web browser. Figure 3 shows the overall implementation of the service.

We use the following technologies.

- Development languages: Python3, HTML5, JavaScript

- Libraries: Django[1], OpenCV[2], dlib [7], Plotly[3]
- Others: Celery[4], RabbitMQ[5]

Django is one of the web application frameworks for Python. We used this to build the service. OpenCV and dlib are image processing libraries that can be used from Python. The service uses OpenCV for basic handlings such as image reading and dlib for face detection and facial feature point extraction. Plotly is a library for drawing graphs. Plotly is used from JavaScript on a Web browser. Celery is a task queue and RabbitMQ is a message broker. These two are used to exchange data with the Facial feature measure module. In this prototype, the database for storing facial features definitions and measurement results is built into the application using Django's default SQLite3[6] handling function.

B. Measure facial features

To extract facial feature points, the service uses dlib's feature point extractors (`dlib.shape_predictor`) and learned files[7]. This file defines 68 points shown in Figure 4 as feature points, and these points are used by the service.

The service can measure three types of the distance between two feature points as features, that is height (vertical distance), width (horizontal distance), and linear distance (vector distance). Before measuring features, the service normalizes the face using the distance d_{eye} between the centers of both eyes in order to obtain a constant value regardless of the face closeness. Specifically, if the distance (one of the three types described above) between two feature points is d, the value of the feature will be d/d_{eye}. In addition, when calculating height or width, to consider the lean of the face, the service rotates the axis using the angle between the centers of both eyes before measuring d. Figure 5 shows a conceptual diagram of this rotation process. When the slope of the line connecting both eyes subtend the horizontal direction of the image is θ, and the coordinates of points in the original coordinate axis are (p, q), the modified coordinates (p', q') is calculated by equation (1).

$$\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \quad (1)$$

The service uses this modified coordinate to measure height or width.

Figure 6 shows the actual screen to measurement. The web browser sends an image to the service, and the service measures the features. For each measurement, the values of features and an image showing the feature points used for the measurement is displayed on the Web browser.

As we mentioned in III-B1, the service can also measure features from video files. Users upload a video file (such as .avi or .mp4 file) to the service. At this time, users create a measurement session in the same way as during real-time measurement and set the measurement interval. After this, the service extracts frames from the video file at specified intervals and measures the features and records them.

[1]https://www.djangoproject.com/
[2]https://opencv.org/
[3]https://plot.ly/
[4]http://www.celeryproject.org/
[5]https://www.rabbitmq.com/
[6]https://sqlite.org/index.html
[7]http://dlib.net/files/shape_predictor_68_face_landmarks.dat.bz2

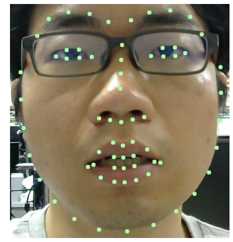

Fig. 4. Feature points that the proposed service extracts.

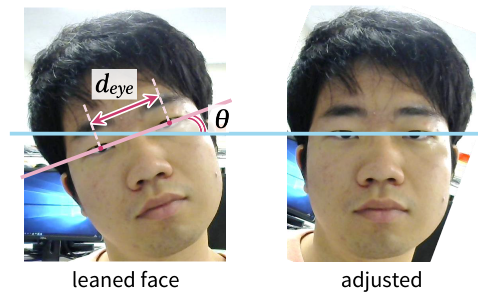

Fig. 5. Conceptual diagram of adjusting leaned face.

C. Utilize the data

All records of features obtained by measurements are saved in the database. Face photos are stored in a particular directory in the application server.

The service offers users the record browsing screen. First, the user selects the name of the session to view the record. Then, a list of features is displayed. Next, the user selects features to check the values. The service loads the selected features' records from the database and sends it to the Web browser. The Web browser uses the Plotly library to create and display a time series line graph from the received data. Figure 7 shows the graph generated by the service. By using the functions of Plotly, users can move and scale the graphs freely. In addition to displaying raw data, the user can view a moving average of any number of frames. If you click the "Download (CSV)" button, a CSV file containing time-series data of all the features in the session is generated, and the user can download it via a Web browser.

V. Preliminary Experiment

A. Summary

We have tested to use facial expression sensing service (in this chapter, from here we refer it as "the proposed service") and emotion analysis as a preliminary experiment. For this, we used the data of five elderly people that had been obtained by the previous experiment of FET mentioned in II-C. This data includes time-series data of emotion values and face images of

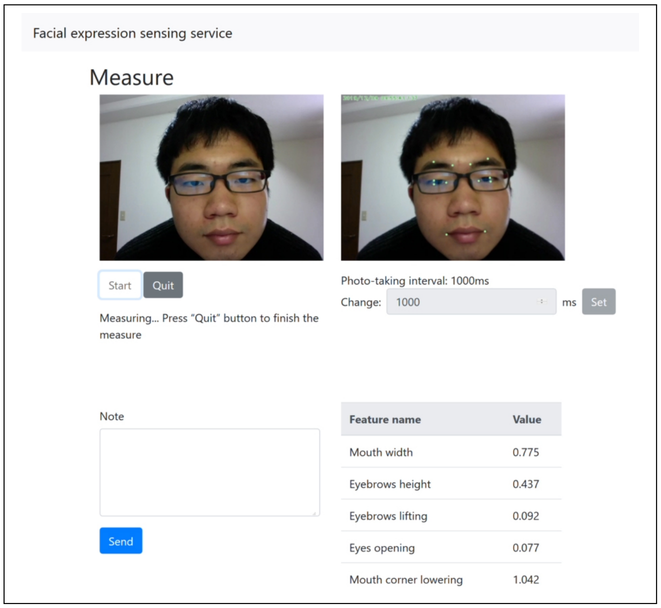

Fig. 6. The measurement screen.

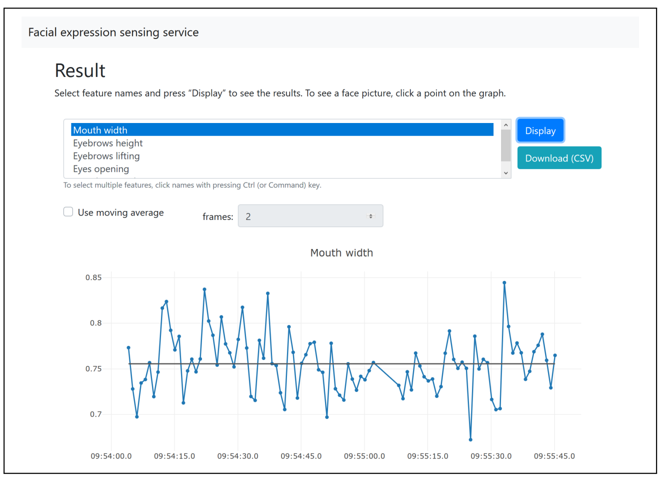

Fig. 7. The result displaying screen.

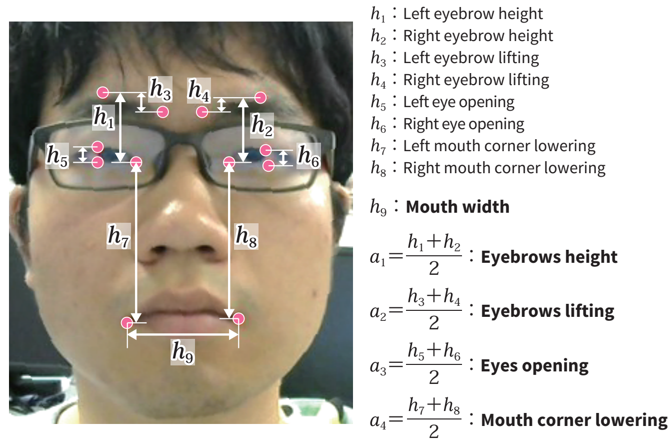

Fig. 8. Defined features for the preliminary experiment.

TABLE I
THE CORRELATION BETWEEN THE "HAPPINESS" VALUE OF FET AND THE VALUES OF FEATURES MEASURED BY PROPOSED SERVICE

	Mouth width	Eyebrows height	Eyebrows lifting	Eyes opening	Mouth corner lowering
Subject A	0.18282	-0.23628	-0.05432	-0.00720	-0.23862
Subject B	0.05201	-0.01742	-0.01372	0.15207	0.22380
Subject C	0.69085	0.35611	-0.35604	0.25550	-0.72197
Subject D	0.69093	-0.15822	0.06826	-0.39886	-0.70356
Subject E	0.11818	-0.01107	0.01582	-0.01585	0.00205

TABLE II
DETAILS OF EACH SUBJECT PERSON

	Age	Sex	Notes
A	80	M	expressive, experienced a stroke (cerebral infarction)
B	99	W	showed little facial expression changes
C	73	W	expressive, needed support
D	84	W	expressive, needed support but no problem with conversation
E	80	W	had severe dementia, always smiling and suspected of euphoria

the care subjects corresponding to each measured value. First, we extracted facial images from the FET data and converted them into video files considering the measurement interval. Next, we input the video files to the proposed service, and measured the 5 features "Mouth width", "Eyebrows height", "Eyebrows lifting", "Eyes opening", "Mouth corner lowering" which are shown in Figure 8 as h_9, a_1, a_2, a_3, a_4. We defined these features based on [3]. Then, we compared emotional values by FET and the values of features by the proposed service. Besides, we excluded some outliers from the FET data. The experiment has been approved by the Research Ethics Committee of Kobe University Graduate School of System Informatics (No. R01-02).

B. Results

We investigated the correlation between the emotional values of FET and features by the proposed service. The results are shown in TABLE II. Subject A and Subject B are the same people shown in figure 1. All subjects are Japanese.

The high correlation was seen on Subject C and D. For these two people, "happiness" was positively correlated with "mouth width", and negatively correlated with "Mouth corner lowering". From this result, we can see that when they laughing they widened mouth and raised mouth corner (because "Mouth corner lowering" will decrease if mouth corner raised), and such facial expression changes appeared to the results. However, the signs of the correlation with the other features do not match between two subjects, and we can assume that the facial movement when laughing slightly varies among individuals.

Subject A was also expressive like Subject C and D, but the correlation was relatively low compared with theirs. However, we can see a slight correlation between "happiness" and the "mouth width" and "Mouth corner lowering", which similar to those of Subject C and D. In fact, we looked at the face

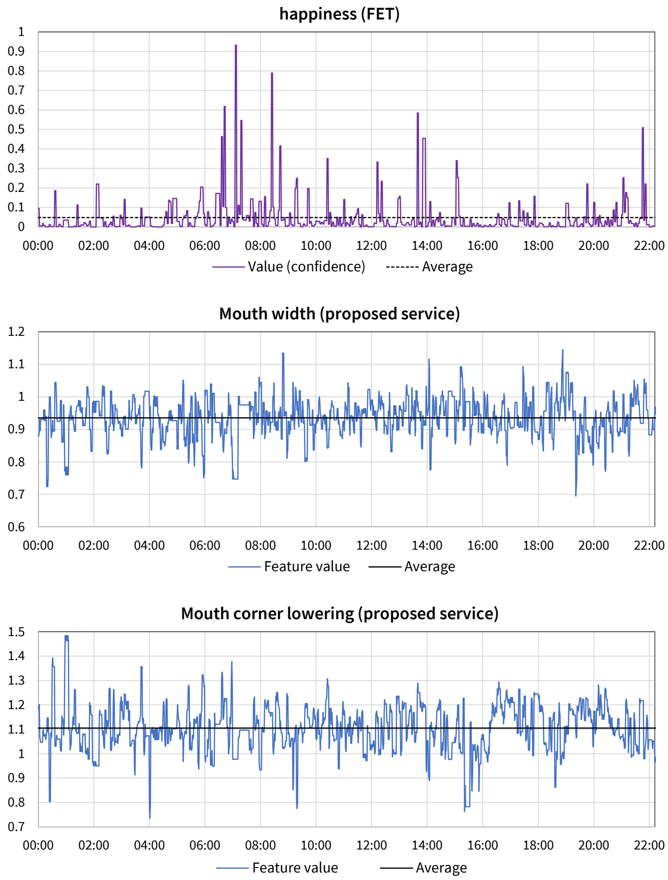

Fig. 9. The graphs of the measurement results of Subject B by FET and the proposed service. The horizontal axis represents elapsed time.

where the mouth widened and we could see a smiling face. Comparing FET emotion values with Subject C and Subject D, the total time that Subject A showed high "happiness" value was relatively short. As a result, it seems that the change in the value of features other than the moment showing "happiness" became noises when taking the correlation.

We could see almost no correlation for Subject B. The reason may be that FET has not been able to capture emotions.

Subject E showed a smiley face almost always. Therefore, the value of "happiness" was always high during care. Like Subject B, the correlation was low.

C. *Discussion*

As the results showed in V-B, for expressive subjects, the correlation between the result of FET and that of the proposed service was higher. In other words, the features of the proposed service followed the FET results, and it seems to further support the FET result. For this result, when targeting expressive people, it is expected that the results of emotion analyses using machine learning can be more quantitatively and explainable by the proposed service. In addition, from the results obtained by the proposed service, we were able to observe the changes in facial expression that occurred when subjects showing feelings of happiness as numerical data. Therefore, it seems that the proposed service also can be applied to analyses performed by human evaluators.

For not expressive subjects, We could not see a high correlation. However, the proposed service was able to observe some subtle facial movements. Figure 9 is the graph of the "happiness" value by FET of Subject B and a part of the values of features by the proposed service. Even when the value of FET hardly changes, the values of the proposed service changed. If we could use this value effectively, we will be possible to analyze fine facial expressions and evaluate emotions that could not be captured in the past. To realize it, further research is needed in the future.

Besides, the feature point extraction of the proposed service often became unstable. This problem was seen especially when the subject person leaned the face. The positions of feature points affect measuring features, therefore, we have to address it as much as possible by improving the implementation.

VI. Conclusion

In this paper, we have considered and implemented "facial expression sensing service" which enables users to define facial features themselves and measure and record the values. We also have conducted a preliminary experiment using the implemented service. By using this service, we can obtain facial expression data efficiently. Furthermore, it seems that it can contribute to realizing the explainable emotion analysis by subtle changes in facial expression.

As future research topics, we would like to operate this system and research how effective it can be in actual situations such as quantitative evaluation of care. In addition, improvements in implementation such as more stable facial feature point extraction and expansion of service function will be needed. We would also like to consider combining factors other than the face, such as the tone of the voice, to realize more powerful care effect evaluation.

Acknowledgment

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP17H00731, JP18H03242, JP18H03342, JP19H04154, JP19K02973.

References

[1] Headquarters for Japan's Economic Revitalization (No.26), "Future investment strategy (2017)," http://www.kantei.go.jp/jp/singi/keizaisaisei/dai26/siryou.pdf.

[2] P. Ekman and W. V. Friesen, *Unmasking the Face*. Malor Books, 1975.

[3] K. Sato, K. Sugino, and T. Hayashi, "Development of an objective evaluation method for the nursing-care education based on facial expression analysis," *Medical and health science research : Bulletin of Tsukuba International University*, vol. 1, pp. 163–170, 2010.

[4] A. Sako, S. Saiki, M. Nakamura, and K. Yasuda, "Developing face emotion tracker for quantitative evaluation of care effects," in *Digital Human Modeling 2018 (DHM 2018), Held as Part of HCI International 2018*, vol. LNCS 10917, July 2018, pp. 513–526, las Vegas, USA.

[5] S. Tokunaga, K. Tamamizu, S. Saiki, M. Nakamura, and K. Yasuda, "VirtualCareGiver: Personalized Smart Elderly Care," *International Journal of Software Innovation (IJSI)*, vol. 5, no. 1, pp. 30–43, Oct 2016, http://www.igi-global.com/journals/abstract-announcement/158780.

[6] "Face API - Facial Recognition Software — Microsoft Azure," https://azure.microsoft.com/en-us/services/cognitive-services/face/.

[7] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.