Internal state estimation by thermal image and identification of face and nose position

Yuta Watanabe, Yoshitsugu Manabe, and Noriko Yata
Chiba University, Chiba, Japan
E-mail: yuuta0407@chiba-u.jp, manabe@faculty.chiba-u.jp, yata@chiba-u.jp
Tel/Fax: +81-43-290-3261

***Abstract*—There is a close relationship between body temperature and activity, such as decreased body temperature when drowsiness occurs in humans. Therefore, we can estimate drowsiness by measuring changes in human body temperature. In particular, this paper focuses on the facial area that includes the forehead, which is thought to reflect the temperature inside the body, and the area around the nose, where changes in skin temperature are significant. At that time, the body temperature is measured non-contact using an infrared camera without attaching electrodes or sensors. The measurement is performed without giving stress to the user. In addition, when expanding the target to multiple people, it is necessary to detect the face positions of multiple people. However, most systems that can measure temperature without contact currently use visible images simultaneously to detect the position of the face. Therefore, in order to estimate drowsiness for multiple people, the positions of the faces and noses of multiple people should be estimated using only infrared cameras. In the experiment results, the drowsiness was estimated to be almost correct. It was shown that position estimation is effective even when it is expanded for multiple people.**

I. Introduction

Human body temperature is not always constant and fluctuates in our lives. In addition, human body temperature is lowest in the early morning, gradually rises over time, and is highest in the evening. There is a daily temperature rhythm called the circadian rhythm. In this way, human body temperature and activity are closely related [1]. Therefore, the state of activity of the target person can be estimated by measuring the change in human body temperature. This study focuses on drowsiness and estimates the drowsiness. Humans are characterized by a decrease in deep body temperature when drowsiness occurs and an increase in body temperature at the edges of the body. Especially in the face area, the forehead part reflects the temperature inside the body, and the change in skin temperature is remarkable around the nose. This study estimates the drowsiness by measuring changes of face area without contact using an infrared camera. This is a response to the problems of conventional measurement methods. Many studies related to the estimation of drowsiness have been conducted. For example, Nakano et al. detected the driver's alertness using a photoelectric pulse wave sensor on the earlobe [2]. However, the method creates a feeling of restraint and leads to stress because the sensor is attached directly to the body. In addition, it may come off depending on the movement of the subject. Therefore, if non-contact measurement is available, states estimation becomes possible without giving stress to the subject. Infrared cameras also have the advantage of being less affected by changes in lighting, so they are also used in systems that capture human behavior [3]. However, for non-contact measurement, some studies using RGB images have also been conducted, and they can measure pulse rate in non-contact using RGB images [4]. By combining this study with the feature that the pulse slows down when humans become drowsy, it is thought that non-contact drowsiness estimation can be realized. However, this study estimates drowsiness in a non-contact manner using thermal images, which are less susceptible to changes in lighting than RGB images and can be considered for privacy because they are not projected clearly.

Currently, drowsiness estimation targets one person, but it is considered necessary to expand the target to multiple people in the future. For this purpose, it is necessary to detect the face positions of multiple people. However, most current systems that can measure temperature without contact detect the face from visible images and not from thermal images. The reasons are that there are little thermal image data, position estimation by machine learning is more complex than with visible images, and infrared cameras are more expensive than ordinary visible light cameras. In addition, figure 1 shows an RGB image, a grayscale image and a thermal image. Comparing these images, the thermal image is not clearly projected as a whole and is an uneven image. In particular, it can be confirmed that the boundaries of parts such as the nose and eyes are ambiguous. From these facts, it is expected that the application of ordinary face detection methods to thermal images will be severe. Jin-Sup Eom et al. proposed a method that enables the position of the face and nose to be detected by thermal images for measuring respiratory rate from thermal images [5]. This method uses a face detection method called the Viola-Jones method [6] to estimate the position by narrowing down the candidates by constraints. However, it does not support multiple people and has not been verified for facial tilt. Also, as shown in Figure 2, there was a problem that detection failed if the difference in body temperature between the nose and its surroundings was small. Therefore, this study proposes a face and nose detection method for multiple people using only thermal images obtained from the infrared camera used for drowsiness estimation.

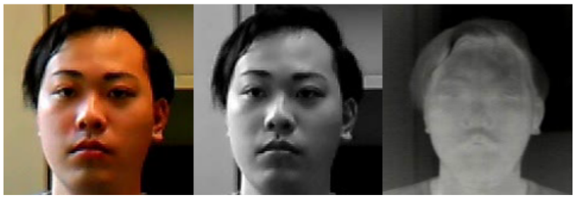

Fig. 1. Comparing images: RGB, grayscale, and thermal image.

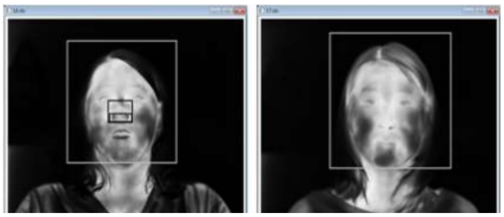

Fig. 2. An example where the temperature difference between the nose and its surroundings is small[5]

II. Methods

A. Estimation of drowsiness

The thermal image obtained from the infrared camera is binarized and labeled to obtain the maximum area, and that part is treated as the face area of the target person. Binarization uses the minimum part near the center of the histogram as the threshold value. At this time, if the number of pixels is too small, that value is not used as the threshold value. Labeling is done 8 neighborhood and uses a look-up table. After that, only the largest area is left and the other areas are blackened. There is a difference in skin temperature between the forehead and the area around the nose on the human face. In the case of drowsy, the internal heat is carried to the end of the body, so the skin temperature on the face rises, especially around the nose. However, the temperature of the forehead decreases because it reflects deep body temperature. As a result, the difference in skin temperature between the forehead and the area around the nose is thought to be small. Therefore, in order to obtain each skin temperature, the face area is vertically divided, as shown in Fig. 3. From the divided image, the average of the brightness values of the forehead and the area around the nose is obtained. The difference between them per unit time is observed, and it is assumed that the state change occurs when there is a large change. Room temperature is also measured to determine whether changes in room temperature affected changes in skin temperature. From the data of internal changes estimated by the above method, it is classified into three states: wakefulness, drowsiness, and sleep. From the obtained data, it is assumed that an internal change may have occurred if the amount of change in skin temperature per unit time exceeds a certain value, and the amount of change in room temperature is investigated to determine whether it affected the change in skin temperature. If it is found that the change in room temperature is not affected by the change in skin temperature, it is considered that an internal change has occurred.

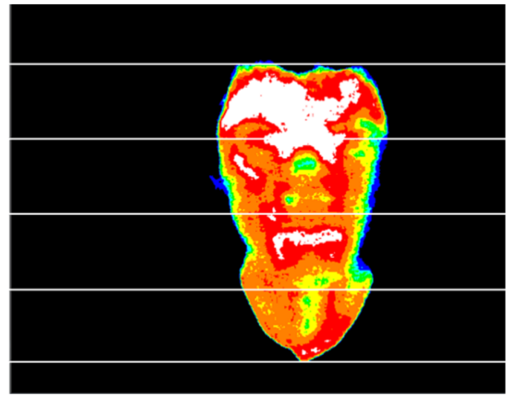

Fig. 3. Example of split image

B. Face/nose detection for multiple people

Since the amount of training data obtained in this study is small, we investigated a method that can detect even a small amount of data, rather than the CNN-based method, which is considered to require a large amount of training data. Face/nose candidates are detected from thermal images by the Viola-Jones method [6]. In addition, the nostrils candidate are detected to estimate the nose position. After that, each candidate is estimated in three steps. First, a minimum size of the face area is set because it is difficult to detect nose area if face size is too small. Next, for the brightness value of the face area, the ratio of the brightness value below a certain value is calculated. If the ratio is out of the allowable range, it is excluded from the candidates to prevent obvious false detection. Finally, the overlap that occurs around the actual face is removed. For the nose, contrast enhancement is first performed on the estimated facial area to deal with cases where the body temperature difference was small, and detection was difficult. After that, the candidates are narrowed down. First, the area where the nose is clearly not located is excluded from the detection target. Next, the candidates are narrowed down based on the mutual positional relationship between the nose and the nostrils. Finally, the pair with the shortest distance between the nose and the nostrils are found, and the position is estimated. Figure 4 shows the face and nose positions obtained by the proposed method. Red and yellow rectangular areas indicate face and nose position, respectively.

Tracking is performed by template matching using the estimated nasal region. At that time, the coordinates of the target face area are also held recorded, and simultaneous tracking for multiple people is possible by taking the correspondence between the template and the face area. Finally, the change in body temperature is acquired by the change in the brightness value in the tracking area.

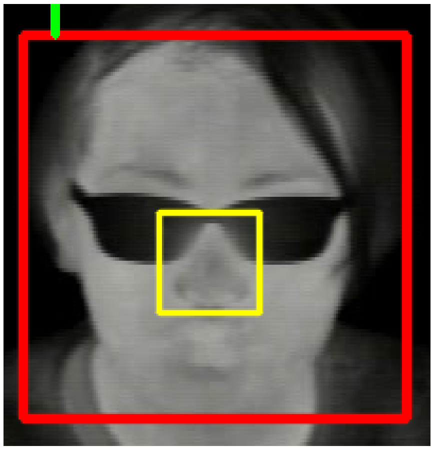

Fig. 4. Example of surrounding the estimated position with a rectangular area

III. Experiment

A. Estimation of drowsiness

First, the pulse rate is used to verify the proposed method for drowsiness estimation. Since the pulse rate decreases when humans become drowsy, the pulse rate during sleep and awakening of the subject is measured, and the pulse rate obtained in the experiment is compared with the pulse rate measured in advance. The measurement results are shown in table I. From this measurement result, drowsiness or sleep state was set at 60 times/min or less, and the awake state was set at 64 times/min or more. Note that 61 to 63 times/min was not used for estimation because it was indistinguishable.

TABLE I
Pulse rate measurement results

	First time	Second time	Third time
awake state	66 times/min	65 times/min	66 times/min
sleep state	57 times/min	58 times/min	58 times/min

Next, we measured the changes in skin temperature of the two data as a preliminary experiment. One was the data when only the internal change occurred. Fig. 5 shows a graph of the change in the difference in skin temperature between forehead and nose per 60 seconds when only internal changes occur. The red lines show that the change becomes large. In this preliminary experiment, it was estimated that the transition to drowsiness/sleep state occurred in the part where the change in the difference in skin temperature per 60 seconds was 1.1 or more, and the change in room temperature was less than 0.36. These values were used in the validation experiment.

In the verification experiment, the room temperature was rapidly raised by heating in a room with a low temperature. The difference in skin temperature until the subject went to sleep and woke up was measured, and the presence or absence of drowsiness of the subject was estimated. In the verification experiment, the acquired data is applied to any of the states using the values obtained in the preliminary

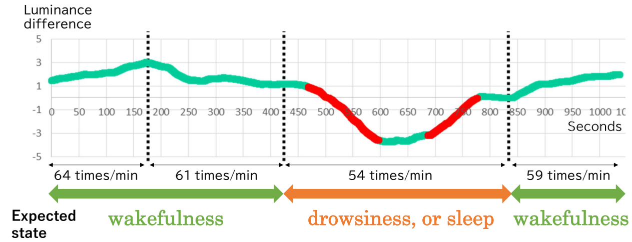

Fig. 5. The change in difference in skin temperature when only internal changes.

experiment. Expressly, after acquiring the difference data, it is confirmed whether the amount of change in the difference in skin temperature changes above the threshold value. If the amount of changes is not above the threshold value, the previous state is maintained. When the difference in skin temperature decreases, if the room temperature does not rise, it is in a drowsy state, and if it is rising, it is in a wakeful state. If the difference in skin temperature is rising, it is put into a sleeping state if the room temperature does not decrease and is in a wakeful state if it is decreased. The processing flow is shown in Fig. 6. Verification is performed by comparing the results with the state expected from the pulse rate, as shown in Fig. 5.

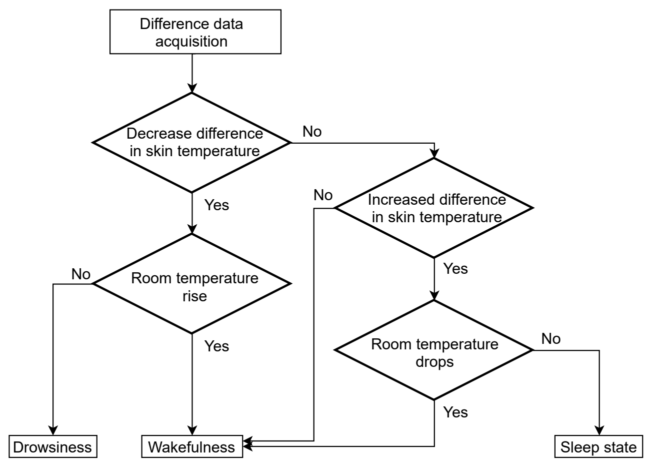

Fig. 6. Processing flow

B. Face/nose detection for multiple people

The experiment was conducted using videos of about 10 seconds each for six males and one female. As training data, 15 images per person were used. Fig. 7 shows an example of learning data for the face, nose, and nostrils. In this study, since the data was not enough for learning, the data for one subject was removed from the learning data in order. Therefore, the evaluation was performed using a discriminator learned by omitting the target subjects. In the experiment, the following points were verified:

- Is it possible to detect the face and nose?
- Is it possible to track the nose?
- Is it possible to detect and track stably?

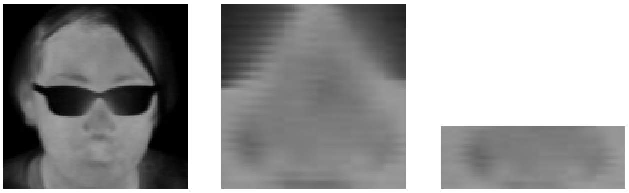

Fig. 7. Example of training data

IV. Results and Discussion

A. Estimation of drowsiness

Fig. 8 shows the results of the drowsiness estimation by the proposed method. The blue line in the graph shows room temperature, and the green, yellow and red color lines show the estimation results. The green line indicates arousal, the yellow line indicates drowsiness, and the red line indicates sleep state. This result shows that the estimation is almost correct. However, even though the sleep state ended once around 580 seconds, the transition to drowsiness and sleep state can be seen again. This is not the expected state. For this reason, the threshold value for detection of transition to drowsiness/sleep state was small. Therefore, we changed the threshold value from 1.1 to 1.4. The results after changing the threshold value are shown in Fig. 9. By changing the threshold, the difference from the expected state was resolved. However, since the judgment is made only by the pulse rate in this experiment, it cannot be said that the estimated state is highly reliable, and since the amount of change in the luminance difference is convex downward again, it is possible that the sleep state has continued. For more accurate estimation, it is necessary to improve the measurement method of the state.

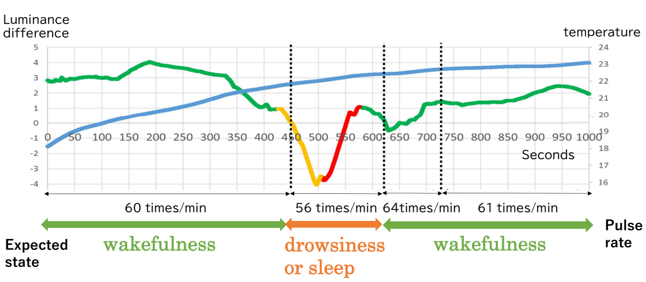

Fig. 8. Estimation result

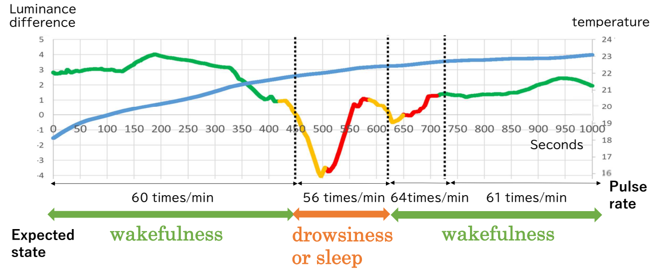

Fig. 9. Estimation results after adjusting threshold value.

B. Face/nose detection for multiple people

The characters that distinguish the subjects in the face / nose detection experiments are the same in all experiments. Table II shows the results in the case of one subject. From the left, this table shows the number of frames to be tested for each subject, the success rate of face detection, and the success rate of nose detection. Here, face detection accuracy represents the success rate of the number of frames to be tested, and the accuracy of nose detection is the success rate of the number of frames in which face detection was successful. The results were divided for each subject. In subjects C, D, and G, the recognition accuracy was high for both the face and nose. Especially in face recognition, high accuracy was obtained in the majority of subjects. However, some subjects had low recognition accuracy, and some became 0%. There are various reasons, one of which is that the detection failed due to correct exclusion due to size conditions. In particular, subject F had a small face size, so it was considered difficult to detect the nose, so it failed. Another reason is the bias of the training data, which is thought to have caused the accuracy of some subjects to decline. Figure 10 shows the nasal regions of subjects E. Looking at this figure, the nasal region of subject E is displayed as if it were divided into two regions. Such cases could not be confirmed by other subjects, and it is considered that the cause of the failure to detect the nose of subject E is such an irregular case. Similarly, regarding the face area, subject A had a longer face in the vertical direction than other subjects, which is considered to be the cause of the failure. If there are features that are found only in some subjects, as in these cases, it is difficult to detect. Therefore, we believe that the accuracy can be expected to improve if the bias of the learning data is reduced and a sufficient amount of data can be prepared.

TABLE II
Recognition rate when alone(%)

	Number of frames	Face detection	Nose detection
A	508	41.7	17
B	142	14.8	57.1
C	411	98.1	92.8
D	119	100	100
E	101	100	0
F	232	0	—
G	207	99.5	90.8

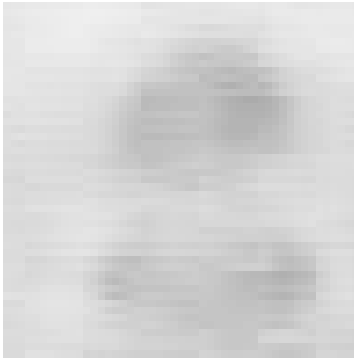

Fig. 10. Nose area in subject E

Table III shows the results for multiple people. From the left, this table shows the subjects included in each video, the number of frames to be tested, the subject data excluded from learning, the success rate of face detection, and the success rate of nose detection. In particular, the subject data excluded from the learning in the third column from the left in Table III is, for example, in the case of subjects A and B on the second line, the first line shows the recognition rate of multiple subjects at the same time by the discriminator learned without including the data of subject A in the training data. In the moving images of subjects A and B, when the data of subject B was excluded from the learning data, the recognition accuracy was relatively high, and in particular, the recognition accuracy of nose detection exceeded 90%. In addition, even in subjects A and G, high accuracy was confirmed in some face detections. However, another accuracy was low, and most of them were 0%.

TABLE III
SIMULTANEOUS RECOGNITION RATE FOR MULTIPLE PEOPLE(%)

	Number of frames	Removed subject data	Face	Nose
A,B	158	A	44.9	0
		B	82.9	91.6
A,F	129	A	0	—
		F	0	—
A,B,F	123	A	0	—
		B	0	—
		F	0	—
A,G	159	A	22	0
		G	83	2.3
A,G	23	A	0	—
		G	0	—

Table IV shows the recognition rate for each subject. It can be seen that the accuracy of face detection is not enough. The reason is that the data with one person is used as learning data. In the case of multiple people, the size of the face becomes smaller, and it is thought that it was excluded depending on the conditions. To solve this problem, the minimum size will be changed according to the case of one person and the case of multiple people. Another reason both face detection and nose detection were considered to be due to the bias of the learning data, as in the case of one person. The failure due to the change in the number of people was not confirmed.

TABLE IV
RECOGNITION RATE FOR EACH SUBJECT IN THE CASE OF MULTIPLE PEOPLE(%)

	Number of frames	Face detection	Nose detection
A	643	43.5	5
B	403	58.1	90.6
F	362	0	—
G	364	48.9	97.8

Next, the average number of failures per video was calculated for subjects with high recognition accuracy, and the tracking was evaluated. Table V shows the results in the case of one person. As a result, Subject D had only one video but no failure. However, subjects C and G had failures, especially in the nose detection. Therefore, in subjects C and G, a change in the luminance value in the tracking area was confirmed. Fig. 11 shows the results for subject C, and Fig. 12 shows the results for subject G. In the graph, the vertical axis is the luminance value, and the horizontal axis is the elapsed time. From these two graphs, it can be confirmed that in the case of subject C, relatively stable tracking is possible, although there are some outliers. In the case of multiple people, there was only one video with high recognition accuracy, so it was evaluated by one video. As a result, stable tracking was possible, as in the case of a single person. The proposed method was almost successful except for the bias of the training data and the exclusion due to the face size and was successful even when there was a certain degree of face direction, as shown in figure 13. Furthermore, since no failure due to changes in the number of people was confirmed, it was shown that the proposed method is effective whether it is one person or multiple people.

TABLE V
AVERAGE NUMBER OF FAILURES

	Face	Nose	Number of videos
C	4	16.5	2
D	0	0	1
G	0.5	10	2

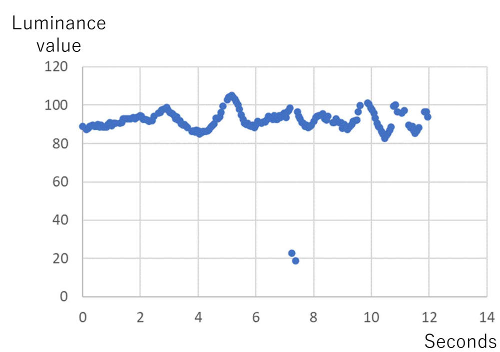

Fig. 11. Changes in the luminance value of the tracking area in subject C

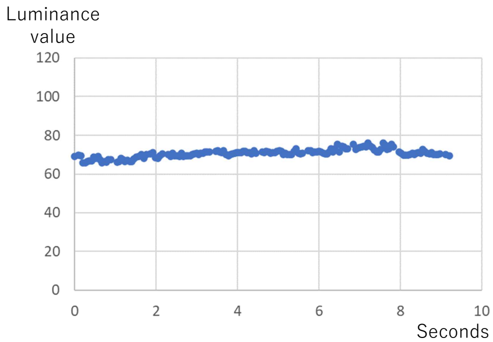

Fig. 12. Changes in the luminance value of the tracking area in subject G

Fig. 13. Detection result for not full-face

V. Conclusions

The purpose of this study was to estimate drowsiness in a non-contact method using only thermal images and to detect and track the faces and noses of multiple people in order to extend to estimate the drowsiness of multiple people. In the drowsiness estimation, we focused on the close relationship between human body temperature and activity. We estimated the drowsiness of the subject by measuring changes in body temperature and room temperature over time. The proposed method has made it possible to make a almost correct estimation. However, in order to further improve reliability, it is necessary to examine a more reliable comparison target. We focused on the estimation of drowsiness by measuring the skin temperature of the forehead and nose, but in the future, to estimate the subject's condition in more detail, it is necessary to consider the acquisition position and acquisition method of skin temperature data.

Regarding the detection of the faces and noses of multiple people, the Viola-Jones method was used, and the obtained results were used as candidates, and the candidates were narrowed down by three steps for each candidate. In addition, based on the information of the estimated position, we performed tracking corresponding to multiple people using template matching. As a result, the success rate of face detection exceeded 98% in the majority of subjects in the case of one subject, and the success rate of nose detection also exceeded 90% in subjects C, D, and G. Regarding tracking, it was confirmed that stable tracking was possible from the average number of failures performed on subjects with a high success rate of nose detection and the measurement results of changes in the luminance value of the tracking area. On the other hand, we also found a failure that seems to require a reconsideration of learning data and the minimum face size in the case of multiple people. However, since no failure was found that was thought to be directly caused by the change in the number of people, it was shown that the proposed method is effective not only for one person but also for multiple people. In future work, it is necessary to increase the number of subjects and require learning data with various facial features to improve learning data bias. In addition, to estimate drowsiness for multiple people, it is necessary to consider a method for estimating the position of the forehead and a method for obtaining changes in body temperature in consideration of room temperature.

In the future, for applications such as alerting when drowsiness is detected, it will be necessary to perform from face detection to drowsiness estimation in real-time. Furthermore, by performing personal authentication simultaneously, it can be applied to the collective management of individual physical conditions.

References

[1] M. Uchiyama, R. Furihata, "Regulation of Body Temperature and Sleep in Humans", The Journal of Balneology, Climatology and Physical Medicine, Vol.78, No.1, pp.6-9(2014)(in Japanese)

[2] Y. Nakano, A. Miyagawa, S.Sano,"Detection of Driver Drowsiness Level", Fujitsu., Vol.59, No.4, pp.416-420(2008)(in Japanese)

[3] M.Shoji, A.Sakata, "Understanding life behavior by changes in temperature distribution of face and palm", Graduate School of Engineering The University of Tokyo Master's Thesis(2002).

[4] X. Lin, Z. Jiang, J. Fromm, X. Xu, S. Patel, D. McDuff, "MetaPhys: Few-Shot Adaptation for Non-Contact Physiological Measurement", ACM Conference on Health, Inference, and Learning(ACM CHIL)(2021).

[5] Jin-sup Eom, Mi-Sook Park, Hye-Ryeon Yang, Young-Ji Eum and Jin-Hun Sohn,"Nose Detection and Breathing Monitoring in Thermal Images", International Journal of Advanced Science and Technology Vol.109, pp.67-76(2017).

[6] P.Viola, and M.Jones,"Robust real-time face detection", International journal of computer vision, vol.57, no.2, pp.137-154(2004).

[7] F. Q. AL-Khalidi, R. Saatchi, D. Burke, and H. Elphick, "Tracking human face features in thermal images for respiration monitoring", IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)(2010).