

Monitoring respiratory state from near-infrared face video images

Koichi Ashida^{1,*}, Yuta Hino^{1,**}, Chawan Koopipat², Keiko Ogawa-Ochiai³ and Norimichi Tsumura^{1,3}

¹Graduate School of Science and Engineering, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan

²Department of Imaging and Printing Technology, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok, 10330 Thailand

³Hiroshima University, Hiroshima University Hospital, 1-2-3, Kasumi, Minami-ku, Hiroshima-shi, Hiroshima, 734-0037 Japan

* The first and second authors are equally contributed to this paper.

**Email: hino0329y@chiba-u.jp

Abstract: In this paper, we propose an algorithm for estimating respiratory state using near-infrared facial video images. Estimation of respiratory state is an important indicator for early detection of respiratory diseases. In particular, there is a demand for monitoring respiratory state during the night. One method of monitoring respiratory state is to use contact-type sensors. However, this method requires the installation of many sensors and a visit to a hospital, which place a burden on patients. Therefore, we propose to acquire respiratory-induced features from near-infrared face video images and investigate their similarity to measurements obtained with a respirometer for non-contact monitoring of respiratory state in the dark. Respiratory-induced features were obtained from pulse wave signals extracted from the face video images. The results showed correlations in several respiratory states. This study opens some perspectives in non-contact monitoring of respiratory states.

Keywords: Near-infrared, respiratory state, image processing, pulse wave estimation

1 INTRODUCTION

Respiration measurement is an important indicator for the early detection of respiratory diseases. Sleep apnea is one of the respiratory diseases that occur during sleep. Sleep apnea is a disease in which breathing stops or becomes shallow many times during sleep, causing hypoxia in the body [1]. This increases the occurrence of ischemic heart diseases such as hypertension, stroke, and myocardial infarction due to stress caused by hypoxia during sleep and sleepiness during the day. Apnea is defined as a condition in which breathing stops for 10 seconds or more, and oxygen saturation drops by 3%. SAS (sleep apnea syndrome) is diagnosed when apnea occurs five or more times per hour during sleep. The number of potential SAS patients in Japan is estimated to be around 5 million [2]. SAS patients are prone to complications such as lifestyle-related diseases and are unable to achieve deep sleep. This makes it difficult to prevent or recover from the disease, which may lead to death by myocardial infarction or heart attack if SAS patients are left untreated [2].

To detect these symptoms at an early state, it is important to monitor respiratory state during sleep. Polysomnography is a common technique for measuring respiratory state during sleep. It uses multiple sensors to comprehensively determine the electroencephalogram, blood oxygenation, electrocardiogram, chest and abdominal movements, and airflow in the trachea during sleep. There are ten types of sensors used in the examination [3]. Three of them are

sensors that measure respiration. However, polysomnography requires a contact-type sensor, which imposes a heavy burden on users. In addition, it requires a visit to a hospital because of the need to inspect precision equipment.

Cho et al. proposed a method for detecting respiration by detecting the nostril region from thermal images and detecting the temperature change of the nostrils due to respiration [4]. However, this method requires a special camera to capture thermal images, which is very expensive. Walter et al. proposed a method for estimating respiratory rate using a photoplethysmography obtained from a contact-type sensor [5]. However, this method uses a contact-type device, which may be uncomfortable to subjects when worn for long periods of time during sleep. In addition, trauma or burns to the fingers make it nearly impossible to use a contact-type sensor [6]. Furthermore, respiratory state, such as respiratory depth, cannot be detected by respiratory rate alone. Kurita et al. proposed a method for estimating pulse wave signal using RGB face video images as a non-contact method for estimating pulse wave signal [7]. However, this method is difficult to use at night during sleep because it uses an RGB camera.

In this paper, we investigate whether respiratory state is reflected in non-contact acquisition of respiratory-induced features for non-contact monitoring of respiratory state in the dark. Specifically, we acquired respiratory-induced features from near-infrared face video images and obtained correlation coefficients with respiratory measurements. The

results showed there was a correlation between normal, deep and shallow breath states. In addition, machine learning utilizing features of respiratory-induced variation was used to estimate apnea state. The results of this study allow a knowledge deepening in the field of non-contact estimation of respiratory state in the dark.

2 METHODS

2.1 Conventional method

A method to estimate respiratory rate by extracting three respiratory-induced variations from the PPG (photoplethysmography) signal has been reported [5]. Respiratory-induced variations are strongly correlated with the PPG signal. Figure 1 shows the characteristics of respiratory-induced variation in the PPG signal. RIAV (respiratory induced amplitude variation) is a characteristic quantity caused by an increase in cardiac output during inspiration due to changes in intrathoracic pressure, and by an increase or decrease in the volume of blood flow. This can be obtained by calculating the difference between the upper and lower peak points. RIIV (respiratory induced intensity variation) is a characteristic quantity caused by an increase in the inflow of blood to the heart due to the dilation of the veins during inspiration caused by a decrease in intrathoracic pressure. This can be obtained by calculating the trend of the line connecting the adjacent peak points. RIFV (respiratory induced frequency variation) is a characteristic value that the heart rate increases during inspiration and decreases during expiration. This can be obtained by calculating peak interval of the pulse wave signal.

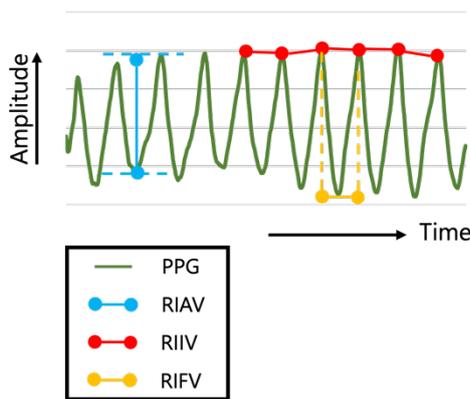


Fig. 1. Respiratory-induced features

2.2 Proposed method

Figure 2 shows the procedure for extracting the pulse wave signal from near-infrared face video images. A near-infrared camera was used because they use wavelengths that are invisible to human eyes. The pulse wave signal can be estimated by setting the ROI (region of interest) from the

face video images and outputting a time series of average pixel values. The ROI was set in the area including the subject's nose and cheeks [8, 9]. By setting the ROI on this part of the face, the pulse wave signal can be effectively acquired. The ROI was set manually. The obtained pulse wave signal was detrended [10]. The detrended pulse wave signal was then bandpass filtered. The passband for bandpass filtering was set to [0.75, 4.0] Hz [11, 12, 13].

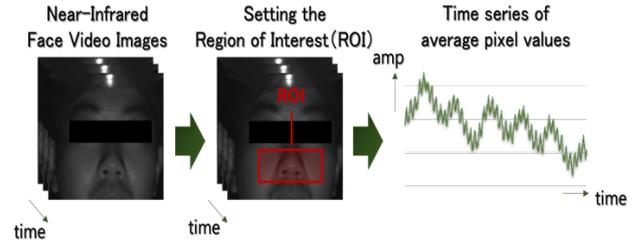


Fig. 2. Procedure for extracting pulse wave signal from near-infrared face video images

After bandpass filtering, respiratory-induced features were obtained. To obtain features, the peak points were detected from the pulse wave signal after bandpass filtering [9]. There are a total of seven acquired features obtained: RIAV, RIIV (obtained from the upper and lower peak points respectively), RIFV (obtained from the upper and lower peak points respectively) and DOPP (difference of peak points) (obtained from the upper and lower peak points respectively). The two features in DOPP are newly introduced in our proposed method. Figure 3 shows these seven features.

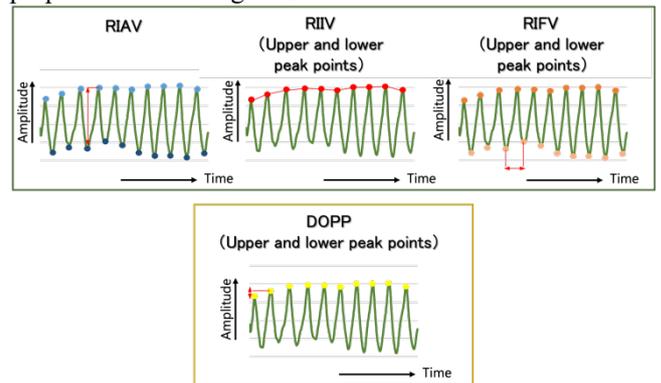


Fig. 3. Seven features used in the proposed method

3 EXPERIMENTAL PROCEDURES

Figure 4 shows the experimental environment. The subjects were three males in their 20s. In this experiment, the camera was placed in a dark room. In the experimental condition, face video images were acquired while the subjects were sitting. The subjects were instructed to remain still during the filming. The subjects' faces were fixed using a chin rest, which minimized the subjects' head motion as

much as possible. A monochrome camera (DMK33UX174, The imaging source) with a near-infrared filter (Edmund Optics) that passes light at wavelengths range of 750-850nm was used as a near-infrared camera. The resolution of this camera's sensor was 1920×1200 pixels. Video images of 1280×960 pixels in the center of 1920×1200 pixels was captured by configuring the settings of capturing software. The exposure time of the camera was $1/60$ seconds. This is the reciprocal of the frame rate. Near-infrared LED was used as the light source in the near-infrared region. The peak wavelength of the near-infrared LED was 840 nm. The ground truth of respiration state was measured by a respirometer attached to the subject's chest.

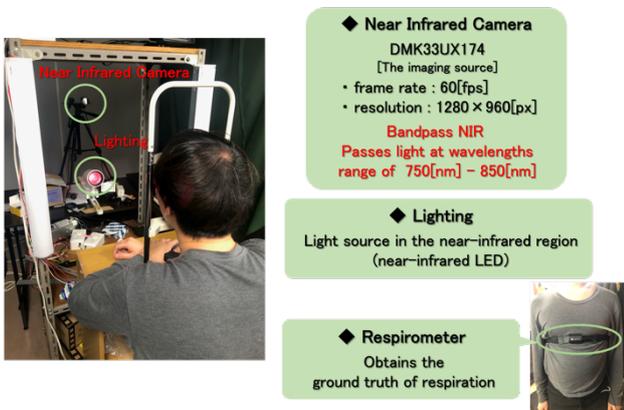


Fig. 4. Experimental environment

In these experimental conditions, the following procedure was used.

Condition1: normal breathing (60 seconds)

Condition2: normal breathing (20 seconds) → apnea (10 seconds) → normal breathing (30 seconds)

Condition3: deep breathing (30 seconds) → shallow breathing (30 seconds).

Under these conditions, the subjects were instructed to change their respiratory state during the imaging. Each subject's respiratory rate for 60 seconds was 12, 13 and 11 breaths respectively in condition 1. In condition 3, subjects were instructed to breathe once every 6 seconds for deep breathing and once every 3 seconds for shallow breathing. Figures 5, 6 and 7 shows the ground truths for one subject under these conditions.

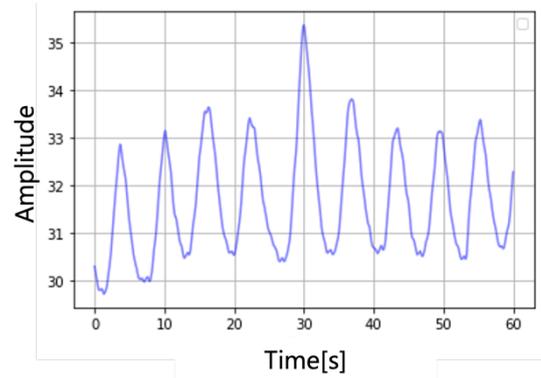


Fig. 5. Ground truth (Condition 1)

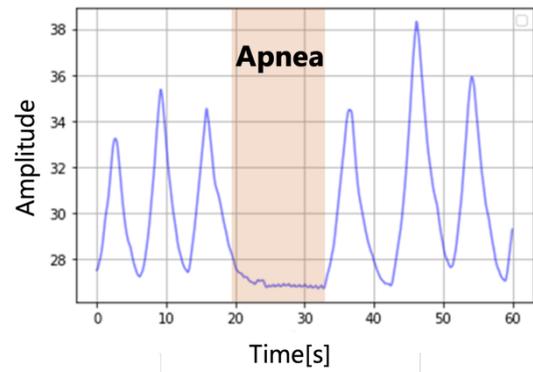


Fig. 6. Ground truth (Condition 2)

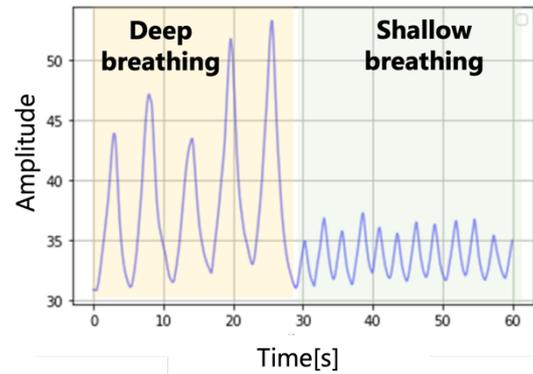


Fig. 7. Ground truth (Condition 3)

4 RESULTS OF EXPERIMENT

4.1 Estimation of pulse wave signals

In the experiment, the correlation coefficients between the pulse wave signals estimated from the face video images and those measured by the pulse wave meter were determined. The mean value and standard deviation of the correlation coefficients for a total of nine pulse wave signals were 0.72 ± 0.06 . This result indicates that the estimated pulse wave signal has a high correlation with the pulse wave signal measured by the pulse wave meter. The peak points

are detected from the estimated pulse wave signal after detrending and bandpass filtering. Figure 8 shows the pulse wave signal after detrending and bandpass filtering and the pulse wave signal measured with a pulse wave meter. This figure shows that the position and the variability of the peak points coincide in the two pulse wave signals. This result indicates that the accuracy of pulse wave estimation in this study does not affect the acquisition of respiratory-induced features.

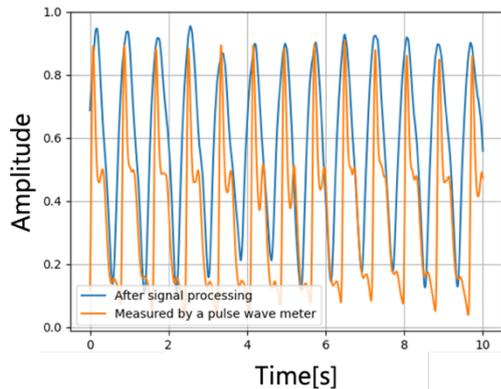


Fig. 8. The pulse wave signal after signal processing and the pulse wave signal measured with a pulse wave meter

4.2 Estimation of respiratory-induced features

Figures 9, 10 and 11 show the results of feature estimation in condition 1, 2 and 3, respectively, for a subject. The points (observed) in the figures indicate the values of the acquired features. The solid lines (fitted) in the figures are straight lines connecting every two points.

Table 1 shows the correlation coefficients between each estimated respiratory-induced feature and the ground truth. The mean and standard deviation of the correlation coefficient were calculated for the three subjects. In this study, the correlation coefficient was considered to be correlated with the ground truth if the mean value of the correlation coefficient was greater than 0.50. RIAV and RIIV were correlated with the ground truth in condition 1. RIIV was correlated with the ground truth in condition 3. On the other hand, no features correlated with the ground truth in condition 2. This result indicates that temporal variations in respiration can be detected only in conditions 1 and 3.

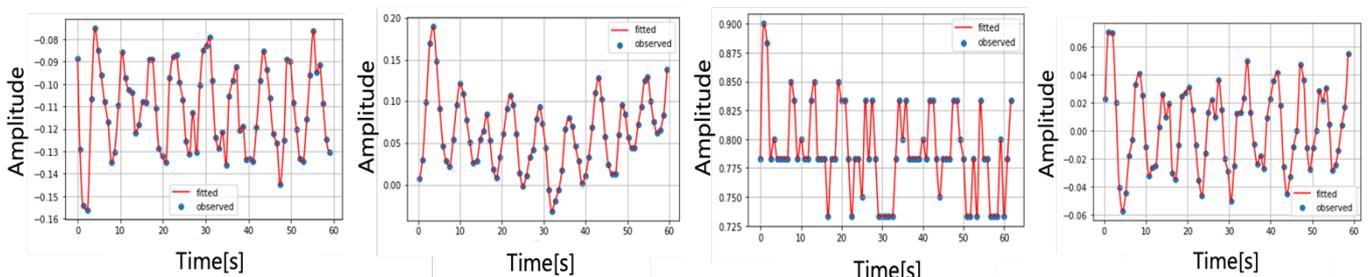


Fig. 9 Results of feature estimation in condition 1. **(a)** RIAV; **(b)** RIIV (obtained from the upper peak points); **(c)** RIFV (obtained from the upper peak points); **(d)** DOPP (obtained from the upper peak points).

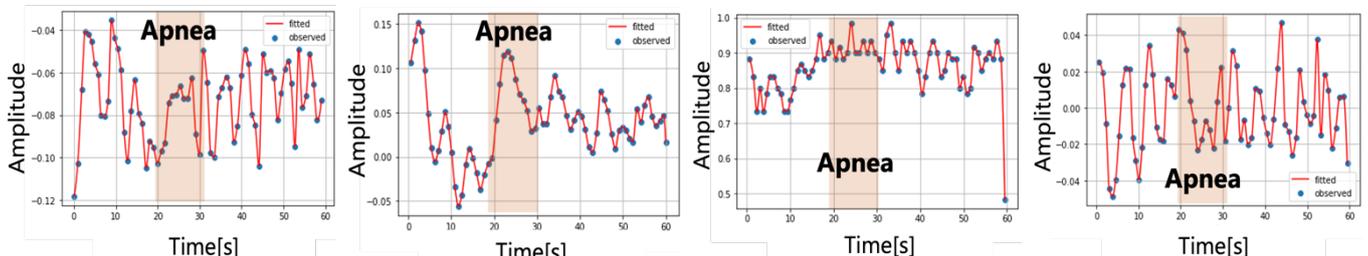


Fig. 10 Results of feature estimation in condition 2. **(a)** RIAV; **(b)** RIIV (obtained from the upper peak points); **(c)** RIFV (obtained from the upper peak points); **(d)** DOPP (obtained from the upper peak points).

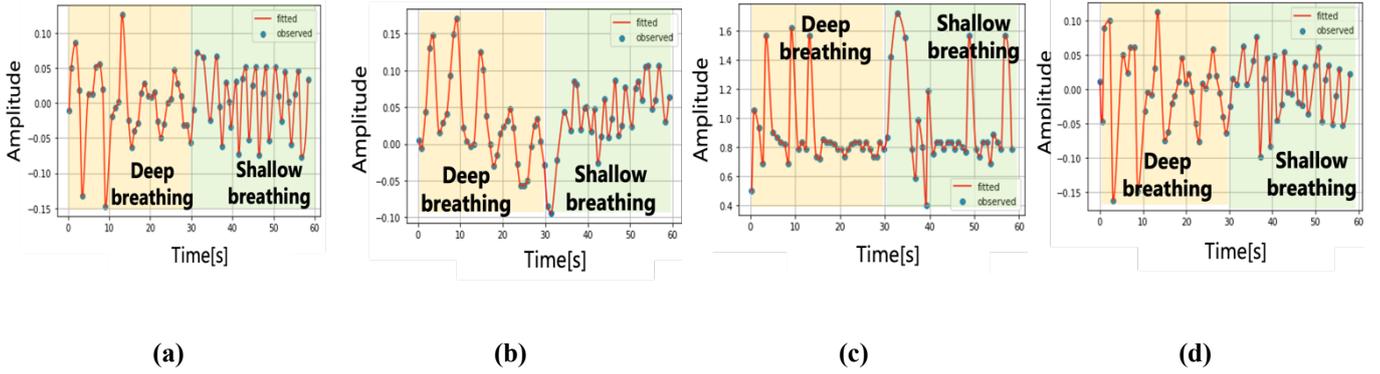


Fig. 11 Results of feature estimation in condition 3. **(a)** RIAV; **(b)** RIIV (obtained from the upper peak points); **(c)** RIFV (obtained from the upper peak points); **(d)** DOPP (obtained from the upper peak points).

Table 1. The correlation coefficients between estimated respiratory-induced features and the ground truth.

Condition	Features	Correlation coefficient
Condition 1	RIAV	0.54 ± 0.15
	RIIV	0.65 ± 0.12
	RIFV	-0.39 ± 0.14
	DOPP	-0.48 ± 0.12
Condition 2	RIAV	0.24 ± 0.13
	RIIV	0.18 ± 0.15
	RIFV	-0.26 ± 0.09
	DOPP	-0.19 ± 0.15
Condition 3	RIAV	0.32 ± 0.14
	RIIV	0.52 ± 0.27
	RIFV	-0.14 ± 0.11
	DOPP	-0.28 ± 0.03

4.3 Estimation of apnea state

To improve the accuracy of apnea state estimation, we attempted to estimate the ground truth obtained from the respirometer using SVR (support vector regression). This was attempted because the apnea state is clearly indicated in the ground truth as shown in Figure 6. The input values used for SVR are the seven features described in section 2.2. The kernels and parameters of the SVR were set to various conditions, and the kernels and parameters of the condition that showed the highest performance were used. In this study, a linear kernel was used. In addition, the regularization parameter C was set to 1.0 and the hyperparameter epsilon to 2.0.

3-fold cross-validation was performed using nine data sets from three subjects. Data from two subjects were used for training, and data from the remaining subject was used for testing. The correlation coefficient between the value estimated by SVR and the ground truth was 0.46 ± 0.07 . Comparing this result with the value of the correlation coefficient for condition 2 in Table 1, the use of SVR resulted

in a higher correlation with the ground truth. However, compared to the results for conditions 1 and 3 in Table 1, the correlation with the ground truth was still low.

Figure 12 shows the ground truth for condition 2 and the estimated values output by the SVR for a subject. The values fluctuate less in the apnea state than in the normal breathing state. This indicates the possibility of apnea state detection.

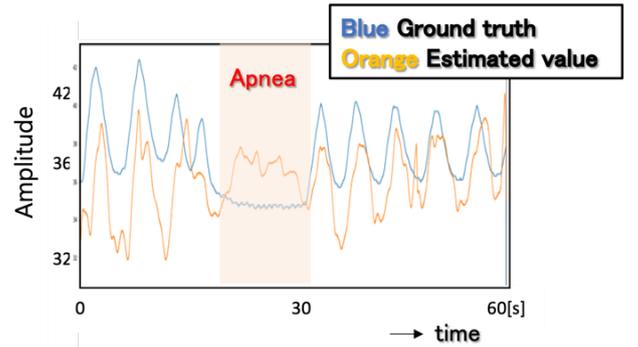


Figure 12. Estimated value from SVR and the ground truth in condition 2.

5 DISCUSSION

From the experimental results, normal breathing, deep breathing and shallow breathing states have correlation with respiratory measurements. However, the apnea state showed a lower correlation than the other respiratory states. This is due to the occurrence of a symptom called compensatory mechanism. Compensatory mechanism is a function to maintain blood flow to the systemic circulation, especially to vital organs, even when cardiac function declines [14]. In the apnea state shown in Figure 13, the blood volume necessary for life is maintained at first, but after reaching the peak, the blood volume gradually decreases. Therefore, the correlation was low in apnea state.

We evaluate polysomnography and the method we implemented in terms of usability, reliability and costs. In terms of usability, the proposed method does not require the

user to wear sensors or equipment. Therefore, the proposed method can detect respiratory state with minimal burden on the user. In terms of reliability, the proposed method uses only a camera as a sensor, whereas polysomnography comprehensively determines respiratory state based on data obtained from various sensors, resulting in lower accuracy. In terms of costs, the proposed method can reduce costs because cameras are less expensive than the equipment used in polysomnography. Potential applications of the proposed method include home healthcare and nursing care facilities. Although the accuracy of respiratory state estimation is lower than that of polysomnography in these situations, the cost of the proposed method can be significantly reduced as described above. Further improvement of the accuracy of respiratory state estimation is needed to achieve the same accuracy as polysomnography and to enable its application in medical settings.

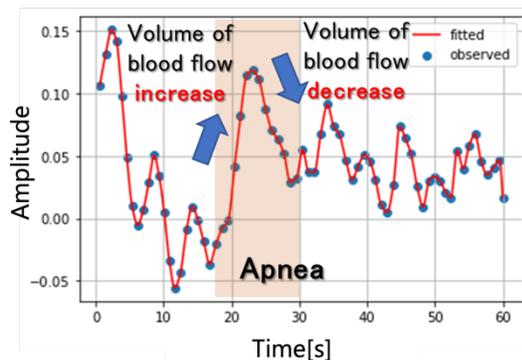


Fig. 13. RIV components in apnea state

6 CONCLUSION AND FUTURE WORKS

In this paper, we investigate whether respiratory state is reflected in non-contact acquisition of respiratory-induced features for non-contact monitoring of respiratory state in the dark. Specifically, we acquired respiratory-induced features from near-infrared face video images and obtained correlation coefficients with respiratory measurements. Experimental results showed there were correlations for normal, deep and shallow breath state in some of the feature values. These indicate that the respiratory state is reflected in some feature values. However, the apnea state could not be detected. Therefore, we attempted to detect the apnea state by learning features related to respiratory-induced variation using machine learning. The results showed that the correlation between estimated and measured respiration values was higher than for the features alone.

One of our future works is to make it possible to determine the respiratory state using only feature values. The results of this study indicate that the respiratory state is

reflected in some feature values. This suggests the possibility of non-contact estimation of respiratory state. It is necessary to determine the respiratory state using the respiratory-induced features by analyzing their amplitudes and frequencies and by performing further experiments.

In our experiment, the face was fixed using a chin rest. Therefore, the artifact caused by facial movement was almost negligible, and the pulse wave signal could be estimated with high accuracy. However, in the actual application of this method, it is expected that the face will move. Therefore, it is necessary to accommodate facial motion when estimating the pulse wave signal. In our method, the ROI used to estimate the pulse wave signal was fixed. To accommodate facial motion, it is necessary to detect the face and automatically set the ROI according to the facial motion.

REFERENCES

- [1] Kinki Central Respiratory Center, Sleep apnea syndrome. https://kcmc.hosp.go.jp/cnt0_000236.html. (Accessed 17 January 2023)
- [2] Zone, About Sleep Apnea Syndrome (SAS), <https://www.sleepzone.co.jp/clinic/sas/index.php>. (Accessed 17 January 2023)
- [3] Rakuwakai Health Care System, Sleep apnea syndrome Polysomnography (PSG) Test. <http://www.rakuwa.or.jp/otowa/shinryoka/sas.html>. (Accessed 17 January 2023)
- [4] Cho Y, Julier S, Marquardt N, et al. (2017), Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging. *Biomedical Optics Express*, Vol. 8, Issue 10, pp. 4480-4503
- [5] Walter K, Srinivas RJ Mark A, et al. (2013), Multiparameter respiratory rate estimation from the photoplethysmogram, *IEEE Transactions on Biomedical Engineering*, Volume 60, Issue 7, pp.1946-1953
- [6] Gambi E, Ricciuti M, and Spinsante S (2018), Sensitivity of the contactless videoplethysmography-based heart rate detection to different measurement conditions. 26th European Signal Processing Conference (EUSIPCO), IEEE, pp.767-771
- [7] Kurita K, Yonezawa T, Kuroshima M, et al. (2015), Non-Contact Video Based Estimation for Heart Rate Variability Spectrogram using Ambient Light by Extracting Hemoglobin Information. *Color and Imaging Conference*, Volume 2015, pp.207-211
- [8] Mayank K. Ashok V and Ashutosh S (2015), Distance PPG: Robust non-contact vital signs monitoring using a camera, *Biomedical Optics Express*, Volume 6, Issue 5, pp.1565-1588
- [9] Rong M and Li K (2021), A Blood Pressure Prediction Method Based on Imaging Photoplethysmography in combination with Machine Learning, *Biomedical Signal Processing and Control*, vol.64, pp.102328
- [10] Tarvainen MP, Ranta-

Aho PO, Karjalainen PA (2002), An advanced detrending method with application to HRV analysis. IEEE Transactions on Biomedical Engineering, vol.49, no.2, pp.172-175

[11] Wang W, den Brink AC and de Haan G (2019), Discriminative Signatures for Remote-PPG. IEEE Transactions on Biomedical Engineering, vol.67, no.5, pp.1462-1473

[12] Poh MZ, McDuff DJ and Picard RW (2010), Non-contact automated cardiac pulse measurements using video imaging and blind source separation. Optics express, vol.18, no.10, pp.10762-10774

[13] Poh MZ, McDuff DJ and Picard RW (2010), Advancements in noncontact, multiparameter physiological measurements using a webcam, IEEE Transactions on Biomedical Engineering, vol.58, no.1, pp.7-11

[14] Levawell Nursing, Compensatory Mechanisms. <https://kango-oshigoto.jp/hatenurse/article/1958/>.

(Accessed 17 January 2023)