Proc. 2019 IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN 2019), New Delhi, India, Oct. 2019.

Fatigue Estimation using Facial Expression features and Remote-PPG Signal

Masaki Hasegawa[1], Kotaro Hayashi[2] and Jun Miura[3]

***Abstract*— Currently, research and development of lifestyle support robots in daily life is being actively conducted. Health-case is one such function robots. In this research, we develop a fatigue estimation system using a camera that can easily be mounted on robots. Measurements taken in a real environment have to be consider noises caused by changes in light and the subject's movement. This fatigue estimation system is based on a robust feature extraction method. As an indicator of fatigue, LF/HF-ratio was calculated from the power spectrum of RR interval in the electrocardiogram or the blood volume pulse (BVP). The BVP can be detected from the fingertip by using the photoplethysmography (PPG). In this study, we used a contactless PPG: remote-PPG (rPPG) detected by the luminance change of the face image. Some studies show facial expression features extracted from facial video are also useful for fatigue estimation. dimension reduction of past method using LLE spoiled the information in the large dimention of feature. We also developed a fatigue estimation method with such features using a camera for the healthcare robots. It used facial landmark points, line-of-sight vector, and size of the ellipse fitted with eyes and mouth landmark points. Therefore, proposed method simply use time-varying shape information of face like size of eyes, or gaze direction. We verified the performance of proposed features by the fatigue state classification using Support Vector Machine (SVM).**

I. INTRODUCTION

The worldwide declining birthrate and society's aging have become increasingly serious problems. Moreover, there are various problems in each country such as the self inflicted problem in the USA or the lifestyle diseases in Japan. The healthcare robot is one of the prevailing solutions to such problems. Robots can perform both physical and cognitive tasks. Among them, we focus on fatigue estimation. A fatigue state is important to get signs of crucial diseases or accidents. If a healthcare robot can estimate the fatigue state, it can advise a user to go to a hospital or report to an administrator in the workplace. It is important that a healthcare robot can say "Aren't you tired?" or warn of signs of danger by fatigue detection. Facial expressions (FE) may reveal causes of the fatigue [1]. Recent studies have proposed detecting driver fatigue using a camera [2], [3], [4], because detection by a contract sensor has a high accuracy but is unsanitary and troublesome to use. Most robots have cameras for self-localization, robot mapping,

[1]Masaki Hasegawa is with the Toyohashi University of Technology, Aichi 441-8580 Japan (phone: +08-0532-446826; e-mail: hasegawa@aisl.cs.tut.ac.jp).

[2]K. Hayashi is with the Toyohashi University of Technology, Aichi 441-8580 Japan (phone: +08-0532-446826; e-mail: hayashik@cs.tut.ac.jp).

[3]J. Miura is with the Toyohashi University of Technology, Aichi 441-8580 Japan (phone: +08-0532-446773; e-mail: jun.miura@tut.jp).

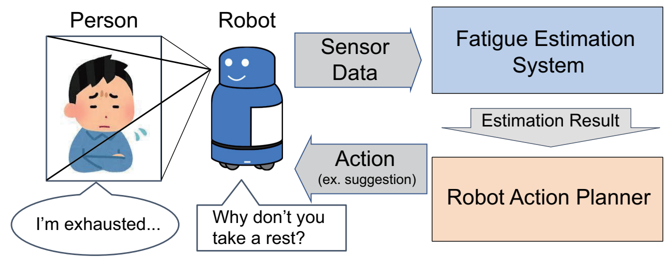


Fig. 1: Image of a robot system for fatigue estimation

object recognition and obstacle avoidance. Therefore, we propose a fatigue estimation method that measures heart rate variability from expression features measured by a camera.

Heart rate variability, or the time fluctuation of electrocardiogram RR variability, is well known as a fatigue estimation indicator. To measure it without using the cardiogram, the reflection of light from the skin can be used to calculate the BVP [5]. There are two ways of contactless measurement with BVP. One is using the waveform received from the Doppler radar, and the other is the light reflected from the skin. Current fatigue estimation studies show availability of the BVP-based method [6]. However, one should avoid having a tired user estimate fatigue. Therefore some fatigue estimation studies focus on the use of facial information obtained by a camera. Ji et al. used FE features to estimate fatigue by using an IR LED and a CCD camera [7]. The FE feature-based fatigue estimation methods must be compact which can be implemented into personal service robots. In this research, we use a simple webcam to get facial information because these robots can not use stereo camera due to the size. Additionally, the FE feature-based fatigue estimation method for healthcare robots need to deal with the various lighting environments and the target movement. The FE feature-based estimation must avoid failing to detect or incorrectly detect these changes. In this study, we aim to develop a robust method overcome these noises.

II. RELATED WORKS

A. Contactless BVP measurement

BVP is a method of analyzing heart rate variability (HRV), because its peak appears in time with the beat. Generically, autonomic nervous activity can be able to estimete by HRV, so that HRV is used as an index of fatigue Photoplethysmography (PPG) is a popular method to measure BVP [5], measuring reflected light and absolution quantity fluctuation as BVP. However, this method can only use optical sensors and a light source such as an LED light. In its resting state, R-R interval acquired from the PPG has a very high correlation relationship with that optained by electrocardiogram [8]. Verkruysse et al. [9] showed a PPG-based method using an RGB camera called the remote-PPG (rPPG). Poh et al. [10] showed a method of extracting BVP from color channels in video recordings by using independent component analysis. This method can extract BVP with a high correlational relationship (0.88) to the one obtained by PPG. Haan et al. [11] showed a normalized color signal model to eliminate the noise of changes in the skin's specular reflection caused by changes of lighting conditions and head movements. They could estimate the heart rate by a correlation of 0.99 from 10000 frame video. Using this model, they extracted BVP by weighted differences of color channels under changes in the head position with high degree of accuracy. Huang et al. [12] improved the accuracy of this model using continuous wavelet transform (CWT) denoising. Mcduff et al. [6] extended the rPPG-based fatigue estimation method using a 5-band camera. We chose rPPG from an RGB camera because it was easy to mount on the robot or was already mounted.

B. Facial expression features-based fatigue estimation

This fatigue estimation method was used on the driver incases [7], [13], [14]. It estimates fatigue by watching the face and recognizing the signs of fatigue such as increased blinking, head movement, and yawns. Ji et al. [7] showed a method to predict fatigue by recognizing the gaze direction, blinking and the head movement. Kawamura et al. [15] proposed a fatigue estimation method from face images for 30 seconds, analyzing feature values and SVM. A feature was obtained from the integration of HOG feature values as the FE and color changes of the cheek region in RGB channel as the BVP feature. Locally Linear Embedding (LLE) [16] was used for the dimension reduction of them. This result showed the effectiveness of the FE image method but it also that it was insufficient to measure LF/HF, and lighting noise will negatively effect the estimation. Past studies focus on the situation under long-term surveillance and can use rich times. We aimed to develop a robust estimation that could account for changes in the lighting and head movements by short-time glancing.

III. METHOD

We propose a fatigue estimation method with a more robust feature extraction to account for the fluctuation of the light. We used an image for 60 seconds as the input for the BVP feature and one for 30 seconds as the input for the FE feature. These features are used for the fatigue estimation by the support vector machine (SVM) binary classification.

A. Extract facial expression feature

We used the OpenFace toolkit [17] to extract facial expression features because this toolkit could run in real time, and was able to compress extraction time and become robust to the change of background. OpenFace can extract four

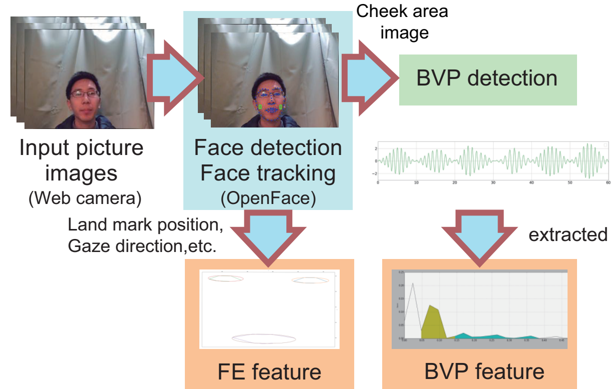


Fig. 2: Extract two features

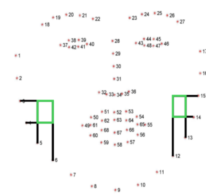


Fig. 3: Cheek region from OpenFace

types of features: 136 two dimensional landmarks, 6 three dimensional line-of-sight vectors and 18 Action Units (AUs) expression levels and AUs intensities. The two dimensional landmarks were given by $sec \times fps \times 136$ dimensional feature vector ($\boldsymbol{V}_l$) from the eye and mouth areas of all picture frames. AUs expression levels and 17 AUs intensities were given by $sec \times fps \times 18$ dimensional feature vector ($\boldsymbol{V}_P$) and $sec \times fps \times 17$ dimensional feature vector ($\boldsymbol{V}_I$), respectively. We used a ellipse fitting by least squares [18] to reduce the dimension of landmark points of the eyes and the mouth shown in Fig. 2. Each 3D feature representing the degrees of eyes or mouth aperture was determined by the ratios of each ellipse (b/a).

B. Extract BVP

We extracted BVP from facial image segments by the rPPG framework with CWT [12], because it is prevent able to noise from the shading of the entire face and there is past performance. Fig. 3 shows the cheek regions for this extraction defined by OpenFace feature points.

The color signal $\boldsymbol{S} = \{\boldsymbol{c}_0, \ldots, \boldsymbol{c}_N\}$ was given by the average pixel value $\boldsymbol{c} = \{\bar{r}, \bar{g}, \bar{b}\}$ of the two cheek regions. In the implementation, we performed eight times fourier interpolation to $\boldsymbol{S}$ and linear interpolation to the time of the data. In [12], the color signal model was given by the color signal $\boldsymbol{S}_i$ of the ith frame:

$$C_i = I_{C_i}(\rho_{C_{dc}} + \rho_{C_i} + s_i), \tag{1}$$

where I_{C_i} was the strength of light, $\rho_{C_{dc}}$ was the stationary part of the reflection coefficient of the skin, ρ_{C_i} was the zero-mean time-varying fraction caused by the pulsation of the BVP and s_i is the additive specular reflection contribution.

$\rho_{C_{dc}}$ could be removed by the normalization of every color channel. We used a moving average for this normalization:

$$C_{ni} = \frac{C_i}{\mu(C_i)}, \tag{2}$$

where $\mu(C_i)$ was a moving average centered around frame index i. We set a window size of forty eight frames (1.6s) so that a window included at least one beat. The color signals $\boldsymbol{R}_n$, $\boldsymbol{G}_n$, $\boldsymbol{B}_n$ were given as elements of the color channels $\boldsymbol{S}_n$. Two differences of color signals $\boldsymbol{X}_s$ and $\boldsymbol{Y}_s$ were given :

$$\begin{cases} \boldsymbol{X}_s = 3\boldsymbol{R}_n - 2\boldsymbol{G}_n, \\ \boldsymbol{Y}_s = 1.5\boldsymbol{R}_n + \boldsymbol{G}_n - 1.5\boldsymbol{B}_n, \end{cases} \tag{3}$$

where coefficients were skin color normalization factors that were defined in [11] to normalize different colors in the various lighting conditions. Then, signals $\boldsymbol{X}_f$, $\boldsymbol{Y}_f$ were given by applying a band-pass filter with a $0.67 \sim 4.0$Hz passband corresponding to $40 \sim 240 BPM$ heart rate. The specular reflection s_i was denoised by

$$\begin{cases} \boldsymbol{S}_{bvp} = \boldsymbol{X}_f - \alpha \boldsymbol{Y}_f, \\ \alpha = \dfrac{\sigma(\boldsymbol{X}_f)}{\sigma(\boldsymbol{Y}_f)}. \end{cases} \tag{4}$$

C. CWT noise rejection

CWT is one of the time frequency analysis methods and expresses a signal as a superposition and parallel transport of various scale wavelets. This method obtained the continuous wavelet coefficients $X_w(\tau, s)$ that represented the similarity ratio between the original signal $\boldsymbol{x}$ and wavelets $\boldsymbol{\psi}_{\tau,s}$. Rach wavelet was parallel translated by τ and rescaled by s from a parent wavelet $\boldsymbol{\Psi}$ at time t:

$$\begin{cases} X_w(\tau, s) = \displaystyle\int_{-\infty}^{\infty} \boldsymbol{x}(t)\boldsymbol{\psi}_{\tau,s}(t)dt, \\ \boldsymbol{\psi}_{\tau,s} = \dfrac{1}{\sqrt{|s|}}\boldsymbol{\Psi}\left(\dfrac{t-\tau}{s}\right). \end{cases} \tag{5}$$

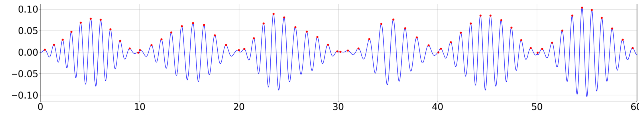
In this paper, we adopted the Morlet wavelets as the parent wavelets. On the assumption that the BVP signal was the dominant wave in the parent wavelets, the maximum total value s^* was given by the sum of the continuous wavelet coefficients:

$$s^* = \arg\max_s \sum_\tau \boldsymbol{X}_w(\tau, s). \tag{6}$$

BVP wave was recovered by the inverse continuous wavelet transform (ICWT) from the maximum total value:

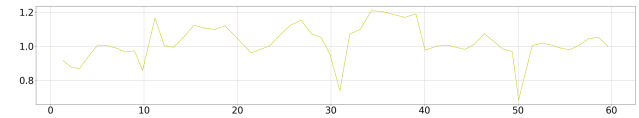
$$\boldsymbol{x}(t) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{\infty} \frac{1}{s^2}\boldsymbol{X}_w(\tau, s)\boldsymbol{\psi}_{\tau,s}(t)d\tau ds. \tag{7}$$

In this study, CWT noise rejection was performed to each of the regions generated by divided signals $\boldsymbol{S}_f$ by the time interval T (10s) after the application of the Hanning window. All regions were connected after decoding.



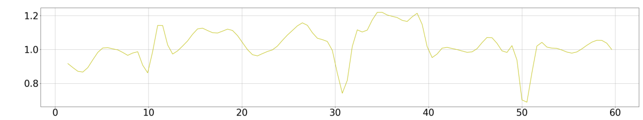
red: peak points, x-axis: time (s)

(a) BVP $\boldsymbol{S}_w$ and peaks



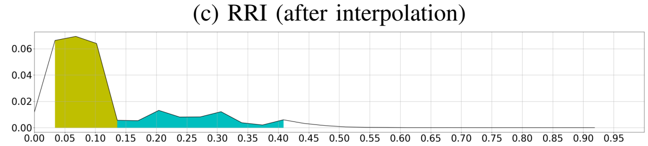
y-axis: RRI(s), x-axis: time (s)

(b) RRI (before interpolation)



y-axis: RRI(s), x-axis: time (s)

(c) RRI (after interpolation)



x-axis: frequency (Hz), yellow: LF, blue: HF

(d) Welch's method power spectrum

Fig. 4: Illustrates the steps of MRI pulse spectrum calculation

D. Calculate RRI power spectrum

RRI was given by the difference between the ith peak and the $i-1$th peak

$$RRI_i = t_{peak_i} - t_{peak_{i-1}}. \tag{8}$$

We used the average value μ_{RRI} between $1.3 * \mu_{RRI}$ and $0.7 * \mu_{RRI}$ to remove the peaks not originally existing in the connecting point. Then, the RRI power spectrum was calculated by Welch's method after data were equally spaced by the cubic spline interpolation.

E. Normalized feature quantity

The heart rate h, and the respiration rate r were obtained by the calculated RRI power spectrum, and LF, HF and LF/HF of this spectrum were used. We use two types of feature vector: 2D $\boldsymbol{V}_{hr} = \{h, r\}$ and 5D $\boldsymbol{V}_{hrv} = \{h, r, lf, hf, lf/hf\}$. The all feature vectors except $\boldsymbol{V}_P$ were normalized by the method in [15]:

$$\begin{cases} \hat{V}_{first_i} = \dfrac{V_{first_i} - \mu_{first_i}}{\sigma_{first}}, \\ \hat{V}_{second_i} = \dfrac{V_{second_i} - \mu_{first_i}}{\sigma_{first}}. \end{cases} \tag{9}$$

IV. EXPERIMENT

We conducted a comparison experiment between our estimation and HRV detection to verify our proposed method. The BVP measured by the PPG was chosen for comparision. 11 university students participated in the next three conditions.

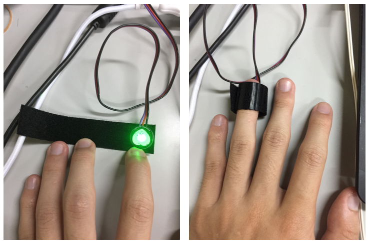


Fig. 5: PPG sensor

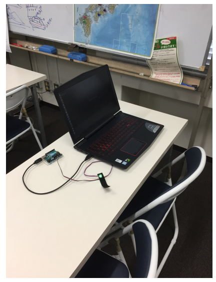


Fig. 6: Cond. 1 booth

- **Condition 1**: Participants execute a fatigue task under the light. A pair of data was measured before and after the task.
- **Condition 2**: Participants executed the same fatigue task in front of the light. A pair of data is measured in the morning and evening of a day (almost 8- hours interval). They attended classes or did research between being measured..
- **Condition 3**: During condition 2, they moved their head.

For 60 seconds, 3 x 2 datasets were collected in each condition. Visual Analog Scale (VAS) [19] was used to verify the subjective fatigue of participants. In this experiment, face data was recorded by a webcam (30 fps, 640×480) and the raw BVP data was detected by a PPG sensor by World Famous Electronics [20]. This sensor was attached to the participant's finger with a touch fastener (Figure 5) and data were collected through an Arduino Uno R3 with 100 Hz.

A. Condition 1

The purpose of this experimental condition was to collect shadowed facial data with the fatigue task the same as the study of Kawamura et al. [15]. We predicted that the luminance noise at the cheek region would effect the fatigue estimation performance. 33 data were collected from 11 participants in three days. Fig. 6 shows a view of experiment site, a meeting room with overhead fluorescent lights.

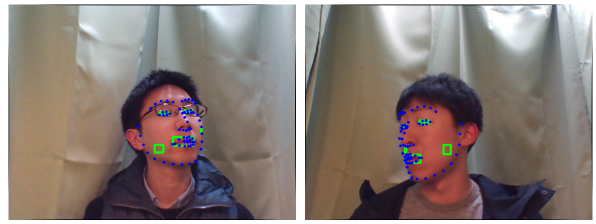
For this experiment, we chose a 15 minute calculation as the fatigue task. This task was a set of repeated determination of the addition-subtraction correctness within a time limit. The time limit was calculated by adding 0.5 seconds to the response time to practice questions before this task. Before and after this task, data was collected for 60 seconds by a GUI application. The experiment flow is as follows:

1) Enter the name in the start screen
2) Adjusts their sitting position, attach the PPG sensor, and click a button.
3) Keep stopping for 60 seconds
4) Answer VAS
5) Execute the fatigue task (10 practice question and the 15 minute task)
6) Repeat the process #2 $\sim$ 4

B. Condition 2

For this condition, as in#1 we measured twice in the day with the same 8-hour interval, but this time the light source was in front of the face. 33 data were collected from 11 participants. We predicted that the accuracy of the estimation would be higher than the result of condition 1. The participants sat in a curtained-off booth and an LED indirect light illuminated the face (Fig. 7). The experiment flow was as follows:

Fig. 7: Cond. 2 booth



(a) Shaking　　(b) Look back

Fig. 8: Condition 3

1) Enter their name and the round of the measurement (first or second) on the day.
2) Adjust their sitting position, attach the PPG sensor, and click a button.
3) Stop for 60 secondsto be measured
4) Answer VAS

C. Condition 3

In this condition, we didn't collect any raw BVP data. The setting was the same as for condition 2 but participants shook their head and looked back 1 to 3 times during the measurement. 24 data from 8 participants and 4 from 2 participants were collected. We predicted head movement would degrade the accuracy of the fatigue estimation.

V. RESULT

A. Comparison with raw PPG value

We evaluated the accuracy of heart rate h, respiration rate r and LF/HF. For the PPG waveform, linear interpolation was performed to obtain 200 Hz sampling rate, and Fourier interpolation was performed for the interpolated waveform. To eliminate noise, we used a threshold of the moving average of the waveform at the window size 75 multipled by 1.3. The scaling factor of the moving average was decreased by 0.05 until at least 10 peaks had been detected. Furthermore, we excluded values outside the range $(0.7\sigma_{rri}, 1.3\sigma_{rri})$ by using the RRI standard deviation σ_{rri}. After the measurement of the heart rate, the respiration rate and LF/HF, the ranges ($\pm\mu + 3\sigma$ for heart rate, respiration rate, and LF/HF) were excluded. We used 27 data in total (15 data of dataset 1 and 12 of dataset 2). Table I shows the measurement

TABLE I: Measurement error of heart rate, respiration rate, and LF/HF by rPPG

	Condition 1			Condition 2		
	HR	RR	LF/HF	HR	RR	LF/HF
Mean Error	4.28	4.34	2.40	5.15	5.37	2.46
Correlation	0.86	0.10	-0.031	0.77	0.10	0.01

TABLE II: Fatigue estimation result by linear SVM (60 seconds, one data set)

Dataset	Feature	D	A	P	R	Fm
1-1	$\widehat{\boldsymbol{V}}_t$	980	0.50	0.0	0.0	0.0
1-1	$\boldsymbol{V}_{hr}$	2	0.65	**1.0**	0.30	0.44
1-1	$\boldsymbol{V}_{hrv}$	5	**0.67**	**1.0**	**0.33**	**0.47**
2-2	$\widehat{\boldsymbol{V}}_t$	980	0.41	0.17	0.12	0.13
2-2	$\boldsymbol{V}_{hr}$	2	0.58	0.8	0.16	0.26
2-2	$\boldsymbol{V}_{hrv}$	5	**0.73**	**1.0**	**0.47**	**0.63**
2-1	$\widehat{\boldsymbol{V}}_t$	980	0.53	0.6	0.05	0.09
2-1	$\boldsymbol{V}_{hr}$	2	0.63	**1.0**	0.27	0.42
2-1	$\boldsymbol{V}_{hrv}$	5	**0.73**	**1.0**	**0.47**	**0.61**
2-3	$\widehat{\boldsymbol{V}}_t$	980	0.59	0.20	0.03	0.06
2-3	$\boldsymbol{V}_{hr}$	2	0.64	0.80	0.28	**0.38**
2-3	$\boldsymbol{V}_{hrv}$	5	**0.69**	**1.0**	**0.38**	**0.53**

The dimention (D). accuracy (A), precision (P), recall (R), F-measure (Fm), , Connected and compressed equency spectrum of cheek erea RGB waves (30 sec) ($\widehat{\boldsymbol{V}}_t$) , and Hat means the standardization

errors of LF/HF by rPPG waveform. Participants reported the subjective fatigue became stronger in the afternoon than in the morning. However, PPG and rPPG detected fatigue increase for just two data of each wave pulse and the accuracy was about 50% to $\sim$ 58%. This result shows a simple PPG or rPPG waveform estimation cannot estimate fatigue adequately.

B. BVP feature fatigue estimation

We evaluated the estimation accuracy for several combinations of BVP features. Feature vectors $\boldsymbol{V}_{hr}$ and $\boldsymbol{V}_{hrv}$ were created from 60- and 30-second facial videos and fatigue estimated by linear SVM referring to [6]. We compared 4 dataset conditions (SVM training-test): two within-dataset 1-1 and 2-2, between-dataset 2-1 and 2-3. 3-3 was not conducted because head moving situation affected the tracking landmmark. Especially, it affected FE features objectionability and we used 3 for just verification. The accuracy (A), precision (P), recall (R), and F-measure (Fm) of positive label were obtained by 5-fold cross validation The cost parameter C of the SVM was optimized by the grid search at each cross validation. As a baseline score, we used the feature vector by the previous method [15] In this method, 30 seconds feature vectors $\hat{\boldsymbol{V}}_t$ were connected to seven $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_t$ compressed to 140 dimensions by LLE with five seconds shifting. Table II shows the result of this evaluaton.

The comparison of fatigue estimation shows the proposed features $\boldsymbol{V}_{hr}$ and $\boldsymbol{V}_{hrv}$ were better than the baseline feature $\boldsymbol{V}_{t+s}$ in all conditions. $\boldsymbol{V}_{hr}$ and $\boldsymbol{V}_{hrv}$ received higher scores than the baseline also in the cross environment dataset. In conclusion, our proposed method can estimate the BVP robustly to the lighting noise. Moreover, in this experiment, $\boldsymbol{V}_{hrv}$ is better for fatigue estimation than g $\boldsymbol{V}_{hr}$.

TABLE III: Fatigue estimation result by RBF kernel SVM

Dataset	Feature	D	A	P	R	Fm
1-1	$\hat{\boldsymbol{V}}_t$	140	0.95	0.97	0.93	0.94
1-1	$\hat{\boldsymbol{V}}_{t+s}$	120	0.93	0.94	0.93	0.93
1-1	$\boldsymbol{V}_I$	15300	0.95	**1.0**	0.90	0.94
1-1	$\boldsymbol{V}_l$	36000	0.88	0.82	**1.0**	0.89
1-1	$\boldsymbol{V}_{gaze}$	5400	0.95	**1.0**	0.90	0.94
1-1	$\hat{\boldsymbol{V}}_{gaze}$	180	**0.96**	**1.0**	0.93	**0.96**
1-1	$\boldsymbol{V}_{size}$	2700	0.80	0.90	0.70	0.78
1-1	$\hat{\boldsymbol{V}}_{size}$	**90**	0.88	**1.0**	0.77	0.84
2-2	$\hat{\boldsymbol{V}}_t$	140	0.90	0.95	0.87	0.89
2-2	$\hat{\boldsymbol{V}}_{t+s}$	120	0.89	0.91	0.88	0.89
2-2	$\boldsymbol{V}_I$	15300	0.92	0.97	0.88	0.92
2-2	$\boldsymbol{V}_l$	36000	0.75	0.72	**0.97**	0.84
2-2	$\boldsymbol{V}_{gaze}$	5400	0.94	**1.0**	0.88	0.93
2-2	$\hat{\boldsymbol{V}}_{gaze}$	180	**0.96**	**1.0**	0.91	**0.95**
2-2	$\boldsymbol{V}_{size}$	2700	0.92	**1.0**	0.83	0.90
2-2	$\hat{\boldsymbol{V}}_{size}$	**90**	0.94	**1.0**	0.89	0.94
2-1	$\hat{\boldsymbol{V}}_t$	140	0.55	0.60	0.09	0.15
2-1	$\hat{\boldsymbol{V}}_{t+s}$	120	0.70	**1.0**	0.41	0.58
2-1	$\boldsymbol{V}_I$	15300	0.95	0.99	0.91	0.94
2-1	$\boldsymbol{V}_l$	36000	0.77	0.73	**0.92**	0.81
2-1	$\boldsymbol{V}_{gaze}$	5400	0.92	1.0	0.84	0.91
2-1	$\hat{\boldsymbol{V}}_{gaze}$	180	**0.96**	**1.0**	**0.93**	**0.96**
2-1	$\boldsymbol{V}_{size}$	2700	0.88	0.98	0.78	0.87
2-1	$\hat{\boldsymbol{V}}_{size}$	**90**	0.91	1.0	0.82	0.90
2-3	$\hat{\boldsymbol{V}}_t$	140	0.58	0.60	0.15	0.24
2-3	$\hat{\boldsymbol{V}}_{t+s}$	120	0.52	0.40	0.04	0.07
2-3	$\hat{\boldsymbol{V}}_I$	15300	0.81	0.92	0.70	0.79
2-3	$\hat{\boldsymbol{V}}_l$	36000	0.64	0.63	**0.85**	0.71
2-3	$\boldsymbol{V}_{gaze}$	5400	0.82	**1.0**	0.63	0.76
2-3	$\hat{\boldsymbol{V}}_{gaze}$	190	**0.86**	**1.0**	0.72	**0.83**
2-3	$\boldsymbol{V}_{size}$	2700	0.78	**1.0**	0.55	0.69
2-3	$\hat{\boldsymbol{V}}_{size}$	**90**	0.82	**1.0**	0.64	0.76

The dimention (D). accuracy (A), precision (P), recall (R), F-measure (Fm), gaze vector ($\boldsymbol{V}_{gaze}$) , degree of opening of eye and mouth ($\boldsymbol{V}_{size}$) , Connected and compressed equency spectrum of eye and mouth erea HoG features (25-30sec) ($\hat{\boldsymbol{V}}_{t+s}$) , and Hat means the standardization

The paper shows that it is possible to estimate fatigue simply by simple heart rate when that the target has post exercise physical fatigue. On the liner SVM fatigue estimation, our method improved the accuracy of estimation by the BVP over the previous method.

C. FE features fatigue estimation experiment

We verified FE features $\boldsymbol{V}_l$, $\boldsymbol{V}_I$, $\boldsymbol{V}_{size}$, $\boldsymbol{V}_{gaze}$, $\hat{\boldsymbol{V}}_{size}$, and $\hat{\boldsymbol{V}}_{gaze}$ for the fatigue estimation in the two within-dataset conditions (1-1, 2-2) and two between-dataset conditions (2-1, 2-3). Since we think the discrimination boundary of FE features will be nonlinear, we used an RBF kernel SVM. As a baseline, the features $\boldsymbol{V}_t$ and $\boldsymbol{V}_{t+s}$ were chosen, which was used by the previous method, and condition2 were chosen, which was best controlled (light and no moving) in the three condition. We used 30 seconds of the facial video first in the datasets. The cost parameter C of SVM and the kernel parameter γ of the SVM were optimized by the grid search in each cross validations. Table III shows the result of this verification.

In the within-dataset conditions, our method exhibited a similarly good performance to the baseline. Meanwhile, in the between-dataset condition, the estimation ability of the baseline was greatly decreased. Particularly, the recall and F-measure were lower than 0.5, and the accuracy with $\boldsymbol{V}_{t+s}$ become lower than with $\boldsymbol{V}_t$. Our proposed method can estimate with the recall and F-measure over 0.5 in the between-dataset condition. In this experiment, the accuracy, precision and the F-measure of $\hat{\boldsymbol{V}}_{gaze}$ were the highest.

This result shows the accuracy rate of the previous method was degraded in 0.6, while the proposed method can estimate with over 0.6 accuracy. Especially, using the average of line-of-sight vectors of per one second, the proposed method achieved over 0.8 accuracy. Therefore, our method improved also the accuracy of estimation by FE over the previous method.

VI. CONCLUSIONS & DISCUSSION

Healthcare will be one of the care jobs fthat service robots can undertake in a hospital, a care home, or a private home. Fatigue estimation is considered important because it could indicate disease. Our proposed method, using the BVP feature and the FE feature shows the ability of fatigue estimation. The method is contactless with an RGB camera, hence it is easy to implement to various existing home robots. This study will therefore improve the potential of home robots. In the future, the stereo camera may become reduced in size and it will improve the recognition accuracy of facial features using this method.

A motion in the robot (vibration) is serious problem of this method. When the robot move strongly, the PPG data will be fluctuant by the distracting facial features. The large vibration in the robot cause psychological effects of target person (heart rate elevation etc.). Additionally, the distance will also effect accuracy or cause psychological effects. We add suppose the necessity to implement our method in a robot and should evaluate it on various distances.

To verify the method, the number of the experiments are small. We will rethink experimental designs to manage the large number of participants. The method in this paper only estimated the relative fatigue by using data from one day. If a one-shot fatigue estimation is required, longitudinal data collection is needed, and some absolute value of fatigue. To accomplish this, we would have to create a quantitative index such as LF/HF using an electrocardiogram. Since long-term data will include various changes over the terms of that could cause errors. The development of the next prediction model should answer these issues. Furthermore, a comparison of the results with past methodologies will be necessary.

In this paper, a method of fatigue estimation was proposed that can account for against changes in lighting condition and head movements. The method uses two features: the BVP feature consisting of heart rate, breathing rate and heart rate variability index (LF/HF), and FE features consisting of face landmark points, the degree of opening/closing, AUs expression/intensity levels and line-of-sight vectors. The method accomplished a more accurate fatigue estimation than the previous method using the frequency spectrum of luminosity value.

REFERENCES

[1] T. Sundelin, M. Lekander, G. Kecklund, E. J. W. Van Someren, A. Olsson, and J. Axelsson, "Cues of Fatigue: Effects of Sleep Deprivation on Facial Appearance," *Sleep*, vol. 36, no. 9, pp. 1355–1360, 2013.

[2] Wang Rongben, Guo Lie, Tong Bingliang, and Jin Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems*. IEEE, pp. 314–319.

[3] M. Eriksson and N. Papanikotopoulos, "Eye-tracking for detection of driver fatigue," in *Proceedings of Conference on Intelligent Transportation Systems*. IEEE, pp. 314–319.

[4] M. S. Devi and P. R. Bajaj, "Driver Fatigue Detection Based on Eye Tracking," in *2008 First International Conference on Emerging Trends in Engineering and Technology*. IEEE, 2008, pp. 649–652.

[5] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, mar 2007.

[6] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, aug 2015, pp. 6398–6404.

[7] Q. Ji, Z. Zhu, and P. Lan, "Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, jul 2004.

[8] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," *Journal of Medical Engineering & Technology*, vol. 32, no. 6, pp. 479–484, jan 2008.

[9] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, p. 21434, dec 2008.

[10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, jan 2011.

[11] G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, oct 2013.

[12] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *BioMedical Engineering OnLine*, vol. 15, no. 1, p. 11, jan 2016.

[13] M. Saradadevi and P. Bajaj, "Driver Fatigue Detection Using Mouth and Yawning Analysis," *International Journal of Computer Science and Network Secuirity*, vol. 8, no. 6, pp. 183–188, 2008.

[14] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692–706, oct 2014.

[15] R. Kawamura, N. Takemura, and K. Sato, "Mental fatigue estimation based on luminance changes in facial images," in *2016 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, dec 2016, pp. 526–531.

[16] S. T. Roweis, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, dec 2000. [Online]. Available: http://www.sciencemag.org/cgi/doi/10.1126/science.290.5500.2323

[17] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," *Technical Report CMU-CS, CMU School of Computer Science*, vol. 16, no. 118, 2016.

[18] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, may 1999.

[19] E. Ahsberg, F. Gamberale, and K. Gustafsson, "Perceived fatigue after mental work: an experimental evaluation of a fatigue inventory," *Ergonomics*, vol. 43, no. 2, pp. 252–268, feb 2000.

[20] W. F. E. llc. (2011), "Pulsesensor.com – world famous electronics llc." https://pulsesensor.com/products/pulse-sensor-amped.