Noncontact heart rate measurement using a high-sensitivity camera in a low-light environment

Genki Okada[1], Ryota Mitsuhashi[1], Keiichiro Kagawa[2], Shoji Kawahito[2], and Norimichi Tsumura[1,2]

[1]Graduate School of Advanced Integration Science, Chiba University, CHIBA, JAPAN
[2]Research Institute of Electronics, Shizuoka University, HAMAMATSU, JAPAN
(Tel: 81-43-290-3262, Fax: 81-43-290-3262)

[1]acaa2640@gmail.com

Abstract: We propose a method for the remote estimation of the heart rate and heart rate variability spectrogram by analyzing the hemoglobin concentration obtained from RGB facial videos taken in a low-light environment. The monitoring of emotion has potential in areas such as market research, safety, and health. In particular, methods of analyzing the heart rate obtained from RGB video are expected to be used practically. However, these studies cannot be applied in dark locations where monitoring is necessary, such as an infant's bedroom, a crime-prone road, and within a car. The proposed method therefore uses a highly sensitivity RGB camera capable of capturing videos at low illuminance. As the result, we could measure the heart rate with accuracy exceeding 99% and estimate the heart rate variability spectrogram with high accuracy for low-light environments of 10 lux, which corresponds to brightness levels of the monitoring environments given above.

Keywords: heart rate, high-sensitivity camera, illuminance, low-light environment, photoplethysmography

1 INTRODUCTION

The monitoring of emotion has great potential application in areas such as market research, safety, health, and robotics. Among such applications, the prevention of potential accidents or crime is important in realizing a safe social system. Monitoring with a surveillance camera, for example, allows the detection of people who are trying to commit a crime or who are excessively excited. Furthermore, accidents due to dangerous driving can be prevented by monitoring the driver of the car and measuring drowsiness and concentration. This technique of emotion monitoring has long been studied. Many researchers have attempted to realize emotion recognition using, for example, facial expressions [1][2], voices [3][4], and physiological signals [5][6][7]. In particular, physiological signals have attracted the most attention in recent years for emotion recognition. In the field of physiological psychology, it is known that there is strong correlation between the physiological response through the action of the autonomic nervous system and the human emotional state. Furthermore, physiological signals are less affected by social and cultural differences [8]. We can estimate original emotions that people were trying to hide or that they could not even recognize in themselves.

Park et al. [5] used electrodes to measure physiological signals of the skin temperature, electrodermal activity, photoplethysmography, and electrocardiogram of 12 healthy participants before and after they watched movies that elicited seven emotions (i.e., happiness, sadness, anger, fear, disgust, surprise, and stress). They were able to classify the seven emotions with around 90% accuracy by selecting useful features for emotion recognition through the particle swarm optimization of features obtained by analyzing the measured physiological signals. In this way, it is possible to classify emotions using physiological signals. However, such an approach is not practical because it requires special measuring devices, such as contact-type devices. Moreover, contact-type devices might be uncomfortable for the participants in that place a burden on participants and thus induce stress.

Kurita et al. [6] and Okada et al. [7] realized a remote heart rate variability (HRV) measurement system using an RGB camera by analyzing the hemoglobin concentration obtained from color facial images. They identified if participants were relaxed or stressed or felt any of five emotions by performing frequency analysis on the HRV. They were able to detect stress without causing unnecessary discomfort to the participants. However, their approach could not be applied to low-brightness images taken in the dark, such as those taken inside a car or at night, because they used an ordinary camera.

Zhao et al. [9] measured heart and respiration rates by applying delay-coordinate transformation and independent component analysis. Because they used a camera that is sensitive to visible and near-infrared light, their method can be applied both during the day and at night. However, their approach requires a near-infrared light-emitting diode.

The present paper proposes a method of measuring a pulse wave using an ultrahigh sensitivity camera in a low-illumination environment, which does not require special light sources or contact devices. Our method deals with the hemoglobin concentration obtained by analyzing facial images captured with an ultrahigh sensitive RGB camera capable of capturing with a high inter-scene dynamic range even under low illuminance.

2 METHOD OF REMOTE MEASUREMENT FOR A PULSE WAVE

Various methods of pulse wave measurement have been proposed using a camera. Kurita et al. [6] and Okada et al. [7] measured the pulse wave without contact by detecting the hemoglobin concentration from a facial image and acquiring the temporal change. This method is based on biological optics and is credible. The present paper therefore applies independent component analysis to the RGB pixel values of the facial image to separate skin pigments and to extract hemoglobin-component images. We treat the change in the average pixel value of the hemoglobin-component images as a pulse wave.

Figure 1 shows the model of human skin. Human skin is a multilayer structure that can be roughly divided into the epidermis, dermis, and subcutaneous tissue. In practice, the boundary surface of each layer has an irregular shape. However, we treat the boundary surface as a planar shape for simplicity. Human skin contains melanin and hemoglobin pigments. The color tone of human skin is greatly affected by these pigments. Melanin pigments exist in the epidermis and hemoglobin pigments exist in the dermis. Melanin and hemoglobin pigments can therefore be regarded as being present with spatially independence by assuming that the epidermis is a melanin layer and the dermis is a hemoglobin layer. Light incident on the human skin can be divided into light reflected at the surface and internally reflected light emitted to the outside of the skin after repeated absorption and scattering within the skin. While surface-reflected light represents the color of the light source, such as in the case of gloss, the internally reflected light represents the color of the skin. In this paper, we take images without surface-reflected light with polarizing plates placed in front of the camera and a light source using the algorithm proposed by Ojima et al. [10]. When the modified Lambert–Beer law is assumed to hold with respect to the observed signal of the reflected light, the observed signal can be represented via logarithmic conversion from the image space to the density space as

$$\boldsymbol{v}^{\log}(x,y) = -\rho_m(x,y)\boldsymbol{\sigma}_m - \rho_h(x,y)\boldsymbol{\sigma}_h + p^{\log}(x,y)\boldsymbol{1} + \boldsymbol{e}^{\log}, \quad (1)$$

where $\boldsymbol{v}^{\log}$ is the converted observation signal, (x, y) is the pixel location, ρ_m and ρ_h are respectively the concentrations of melanin and hemoglobin pigments, $\boldsymbol{\sigma_m}$ and $\boldsymbol{\sigma_h}$ are respectively the absorption cross sections of melanin and hemoglobin pigments, $p^{\log}$ is a shading parameter relating to the shape of the skin, $\boldsymbol{1}$ is a vector of the strength of the shading, and $\boldsymbol{e}^{\log}$ is a bias vector. We can therefore regard melanin and hemoglobin pigments as independent signals as shown in Figure 2. It is thus possible to obtain the melanin and hemoglobin pigment concentration distributions from RGB values of facial images.

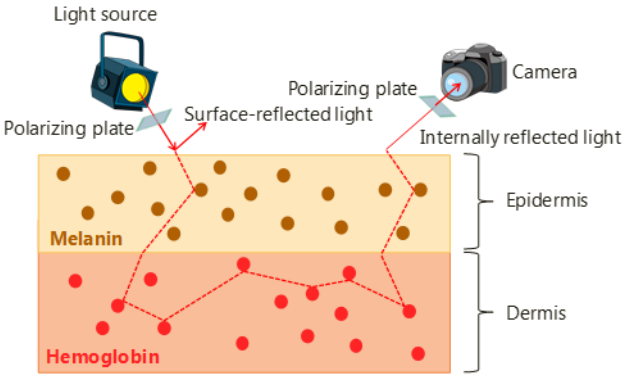


Fig. 1. Movement of light incident on the skin

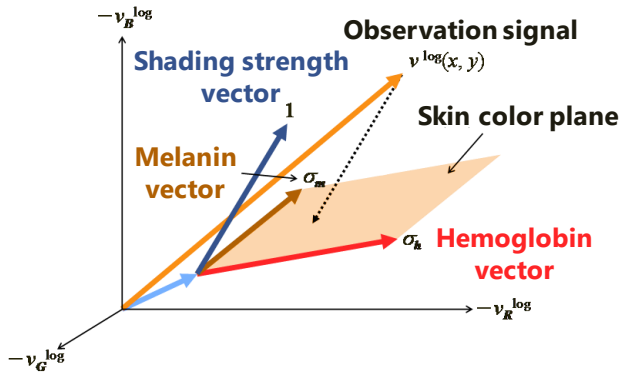
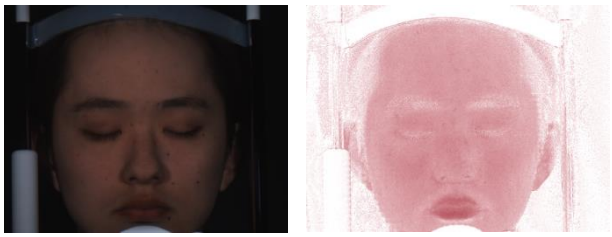


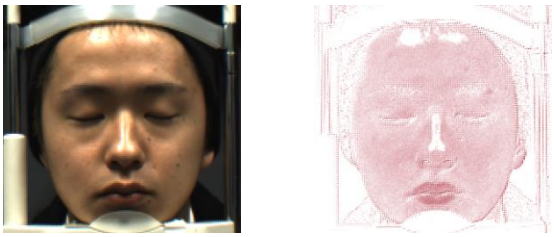
Fig. 2. Obtained signals and three independent signals

Figures 3 (b) shows the hemoglobin pigments extracted by independent component analysis of the internal-reflection facial image shown in Figure 3 (a). Figure 4 (a) is a facial image taken under fluorescent light. In the case that the facial image contains surface-reflected light, we can separate skin pigments as shown in Figure 4 (b) using each pigment component color vector estimated from the internal-reflection image.



(a) Original (b) Hemoglobin

Fig. 3. Result of skin pigment separation for an internal-reflection facial image: (a) original, (b) hemoglobin



(a) Original (b) Hemoglobin

Fig. 4. Result of skin pigment separation for a fluorescent lamp image: (a) original, (b) hemoglobin

The change in the average pixel values of the hemoglobin-component images is the signal of the blood volume change. However, temporal changes in pixel values acquired from the entire image also include changes due to the blinking of eyes and body movements. Therefore, a region excluding the eyes in the facial region detected employing the Viola–Jones method [11], the region of the forehead (i.e., half the width and top 25% of the face), and the region around the mouth (i.e., half the width and the bottom half of the face) were set as the region of interest (ROI). Figure 5 shows sets of the ROI in the hemoglobin-component image while Figure 6 shows the change in time of the average pixel values of the ROI in the hemoglobin-component images. The peaks of the signal of the blood volume change correspond to the peaks of the electrocardiogram waveform called an R wave. The intervals between R waves are called RR intervals and are important in heart-rate analysis. The signal was detrended to allow easy peak detection [12]. Subsequently, the detrended signal was multi-band-pass filtered with a Hamming window to reduce noise in the original wave. The multi-band-pass filter was adjusted for each signal by setting the width of the peak closest to 1 Hz and the width of its second harmonic in the frequency-converted detrended signal to the pass band. A frequency of 1 Hz corresponds to 60 beats per minute (bpm), which is a general heart rate (HR) for a normal state. The filtered signal was interpolated with a cubic spline function at 50 Hz to match the sampling frequency of the electrocardiogram measured as the correct value. The RR intervals were calculated by peak detection with respect to the filtered signal. Figures 7 and 8 show the detrended, filtered signal and RR intervals.

Fig. 5. Area of the set ROI

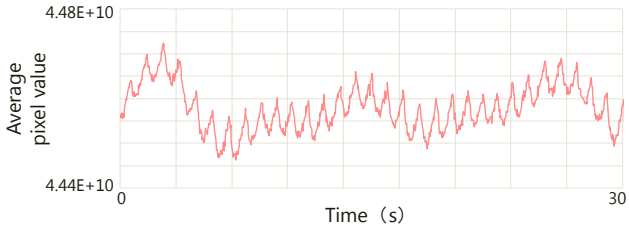


Fig. 6. Average pixel values of hemoglobin-component images

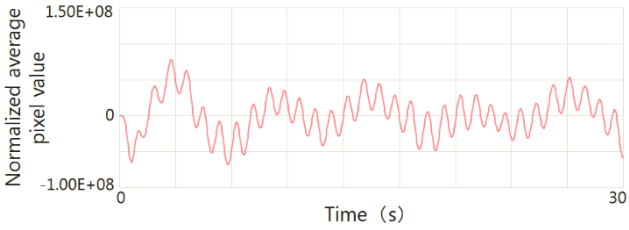


Fig. 7. Normalized, detrended and filtered signal

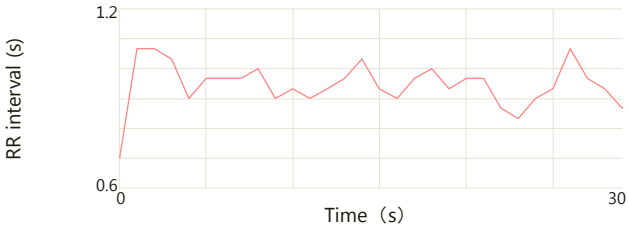


Fig. 8. RR intervals

The noise from the camera due to low illuminance affects the accuracy of HR analysis. The present paper filters RR intervals using the non-causal of variable threshold algorithm [13] to remove such noise.

We calculated the HR and HRV to compare the performances of the electrocardiogram and camera methods. The HR was calculated as $60/RR$, where RR denotes the mean of RR intervals. The HRV, which is the variation of consecutive heartbeats, is modulated by both the sympathetic and parasympathetic branches of the autonomic nervous

system. The most conspicuous periodic component of the HRV is considered to range from 0.15 to 0.4 Hz. In addition to the physiological effect of breathing on the HRV, the high-frequency (HF) component in this range is generally believed to be of parasympathetic origin. Another widely studied component of the HRV is the low-frequency component ranging from 0.04 to 0.15 Hz, which has been thought to be of both sympathetic and parasympathetic origin. The components of the HRV have been found to correlate with, for example, age, mental and physical stress, and attention [14]. The highly accurate estimation of the HRV spectrum is therefore important. We created an HRV spectrogram by calculating the power spectral density of the RR intervals for each moving window. The power spectral density was calculated using the Lomb periodogram [15]. Conventional spectral analysis techniques, such as Welch's method [16], require that the input signal be uniformly sampled. If the sampling is not uniform, such as in the case of the RR interval, the signal needs to be resampled or interpolated to a uniform sample rate. However, such processing can add undesirable artifacts to the spectrum, leading to parsing errors. Because the Lomb periodogram directly processes nonuniform samples, resampling and interpolation are unnecessary, and it is useful for the spectral analysis of RR intervals. We used a window of 1 minute and a step size of 1 s.

3 EXPERIMENT CONDUCTED UNDER VARIOUS INTENSITIES OF ILLUMINANCE

The experiments were conducted in a dark room as shown in Figure 9 and one healthy Asian male student participated. An 18-bit camera [Xviii: ViewPLUS] and dimming light source were placed 1 m from the participant. We used a camera that has ultrahigh sensitivity and a high inter-scene dynamic range even if the exposure conditions hardly change and that can simultaneously capture subjects with illuminance of 0.01 and 400 lux or more at 30 fps. Illuminance of 0.01 lux is typically that of a half moon or starlight while illuminance of 400 lux is typically that under fluorescent lighting.

The present study acquired a pulse wave employing the aforementioned method and the pixel values of the ultrahigh sensitivity camera with 18-bit output. The dimming light source was used to continuously adjust the brightness manually in accordance with the illuminance of the participant's face. Face was fixed using a chin rest in this study.

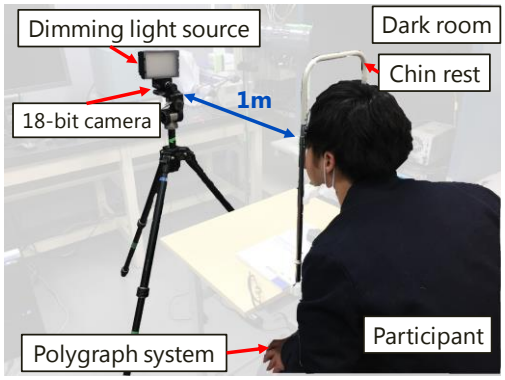


Fig. 9. Experimental environment

Prior to the experiment, participant was introduced to the procedure of the experiments and had time to adapt and feel comfortable in the experimental environment. The brightness of the light source, measured with an illuminometer, was adjusted so that the illuminance of the face was 5, 10, 50, 100, 200, and 300 lux. The participant' face was captured for 2 minutes at each illuminance.

We measured the correct value of the HR from the electrocardiogram using a polygraph system [RMT-1000: Nihon Kohden Inc]. We set the measurement resolution at 50 Hz and applied a low-pass filter. The cut-off frequency was set at 15 Hz. This was sufficient to get the peaks of R waves. From the signals, we obtained RR intervals and calculated the HR as ground truth data employing the method used for the hemoglobin-component signal.

4 RESULTS

Figure 10 shows the accuracy of the HR at each illuminance. The accuracy is calculated as

$$\text{The accuracy} = 100 - \frac{|E - G|}{G} \times 100, \quad (2)$$

where E is the estimated HR and G is the ground truth data obtained from the electrocardiogram. Figure 11 shows the noise in the captured images by the ultrahigh sensitivity camera at low illuminance of 5 and 10 lux.

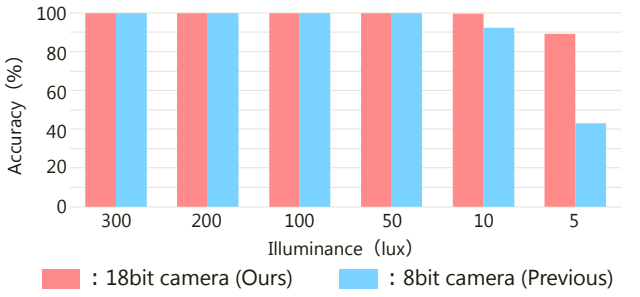
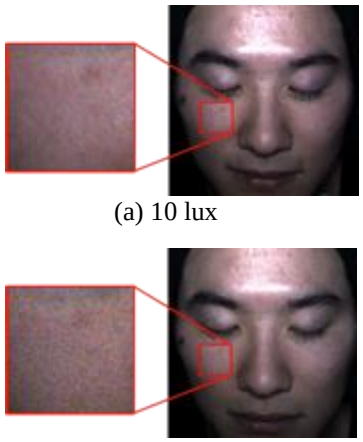


Fig. 10. Accuracy of HR at each illuminance

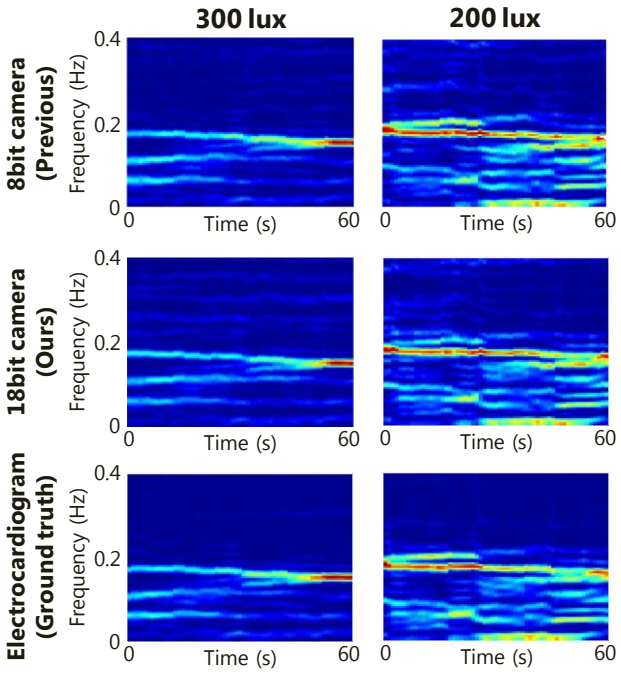


(a) 10 lux

(b) 5 lux

Fig 11. Noise in the captured images at low illuminance

Figure 12 compares the HRVs obtained from the previous 8-bit camera, the ultrahigh sensitivity camera, and the electrocardiogram at each illuminance. When the illumination environment was between 50 and 300 lux, the results obtained by the cameras were almost the same as those calculated from the electrocardiogram. In the case of 10 lux, the noise in the estimation result of the 18-bit camera increased slightly but the frequency with the highest power spectral density was similar to the ground truth.



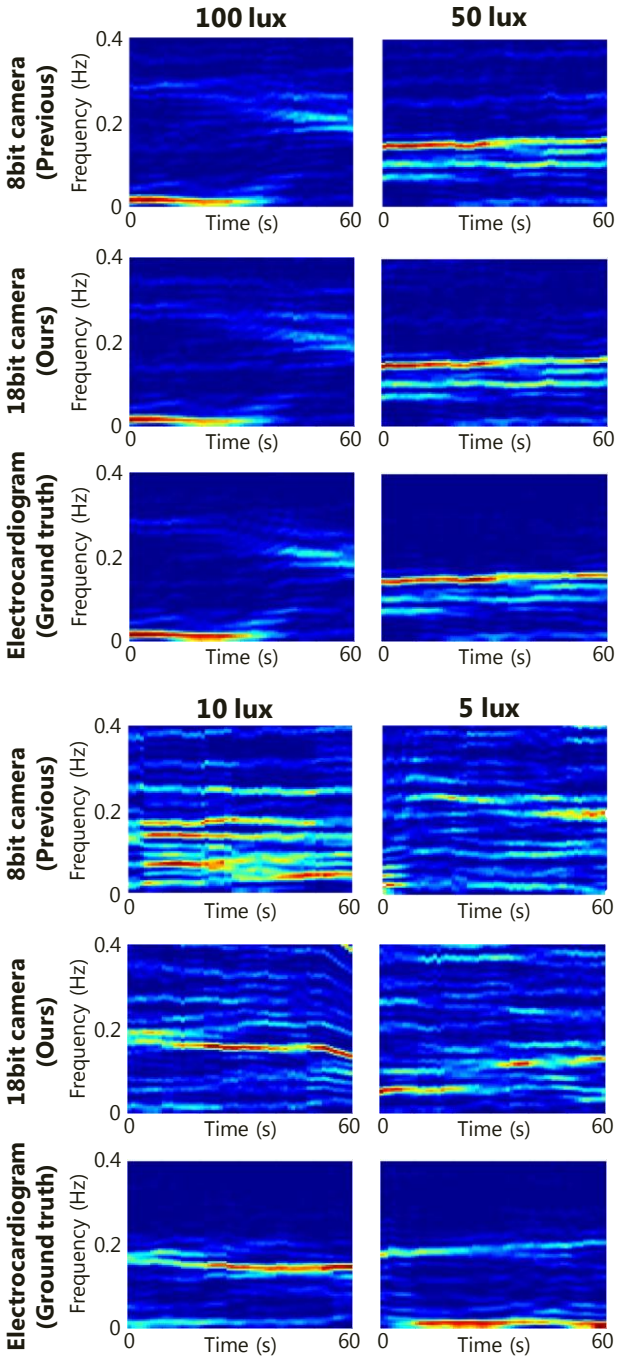


Fig. 12. Comparison of HRVs

5 DISCUSSION

Illuminances of 300, 200, 100, 50, 10 and 5 lux respectively correspond to the brightness of a general office under fluorescent lighting, family living room, night entrance, the ground under a street light, twilight, and a residential street at night. The HR was estimated with nearly 100% accuracy using the 18-bit camera in lower-illuminance environments. The results show that our method can be applied to general household environments and roads.

Measurement of the HR in such dark places has applications relating to sudden infant death syndrome, driver monitoring, and crime deterrence. The estimation at 5 lux was less accurate than that for the brighter illumination environments. This low accuracy seems to be due to noise from the camera being dominant as shown in Figure 11. However, the result shows that the accuracy was considerably better than that achieved with an 8-bit camera.

The HRV is modulated by the autonomic nervous system. Spectral analysis of the HRV is important because the mental state, such as a stressed or drowsy state, can be estimated from the function of the autonomic nervous system. The calculated HRV spectrograms were almost perfectly reproduced at and above 50 lux. Even at 10 lux, although noise was slightly stronger, the strongest spectra were almost the same as the ground truth. The result therefore seems sufficiently accurate for the detection of stress. The value estimated when the face illuminance was 5 lux was different from the true value. This result seems to be affected by the predominant noise that was not removed even by processing, such as spatial averaging, multi-band-pass filtering, and use of the non-causal of variable threshold algorithm. To effectively handle such noise, additional processing should be considered before extracting the hemoglobin content.

6 CONCLUSION AND FUTURE WORK

We measured the pulse wave from the hemoglobin concentration obtained by analyzing an RGB image of the face taken with a high-sensitivity camera in a low-illuminance environment. The HR and HRV spectrogram useful for emotion estimation were calculated from the measured pulse wave. The accuracies of the estimation based on the HR and HRV spectrogram were almost 100% under illumination exceeding 10 lux.

Our future work is to improve the accuracy at low illuminance of 5 lux or less by eliminating random noise due to the camera and to apply the technique to the monitoring of driver emotion. Emotion can be estimated regardless of the lighting condition using a camera having ultrahigh sensitivity.

REFERENCES

[1] Darwin C, Cummings MM, Duchenne GB (1872) The expression of the emotions in man and animals. London: John Murray

[2] Ekman P (1993) Facial expression and emotion. American Psychologist, 48(4):384

[3] Zhang T, Johnson MH, Levinson SE (2004) Children's Emotion Recognition in an Intelligent Tutoring Scenario. Proc. Eighth European Conf. Speech Comm. and Technology (INTERSPEECH)

[4] Kwon OW, Chan K, Hao L, Lee TW (2003) Emotion Recognition by Speech Signals. Proc. Eighth European Conf. Speech Comm. and Technology (EUROSPEECH)

[5] Park B-J, Jang E-H, Kim S-H, Huh C, Sohn J-H (2012) Seven emotion recognition by means of particle swarm optimization on physiological signals: Seven emotion recognition. Proc. 9th IEEE ICNSC, Apr. 2012, pp. 277–282

[6] Kurita K, Yonezawa T, Kuroshima M, Tsumura N (2015) Non-contact video based estimation for heart rate variability spectrogram using ambient light by extracting hemoglobin information. Color Imag Conf 2015(1): 207-211

[7] Okada G, Yonezawa T, Kurita K, Tsumura N (2018) Monitoring emotion by remote measurement of physiological signals using an RGB camera. ITE Transactions on Media Technology and Applications, 6(1):131–137

[8] Alaoui-Ismaili O, Robin O, Rada H, Dittmar A, Vernet-Maury E (1997) Basic emotions evoked by odorants: Comparison between autonomic responses and self-evaluation. Physiol Behav 62:713-720

[9] Zhao F, Li M, Qian Y, Tsien JZ (2013) Remote measurements of heart and respiration rates for telemedicine. PLoS ONE 8(10):e71384.

[10] Ojima N, Minami T, Kawai M (1997) Transmittance measurement of cosmetic layer applied on skin by using processing. Proceeding of The 3rd Scientific Conference of the Asian Societies of Cosmetic Scientists

[11] Viola P, Jones M (2001) Rapid Object Detectio n using a Boosted Cascade of Simple Features. Procee dings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume: 1, pp.511–518.

[12] Tarvainen MP, Ranta-aho PO, Karjalainen PA (2002) An advanced detrending method with application to HRV analysis. IEEE T Bio-med Eng 49(2):172-175

[13] Vila J, Palacios F, Presedo J, Ferna´ndez-Delgado M, Felix P, Barro S (1997) Time-frequency analysis of heart-rate variability. IEEE Eng Med Biol Mag 16(5):119-126

[14] Berntson GG, Thomas Bigger J, Eckberg DL, Grossman P, Kaufmann PG, Malik M, Nagaraja HN, Porges SW, Saul JP, Stone PH, Van Der Molen MW (1997) Heart rate variability: Origins, methods, and interpretive caveats. Psychophysiology 34:623-648

[15] Press WH, Teukolsky SA, Vetterling WA, Flann ery BP (2001) Numerical recipes in Fortran 77: the art of scientific computing. 2nd ed., vol. 1, Cambridge U niversity Press, NY, USA.

[16] Welch PD (1967) The use of Fast Fourier Tran sform for the estimation of power spectra: A method b ased on time averaging over short, modified periodogra ms", IEEE Transactions on Audio and Electroacoustics, AU-15 (2): 70–73