2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Non-contact Measurement of Pulse Rate Variability Using a Webcam and Application to Mental Illness Screening System

Maho Nishikawa, Batbayar Unursaikhan, Takuya Hashimoto, *Member, IEEE*, Masaki Kurosawa, Tetsuo Kirimoto, *Senior Member, IEEE*, Toshikazu Shinba, Takemi Matsui, and Guanghao Sun, *Senior Member, IEEE*

Abstract— **The COVID-19 pandemic is a global health crisis. Mental health is critical in such uncertain situations, particularly when people are required to significantly restrict their movements and change their lifestyles. Under these conditions, many countries have turned to telemedicine to strengthen and expand mental health services. Our research group previously developed a mental illness screening system based on heart rate variability (HRV) analysis, enabling an objective and easy mental health self-check. This screening system cannot be used for telemedicine because it uses electrocardiography (ECG) and contact photoplethysmography (PPG), that are not widely available outside of a clinical setting. The purpose of this study is to enable the extension of the aforementioned system to telemedicine by the application of non-contact PPG using an RGB webcam, also called imaging-photoplethysmography (iPPG). The iPPG measurement errors occur due to changes in the relative position between the camera and the target, and due to changes in light. Conventionally, in image processing, the pixel value of the entire face region is used. We propose skin pixel extraction to eliminate blinks, eye movements, and changes in light and shadow. In signal processing, the green channel signal is conventionally used as a pulse wave owing to the absorption characteristics of blood flow. Taking advantage of the fact that the red and blue channels contain noise, we propose a signal reconstruction method for removing noise and strengthening the signal in the pulse rate variability (PRV) frequency band by weighting the three signals of the RGB camera. We conducted an experiment with 13 healthy subjects, and showed that the PRV index and pulse rate (PR) errors estimated by the proposed method were smaller than those of the conventional method. The correlation coefficients between estimated values by the proposed method and reference values of LF, HF, and PR were 0.86, 0.69, and 0.96, respectively.**

I. Introduction

Given the COVID-19 pandemic, telemedicine, offering the benefit of preventing contact with others, is drawing attention [1]. Some hospitals are flooded with patients experiencing symptoms like fevers and coughs that are not due to COVID-19 [2]. Moreover, pandemic-related stressors are causing people to develop anxiety, depression, sleep disorders, and other signs of stress [3]. In this situation, providing telemedicine to patients who are not infected with COVID-19 will reduce the hospital burden.

The inter-beat interval (IBI) is regulated by the autonomic nervous system and varies in response to changes in physical and emotional states. The temporal change in IBI is called heart rate variability (HRV); this measure is used in the objective and quantitative evaluation of autonomic nervous system (ANS) activity. The ANS comprises the sympathetic nervous system and the parasympathetic nervous system, the former being active during tension and the latter during relaxation. Both the sympathetic and parasympathetic nerves act on the low-frequency component (LF: 0.04~0.15 Hz) of HRV, and the parasympathetic nerves act on the high-frequency component (HF: 0.15~0.40 Hz) of HRV [4]. The low-frequency component value divided by the high-frequency component value, LF/HF, is used as an index of the sympathetic nervous system. Imaging-photoplethysmography (iPPG) is a non-contact technique that uses a camera to measure the blood volume pulse (BVP). Since BVP corresponds to the beating of the heart, heart rate (HR) and HRV can be replaced by pulse rate (PR) and pulse rate variability (PRV) obtained from BVP [5].

Our research group previously revealed changes in the responsiveness of autonomic nervous system activity in mental illness by measuring HRV at rest and during a task [6]. Building on this work, we have developed a mental illness screening system using HRV indices and machine learning [7,8]. The system uses an ECG and contact PPG to measure IBI. Because ECG equipment is usually only available in clinical settings, replacing ECG with iPPG extends the system availability to many more people. As such, the proposed iPPG-based mental illness screening system is expected to support and promote telepsychiatry. Nevertheless, iPPG does pose a challenge: unlike contact-based methods of HRV measurement, iPPG is affected by changes in the relative position of the measurement target and the sensor, and by changes in light. The resultant low accuracy of the PRV indices poses a problem when applying iPPG to the system. In this study, we propose two new analysis methods to improve the accuracy of PRV and PR values obtained from iPPG signals. At the image processing stage, we propose skin pixel extraction to extract only skin pixels from the facial image

*Research supported by JSPS Grant-in-Aid for Scientific Research (B) No. 19H02385.

M. Nishikawa, M. Kurosawa, T. Kirimoto and G. Sun are with the Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan. (e-mail: nishikawa@secure.ee.uec.ac.jp, Guanghao.Sun@uec.ac.jp).

B. Unursaikhan and T. Matsui is with the Graduate School of System Design, Tokyo Metropolitan University, 6-6, Asahigaoka, Hino, Tokyo, Japan.

T. Hashimoto is with Department of Mechanical Engineering, Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo, Japan.

T. Shinba is with Department of Psychiatry, Shizuoka Saiseikai General Hospital. 1-1-1 Oshika, Suruga-ku, Shizuoka, Shizuoka, Japan.

captured by the camera, eliminating remove motion artifacts and specular reflection. At the signal processing stage, we propose signal reconstruction using spectral information to strengthen the signal in the PRV frequency band by the weighted sum of RGB signals. This is inspired by SoftSig, which utilizes the fact that the iPPG signal amplitude depends on the wavelength [9]. To assess the proposed method, we conducted an experiment on 13 healthy subjects in a laboratory setup.

II. Materials and Methods

Figure 1 shows an overview of the entire analysis method. First, a front view of the person is captured by the camera, and OpenCV face detection is performed. Next, the pulse signal is estimated. At this stage, the proposed methods, that are skin pixel extraction and signal reconstruction using spectral information, are performed. The estimated PR is calculated from the frequency with the strongest power in the 0.7~2 Hz range of the power spectral density (PSD) of the estimated pulse signal of each method. Subsequently, the low-frequency component is removed by subtracting the estimated pulse signal filtered with a moving average of the 1.43 second window from the original estimated pulse signal. Then, the beats are detected from the signal, and the IBI is calculated. Raw IBI is unequally spaced, so a 100 Hz cubic spline interpolation is performed on the data for frequency analysis. Then, IBI is filtered with a moving average of the 2.5 second window. LF and HF are obtained by integrating the PSD of IBI over the 0.04~0.15 Hz and 0.15~0.40 Hz ranges, respectively. We use a fast Fourier transform (FFT) to calculate all PSDs.

A. Skin Pixels Extraction

First, the thresholds of the pixel values of the skin are calculated from a certain frame of the video (automatic threshold calculation). Next, skin pixels are extracted from each frame using the thresholds.

The origin is in the upper left corner of the image taken by the camera, with horizontal coordinates increasing toward the right, and vertical coordinates increasing toward the bottom. OpenCV face detection is performed on the image after the camera is activated and automatic settings such as focus and exposure are stable. The detected face is represented by the coordinates of the rectangular area. The coordinates of the vertices of the rectangular face region are expressed as follows, clockwise from the upper left:

$$\left(l_{face}, t_{face}\right), \left(r_{face}, t_{face}\right), \left(r_{face}, b_{face}\right), \left(l_{face}, b_{face}\right). \quad (1)$$

From the detected face, the upper part of the nose, which contains only the skin, is selected as the region of interest for automatic threshold calculation. The coordinates of this skin region are calculated by the following equations, using the coordinates of the face:

$$l_{skin} = l_{face} + \left(r_{face} - l_{face}\right) \times 0.3, \quad (2)$$

$$r_{skin} = r_{face} - \left(r_{face} - l_{face}\right) \times 0.3, \quad (3)$$

$$t_{skin} = t_{face} + \left(b_{face} - t_{face}\right) \times 0.48, \quad (4)$$

$$b_{skin} = b_{face} - \left(b_{face} - t_{face}\right) \times 0.42, \quad (5)$$

where l_{skin}, r_{skin}, t_{skin}, and b_{skin} represent the left, right, top, and bottom coordinates of the skin region, respectively, and l_{face}, r_{face}, t_{face}, and b_{face} represent the left, right, top, and bottom coordinates of the detected face, respectively. The RGB image of the skin region is converted into the HSL image. The pixel values of the 1.5 percentile and 98.5 percentile of each histogram of H, S, and L are selected as the lower and upper thresholds, respectively.

The following operations are performed for each frame of the face video. A window is created that holds pixels within the skin-pixel thresholds. The window is applied to the RGB

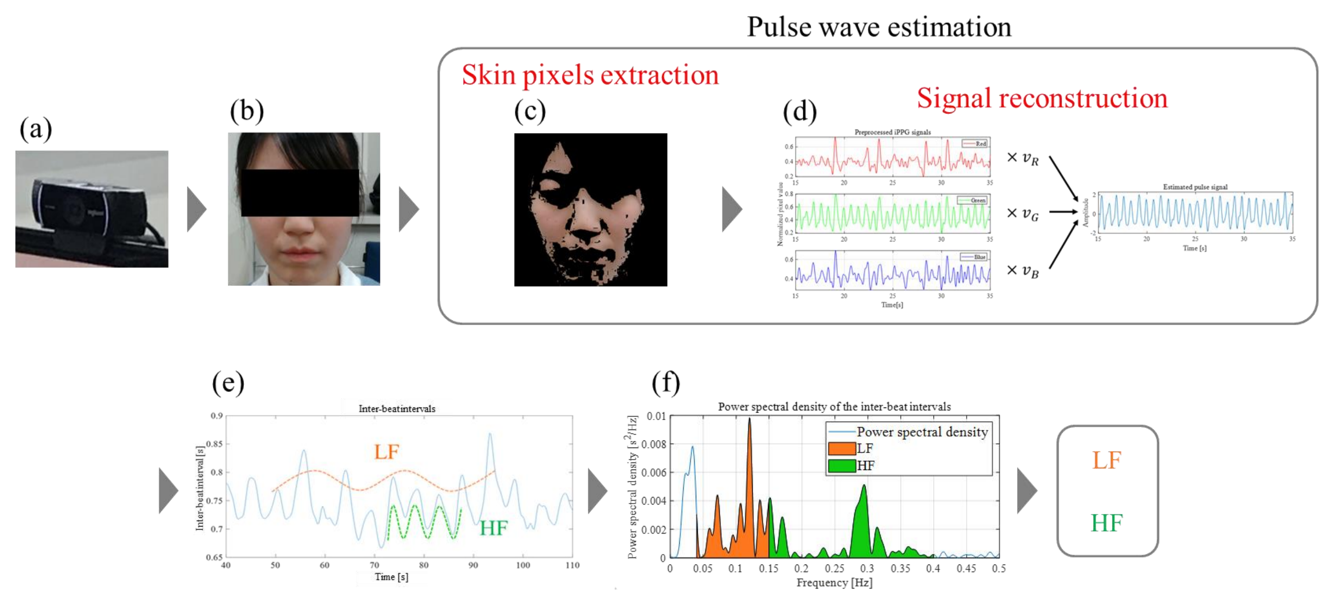

Figure 1. An overview of the analysis method. (a) RGB camera, (b) An example of the RGB image of the OpenCV detected face, (c) An example of the RGB image of the extracted skin pixels, (d) The conceptual diagram of the signal reconstruction using the spectral information, (e) An example of the waveform of the IBI, (f) An example of the PSD of the IBI.

image of the detected face, and only skin pixels are extracted as shown in Figure 1 (c).

B. *Signal Reconstruction Using the Spectral Information*

The relative PPG amplitude is G>B>R. As such, the green signal is considered to have a strong pulse wave component, the red signal to have a strong noise component, and the blue signal to include both. The green signal is multiplied by a coefficient selected in the range 0.1 to 1, and the red and blue signals by a coefficient selected in the range -1 to 1; then the results are added. This effectively removes noise and enhances the pulse wave component. The weighted sum of RGB signals is shown below.

$$C = [v_R \quad v_G \quad v_B]\begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{6}$$

where C is the weighted sum signal, v_R, v_G, and v_B are the weighting coefficients (scalar values), and R, G, and B are the signals that are calculated from the spatial mean values of the RGB images of the extracted skin video. The combination of coefficients that maximizes the signal-to-noise ratio (SNR) of the PRV frequency band in the PSD of the weighted sum signal is selected as the optimum weighting coefficient. The SNR is defined by the following formula:

$$S = \frac{\int_{f_p-0.4}^{f_p+0.4} P(f)df}{\int_0^{f_p-0.4} P(f)df + \int_{f_p+0.4}^{f_n} P(f)df}, \tag{7}$$

where S is the SNR, f is the frequency, $P(f)$ is the PSD of the weighted sum signal C, f_p is the frequency with the strongest power in the 0.7~2 Hz range, and f_n is the Nyquist frequency of C. Unlike SoftSig [9], the characteristic feature of this method is that it strengthens the signal component not only in the PR frequency f_p Hz but also in the PRV frequency band $f_p \pm 0.4$ Hz. The estimated pulse signal is obtained by substituting the combination of weighting coefficients with the largest SNR into (6). Similarly, the POS [10] method referred in the paper [11] extracts a pulse signal by a weighted sum of RGB signals. It changes the weighting coefficients slightly using a fixed projection matrix and "alpha tuning" [10]. On the other hand, the signal reconstruction method in this paper freely changes the balance of the coefficients according to signal conditions.

C. *Experimental Setup*

To evaluate the accuracy of the estimated indices, the simultaneous measurement of iPPG and ECG was performed for 13 healthy students and faculty members (10 males and 3 females) of The University of Electro-Communications and Tokyo Metropolitan University. The experimental setup is shown in Figure 2. For iPPG acquisition, the RGB videos of the faces were recorded at 30 fps with a resolution of 1440 × 1080 using a Logicool C922 Pro Stream WebCam. The 30 Hz raw RGB signals were resampled to 300 Hz at the beginning of the signal processing stage. The distance between the camera and the subjects was about 50 cm. ECG was recorded at 300 Hz using a GMS memory heart rate monitor LRR-03 and a National Instruments DAQ USB-6003.

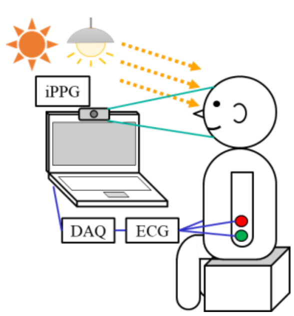

Figure 2. Experimental setup.

Three ECG electrodes were attached to each subject's wrists. We measured iPPG and ECG twice each for 2 min at rest and used the central 90 seconds of them for analysis. As the measurements were taken in a room with windows, the subjects' faces were exposed to light emanating from two sources, the ceiling lights and sunlight. To evaluate the performance of the proposed method, the accuracy of the results of each PRV index calculated by the proposed method were compared with those calculated by the conventional method. The difference between these two methods is the pulse wave estimation. In the conventional method, the signal of the spatial mean value of the green channel image of the entire face region detected by OpenCV is used as the estimated pulse signal. The correlation coefficient and the root mean square error (RMSE) with the reference value were used as evaluation indices. This study was approved by the Ethics Committee of Tokyo Metropolitan University and The University of Electro-Communications. All subjects gave their informed written consent.

III. Results and Discussion

Tables 1 and 2 show the correlation coefficient and RMSE, which represent the estimation accuracy of each index. Figure 3 (a), (b), and (c) show scatter plots depicting the relationship between the estimated value and the reference value of each index. These results indicate that the proposed method is effective in removing noise for all indices. Figure 3 (d), (e), and (f) show the Bland-Altman plots of each index. The 95% confidence interval represented by the chain line is narrower in the proposed method than in the conventional method. From this, it can be concluded that the proposed method improves the estimation of each index. When LF and HF were estimated by the conventional method, the error was mostly positive, due to a slight deviation of the estimated value of the pulse time causing vibration of IBI and strengthening the power in the LF or HF frequency band. However, when LF and HF were estimated by the proposed method, the error was neither positively nor negatively biased. Thus, it can be concluded that the proposed method improved the estimation of the beat timing.

TABLE I. The Correlation Coefficients Between Estimated Values and Reference Values of Each Index.

Index	**Correlation Coefficient**	
	Proposed method	***Conventional method (Green channel)***
LF	0.75	0.03
HF	0.50	0.05
PR	0.96	-0.06

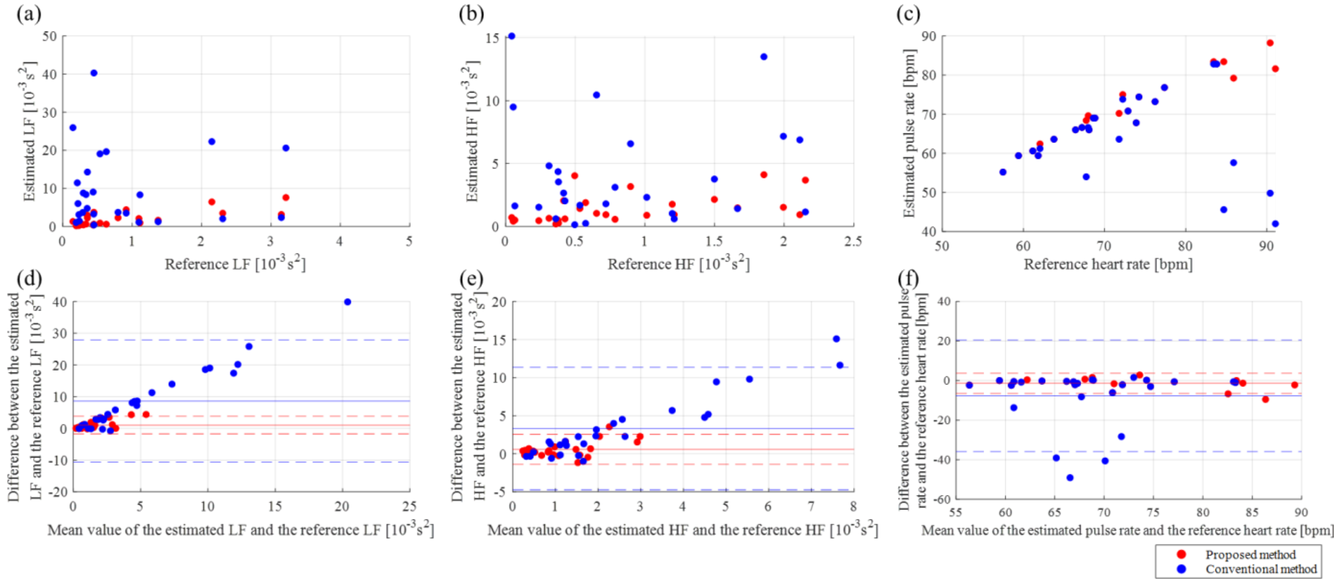

Figure 3. The result of estimation of each index. (a), (b), and (c) Scatter plots of LF, HF, and pulse rate, (d), (e), and (f) Bland-Altman plots of LF, HF, and pulse rate.

TABLE II. THE ROOT MEAN SQUARE ERROR OF EACH INDEX.

Index	Root Mean Square Error	
	Proposed method	*Conventional method (Green channel)*
LF $[10^{-3}s^2]$	1.73	12.94
HF $[10^{-3}s^2]$	1.15	5.20
PR [bpm]	2.9	16.1

IV. CONCLUSION

In this study, we proposed an iPPG analysis method using skin pixel extraction and signal reconstruction using spectral information to improve the accuracy of the estimated PRV indices and PR. We conducted the experiment in a laboratory setup to evaluate the proposed method. The results showed that the proposed method is effective in reducing the estimation error in the PRV indices and PR.

Because skin pixels are extracted from the entire video using skin color information from a certain point in time, no skin pixels may be extracted if the subject moves significantly or the light changes. It is, therefore, necessary to reconsider how to select the skin image to which the threshold calculation refers, so that the skin pixel extraction approach can be applied across a greater variety of situations. There is room for reconsidering the definition of the PRV frequency band used in signal reconstruction. In this study, we used constants empirically. As the PRV tends to weaken as the PR increases, the PRV frequency band can be defined using the PR.

In future work, we will evaluate the proposed method during utterances because our mental illness screening system uses utterance as a mental task. We also aim to improve the estimation accuracy of LF/HF, which is an index of only the sympathetic nervous system.

REFERENCES

[1] C. Duncan, Reimund Serafi ca, Denis Williams, Michael Kuron, and Alana Rogne, “Telepsychiatry during the COVID-19 pandemic,” in *The Nurse Practitioner*, vol. 45, no. 12, December, 2020.

[2] Shuangyi Sun, Y. Keting, Z. Xie and X. Pan, “China empowers Internet hospital to fight against COVID-19,” in *Journal of Infection*, vol. 81, no. 1, pp. E67-E68, July 2020.

[3] R.P. Rajkumar, “COVID-19 and mental health: A review of the existing literature,” *Asian Journal of Psychiatry*, vol. 52, article 102066, August 2020.

[4] M. Malik, J.T. Bigger, A.J. Camm, R.E. Kleiger, A. Malliani, A.J. Moss and P.J. Schwartz, “Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,” in *European Heart Journal*, vol. 17, no. 3, pp. 354-381, 1996.

[5] M.Z. Poh, D.J. McDuff and R.W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” in *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7-11, January 2011.

[6] T. Shinba, “Altered autonomic activity and reactivity in depression revealed by heart-rate variability measurement during rest and task conditions,” in *Psychiatry and Clinical Neurosciences*, vol. 68, no. 3, pp. 225-233, 2014.

[7] M. Kobayashi, G. Sun, T. Shinba, T. Matsui and T. Kirimoto, “Development of a mental disorder screening system using support vector machine for classification of heart rate variability measured from single-lead electrocardiography,” in *2019 IEEE Sensors Applications Symposium (SAS)*, Sophia Antipolis, France, 2019, 1-6.

[8] Guanghao Sun, Toshikazu Shinba, Tetsuo Kirimoto, and Takemi Matsui, “An Objective Screening Method for Major Depressive Disorder Using Logistic Regression Analysis of Heart Rate Variability Data Obtained in a Mental Task Paradigm,” in *Frontiers in Psychiatry*, 7(180), 2016.

[9] W. Wang, A.C. den Brinker and G. de Haan, “Single Element Remote-PPG,” in *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2032-2043.

[10] W. Wang, A. C. den Brinker, S. Stuijk and G. de Haan, “Algorithmic Principles of Remote PPG,” in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479-1491, July 2017.

[11] A. Gudi, M. Bittner, and J. van Gemert, “Real-Time Webcam Heart-Rate and Variability Estimation with Clean Ground Truth for Evaluation,” *Applied Sciences*, vol. 10, no. 23, p. 8630, Dec. 2020.