Noncontact Sensing of Heart Rate Variability from Facial Video Using the Topology Algorithm

Takuya Sakamoto\(^1\) and Tsuyoshi Nakamura\(^2\)

1 Graduate School of Engineering, Kyoto University
Kyotodaiaku-Katsura, Nishikyo-ku, Kyoto 615–8510, Japan
2 Innovation Center, Connected Solutions Company, Panasonic Corporation
4-1-62, Minoshima, Hakata-ku, Fukuoka, 812–8531, Japan
\(^a\) sakamoto.takuya.8n@kyoto-u.ac.jp

Abstract: This study proposes a noncontact, accurate sensing method for heartbeat R–R intervals and heart rate variability from the video of a target person’s face. The proposed method comprises three steps: (1) automatic face recognition and tracking, (2) the detection of skin-tone pixels in the face region, and (3) heart rate estimation using the topology algorithm. We applied the proposed method to facial video recordings of 55 participants and evaluated its accuracy. The results demonstrate that the root-mean-square error in estimating the heart rate variability was 2.7 ms using the proposed method, demonstrating the effectiveness of the proposed approach.

Keywords: facial video, topology algorithm, heart rate variability

Classification: Sensing

References

[1] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, “Video-based heart rate measurement: Recent advances and future prospects,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600—3615, Oct. 2019. DOI:10.1109/JM.2016.2615472

[2] T. Sakamoto and K. Yamashita, “Noncontact measurement of autonomic nervous system activities based on heart rate variability using ultra-wideband array radar,” *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 3, pp. 208–215, Sep. 2020. DOI:10.1109/JERM.2019.2948827

[3] 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. DOI:10.1109/TBME.2013.2266196

[4] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote PPG,” *IEEE Transactions on Biomedical Engineering*, vol. 64 no. 7, pp. 1479–1491, Jul. 2017. DOI:10.1109/TBME.2013.2266196

[5] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, “Estimation of respiratory rate from photoplethysmographic imaging videos compared to pulse oximetry,” *IEEE Journal of Biomedical
and Health Informatics, vol. 19, no. 4, pp. 1331–1338, Jul. 2015. DOI: 10.1109/JBHI.2015.2429746

[6] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” Optics Express, vol. 18, no. 10, pp. 10762–10774, Oct. 2010.

[7] H. Monkaresi, R. A. Calvo, and H. Yan, “A machine learning approach to improve contactless heart rate monitoring using a webcam,” IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pp. 1153–1160, Jul. 2014. DOI:10.1109/JBHI.2013.2291900

[8] H. Ghanadian, M. Ghodratigohar, and H. Al Osman, “A machine learning method to improve non-contact heart rate monitoring using an RGB camera,” IEEE Access, vol. 6, pp. 57085–57094, Oct. 2018. DOI:10.1109/ACCESS.2018.2872756

[9] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, “Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition,” IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 5, pp. 1422–1433, Sep. 2017. DOI:10.1109/JBHI.2016.2615472

[10] K. Alghoul, S. Alharthi, H. Al Osman, and A. El Saddik, “Heart rate variability extraction from videos signals: ICA vs. EVM comparison,” IEEE Access, vol. 5, pp. 4711–4719, Mar. 2017. DOI:10.1109ACCESS.2017.2678521

[11] S. Sanyal and K. K. Nundy, “Algorithms for monitoring heart rate and respiratory rate from the video of a user’s face,” IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, Apr. 2018. DOI:10.1109/JTEHM.2018.2818687

[12] S. Kado, Y. Monno, K. Yoshizaki, M. Tanaka, and M. Okutomi, “Spatial-spectral-temporal fusion for remote heart rate estimation,” IEEE Sensors Journal, vol. 20, no. 19, pp. 11688–11697, May 2020. DOI:10.1109/JSEN.2020.2997785

[13] X. Yu, T. Laurentius, C. Bollheimer, S. Leonhardt, and C. H. Antink, “Noncontact monitoring of heart rate and heart rate variability in geriatric patients using photoplethysmography imaging,” IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1781–1792, May 2020. DOI:10.1109/JBHI.2020.3018394

[14] R. Favilla, V. C. Zuccala, and G. Coppini, “Heart rate and heart rate variability from single-channel video and ICA integration of multiple signals,” IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 6, pp. 2398–2408, Nov. 2019. DOI:10.1109/JBHI.2018.2880097

[15] N. Sugita, M. Akay, Y. M. Akay, and M. Yoshizawa, “Noise reduction technique for single-color video photoplethysmography using singular spectrum analysis,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 6, pp. 1788–1795, Jun. 2020. DOI:10.1109/JBHI.2019.2949883

[16] T. Sakamoto, R. Imasaka, H. Taki, T. Sato, M. Yoshikawa, K. Inoue, T. Fukuda, and H. Sakai, “Feature-based correlation and topological similarity for interbeat interval estimation using ultrawideband radar,” IEEE Transactions on Biomedical Engineering, vol. 63, no. 4, pp. 747–757, Apr. 2016. DOI:10.1109/TBME.2015.2470077

[17] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” Proc. 7th International Joint Conference on Artificial Intelligence, vol. 2, pp. 674—679, Vancouver, BC Canada, Apr. 1981.
1 Introduction

Noncontact heart rate (HR) and heart rate variability (HRV) measurement is of great interest because it can be used in various applications of healthcare monitoring [1, 2] without any sensors attached to the target person’s body. In particular, camera-based HR measurement using facial video is promising because it requires only a common digital camera that is compact, low-cost, and ubiquitous. In video-based HR sensing, after detecting the facial region from the video, motion artifacts are suppressed [3, 4], and pixels with skin tone are extracted to generate a signal that represents the heartbeat [5–9]. For estimating the HR from the signal, however, additional signal processing is required, for which most existing studies adopt a frequency-domain approach (e.g., the Fourier transform) to obtain a frequency spectrum whose peak is often used to estimate the HR [10–14].

Time-domain approaches are a promising alternative to frequency-domain approaches. Because they do not require time-frequency analysis, they have the potential to achieve higher time resolution than frequency-domain approaches. Despite this, there are only a few studies on time-domain HR estimation (e.g., [15]), and thus, this study focuses on the applicability of a time-domain method called the topology algorithm [16] to facial video data. Although the topology algorithm was originally developed for radar-based HR measurements, the algorithm may also be applied to other types of data. This study is the first to investigate the applicability of the topology algorithm to facial video data. In this study, we apply the topology algorithm to video recordings of 55 participants to estimate the HR (R–R interval: RRI) and HRV (standard deviation of the average normal-to-normal intervals: SDNN), whose accuracies are quantitatively evaluated.

2 Heart Rate Estimation Using the Topology Algorithm

In this section, we present a concise explanation of the topology algorithm [16]. A real-valued signal \( s(t) \) is sampled and \( s(t_i) \) for \( t_i = i\Delta t \) \((i = 0, 1, \cdots, N)\) are obtained, where \( \Delta t \) is a sampling interval. We define six kinds of feature labels \( g_i \) for \( t_i = t_n \) that are characterized by \( \dot{s}(t) = ds(t)/dt \), \( \ddot{s}(t) = d^2s(t)/dt^2 \), and \( \dddot{s}(t) = d^3s(t)/dt^3 \) as

\[
 g_i = \begin{cases} 
 1 & \text{for } \dot{s}(t_i) = 0, \ddot{s}(t_i) < 0 \\
 2 & \text{for } \dot{s}(t_i) = 0, \ddot{s}(t_i) > 0 \\
 3 & \text{for } \dot{s}(t_i) > 0, \ddot{s}(t_i) = 0, \dddot{s}(t_i) < 0 \\
 4 & \text{for } \dot{s}(t_i) > 0, \ddot{s}(t_i) = 0, \dddot{s}(t_i) > 0 \\
 5 & \text{for } \dot{s}(t_i) < 0, \ddot{s}(t_i) = 0, \dddot{s}(t_i) < 0 \\
 6 & \text{for } \dot{s}(t_i) < 0, \ddot{s}(t_i) = 0, \dddot{s}(t_i) > 0, 
\end{cases} \quad (1)
\]

and the label \( g_i = 0 \) is given to samples that do not satisfy any of the above conditions.

First, a vector with a length of \( 2K + 1 \) comprising data samples whose center is at \( t_n \) is generated as

\[
 v_n = [s(t_n - K\Delta t), s(t_n - (K - 1)\Delta t), \cdots, s(t_n + K\Delta t)]^T, \quad (2)
\]
only if \( g_n \neq 0 \) \((n = 1, \cdots, N)\). The correlation coefficient \( c_{m,n} \) between \( v_m \) and \( v_n \) is calculated, where \( c_{m,n} \) is defined only for a pair of feature points with the same non-zero label \((g_m = g_n, g_m, g_n \neq 0)\). We then generate another vector \( u_n \) that takes a complex value determined by a sequence of labels \( g_i \) \((i = 1, \cdots, N)\) [16]. The topological similarity \( M_{m,n} \) is calculated as the correlation coefficient between \( u_m \) and \( u_n \).

In the topology algorithm, the normal correlation coefficient \( c_{m,n} \) is used to detect the global waveform similarity between adjacent pulses, whereas the topological similarity \( M_{m,n} \) is used to detect the local waveform similarity over a short duration, which allows us to select reliable pairs of feature points. Using the topology method, which uses both \( c_{m,n} \) and \( M_{m,n} \), the HR estimation accuracy can be substantially improved, as has been confirmed using radar signals [16]. Its applicability to camera-based signals has not yet been established.

For comparison, we also use a conventional method [10–14] based on the fast Fourier transform (FFT) to estimate the HR. This method uses a rectangular window \( w(t) \) that is defined as \( w(t) = 1 \) for \( |t| \leq T_w / 2 \) and \( w(t) = 0 \) for \( |t| > T_w / 2 \), where we set \( T_w = 3.3 \) s, and calculates a spectrogram \( S(t, f) \) as
\[
S(t, f) = \left| \int_{-\infty}^{\infty} w(\tau) s(\tau - t) e^{-j2\pi f \tau} d\tau \right|^2. \tag{3}
\]
The frequency \( f_{\text{max}}(t) = \arg \max_f \ S(t, f) \) is then used to estimate the instantaneous RRI as \( \bar{T}_{\text{RR}}(t) = 1 / f_{\text{max}}(t) \).

3 Experimental Performance Evaluation of the Proposed Method

In this section, we evaluate the performance of the topology algorithm through experiments involving participants. A total of 55 healthy people (51 men and 4 women) participated in this study. The ages of the participants ranged from 24 to 59 years, and the average age was 42.8 years. Figure 1 shows the experimental setting with a participant seated with eyes closed. We used a video camera (HC-WX970M-K, Panasonic Corp., Osaka, Japan) and an electrocardiogram (ECG) device (Check My Heart, TRYTECH Co., Ltd., Tokyo, Japan). The video was recorded with a resolution of qHD \((960 \times 540 \) pixels\), a frame rate of 30 fps, and the digital video format iFrame with intraframe compression. The face of each participant was placed approximately 1.0 m from the video camera. We adjusted the lighting so that the illuminance on the participant’s face was between 500 and 600 lux. The measurement time for each participant was \( T_0 = 300 \) s.

First, to detect and track the face region in the video, the Kanade–Lucas—Tomasi feature tracker [17] was used. Next, pixels with skin tones were detected by applying a color space conversion so that specular reflection and diffuse light components could be separated. The extracted skin-tone pixel values were averaged to generate a time-domain signal. Finally, the topology algorithm was applied to the time-domain signal and the HR and HRV estimates were obtained. For each participant, instan-
Fig. 1: Experimental setting with a video camera and ECG with limb leads.

Instantaneous RRI \((T_{RR}(t))\) was estimated, and we then calculated the average RRI \((\bar{T}_{RR})\) and average SDNN \((\bar{T}_{SDNN})\) as \(\bar{T}_{RR} = \frac{1}{T_0} \int_0^{T_0} T_{RR}(t)dt\) and \(\bar{T}_{SDNN}^2 = \frac{1}{T_0} \int_0^{T_0} (T_{RR}(t) - \bar{T}_{RR})^2 dt\), respectively.

Fig. 2 (a) shows an example of the instantaneous RRI estimated using the topology algorithm applied to a signal from participant 35. In the figure, a reference RRI measured using the ECG device is also displayed for comparison. According to the ECG data, the average RRI and SDNN for this participant were 953.5 and 47.5 ms, respectively. The errors in estimating the RRI and SDNN using the facial video and topology algorithm were 2.8 and 2.6 ms, respectively. These values are much smaller than the ECG-based RRI and SDNN estimates.

Figs. 2 (b) and (c) respectively show the RRIIs and SDNNs estimated using the conventional method (blue), the proposed topology algorithm (red), and reference ECG (black) for all 55 participants. The average ECG-based RRI and SDNN estimates were 868.3 and 42.0 ms, respectively. The average errors in estimating the RRI using the conventional method and topology algorithm, calculated and averaged over 55 participants, were found to be 9.5 and 2.7 ms, respectively. Moreover, the average errors in estimating the SDNN using the conventional method and topology algorithm were 15.9 and 2.8 ms, respectively. The relative errors in estimating the average RRI and SDNN were 1.1% and 37.9% using the conventional method, and they were 0.3% and 6.7% using the topology algorithm, respectively. These results demonstrate the effectiveness of the proposed method for the measurement of the HR and HRV using facial video recordings.

4 Conclusion

In this study, we proposed a noncontact HR sensing method using facial video. The proposed method automatically detects the facial region from the video frames; pixels with skin tone are then extracted and their values are averaged to generate a signal represented as a time series. The signal is then processed by the topology algorithm to extract the feature points and then calculate the correlation and topological similarity coefficients, resulting in the estimation of an instantaneous RRI. We performed experiments
Fig. 2: Example of instantaneous RRI (a), average RRIs (b), and average SDNNs (c) estimated using the video-based methods and ECG.

involving 55 participants using a video camera and ECG recording device. The experimental results showed the average errors in estimating the RRI and SDNN were 2.7 and 2.8 ms, respectively, demonstrating the effectiveness of the proposed noncontact HR measurement method. In particular, the applicability of the topology algorithm to facial video data has been confirmed for the first time.

Acknowledgments
This work was supported in part by Commissioned Research Fund (Panasonic Corporation), SECOM Science and Technology Foundation, JSPS 19H02155 and 21H03427, JST JPMJPR1873, and JST COI JPMJCE1307.

Ethics Declarations
The study protocol was approved by the Ethics Committee of the University of Occupational and Environmental Health. Written informed consent was obtained from all participants.