Inter-Beat Interval Estimation from Facial Video Based on Reliability of BVP Signals

Yuichiro Maki[1], Yusuke Monno[1], Kazunori Yoshizaki[2], Masayuki Tanaka[1,3], and Masatoshi Okutomi[1]

***Abstract*— Inter-beat interval (IBI) and heart rate variability (HRV) are important cardiac parameters that provide physiological and emotional states of a person. In this paper, we present a framework for accurate IBI and HRV estimation from a facial video based on the reliability of extracted blood volume pulse (BVP) signals. Our framework first extracts candidate BVP signals from randomly sampled multiple face patches. The BVP signals are then assessed based on a reliability metric to select the most reliable BVP signal, from which IBI and HRV are calculated. In experiments, we evaluate three reliability metrics and demonstrate that our framework can estimate IBI and HRV more accurately than a conventional single face region-based framework.**

I. Introduction

Heart rate (HR) and inter-beat interval (IBI) are essential cardiac parameters that provide physiological and emotional states of a person. HR represents the number of heartbeats in a certain time window and is typically described in the form of mean HR for the window. IBI represents a time interval between two successive heartbeats and thus indicates an instant cardiac state in a finer scale than mean HR[1]. Furthermore, many heart rate variability (HRV) parameters can be calculated from beat-to-beat change of IBIs, which provides an important clue of many physiological and emotional conditions such as heart rhythm abnormalities and stress levels. Therefore, accurate measurement of IBI enables a lot of applications such as health monitoring of neonates [1] and prediction of cardiac diseases [2].

HR and IBI are typically measured using an optical contact photoplethysmography (cPPG) sensor that measures a blood volume pulse (BVP) signal derived from the change of blood volume in vessels due to heartbeats [3], [4]. Since the cPPG sensor needs to be attached to human skin, it poses restriction on the subject and precludes many applications that non-contact measurement is preferable or necessary.

To overcome the limitation of cPPG sensors, imaging or remote PPG (iPPG or rPPG), which measures the BVP signal from a facial video in a non-contact manner, has received increasing attention in recent years (see [4]–[6] for a survey). Most of existing iPPG methods focus on mean HR measurement and estimate the most dominant frequency of the BVP signal that corresponds to the mean heartbeats frequency. Previous studies have demonstrated that HR can be robustly estimated from a facial video in relatively stable conditions without large face movements and illumination changes. However, accurate estimation of IBI is still a challenging task even in such stable conditions because it requires accurate estimation of beat-to-beat peak positions of the BVP signal, not only the most dominant frequency.

In this paper, we present a framework for accurate IBI estimation from a facial video. Inspired by recent state-of-the-art iPPG methods (e.g., [7]–[10]), our framework first extracts candidate BVP signals from sampled multiple face patches. The candidate BVP signals are then assessed based on a reliability metric to select the most reliable BVP signal, from which IBI is calculated. Main contributions of this work are summarized as follows.

- While the existing methods based on face patches [7]–[10] mainly focus on mean HR estimation, we present an extended framework for IBI estimation that includes a signal processing pipeline to robustly calculate beat-to-beat peak positions.
- We present a new dataset[2] with high frame-per-second (300-fps) face videos, which demonstrates a better correlation with a reference cPPG sensor than standard 30-fps videos.
- We experimentally evaluate three reliability metrics and demonstrate that our framework based on the reliability metric can significantly improve the IBI estimation accuracy compared with a conventional single face region-based framework.

II. IBI Estimation Framework

Figure 1 presents the overall flow of our IBI estimation framework. In our framework, we apply an algorithm to extract a BVP signal from a pair of temporal intensity traces of the face. Firstly, a pair of two face patches is randomly sampled within the detected face region. This sampling process is repeated until a sufficient number of pairs is acquired. Then, for each pair, a candidate BVP signal

This work was supported by the MIC/SCOPE #141203024.

Y. Maki, Y. Monno, M. Tanaka, and M. Okutomi are with the Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan (e-mail: ymaki@ok.sc.e.titech.ac.jp; ymonno@ok.sc.e.titech.ac.jp; mtanaka@sc.e.titech.ac.jp; mxo@sc.e.titech.ac.jp).

M. Tanaka is also with Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo 135-0064, Japan.

K. Yoshizaki is with Olympus Corporation, Hachioji, Tokyo 192-8512, Japan (e-mail: kazunori_yoshizaki@ot.olympus.co.jp).

[1]Although the terms such as beat-to-beat, instant, and continuous HR can be used interchangeably with IBI, in this paper, we use the term "HR" to represent mean HR for a certain time window unless otherwise noted.

[2]The dataset and the code is publicly available at the following website. http://www.ok.sc.e.titech.ac.jp/res/VitalSensing/

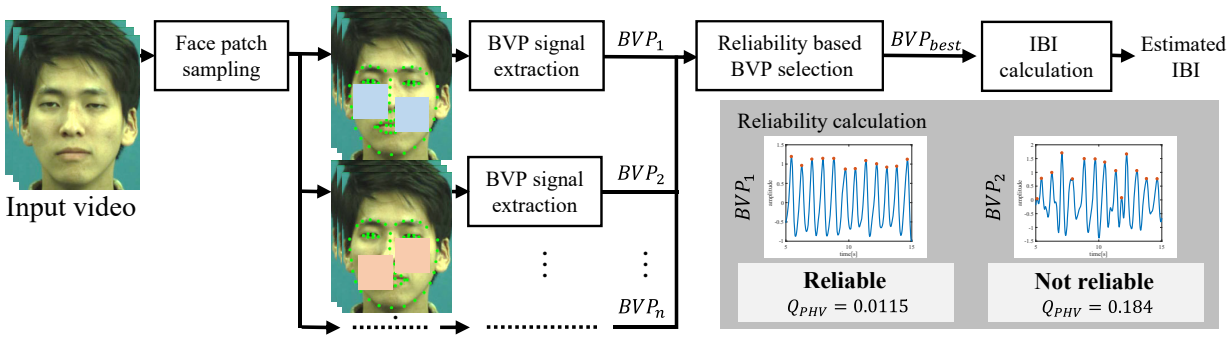


Fig. 1. The overall flow of our IBI estimation framework based on the reliability evaluation of BVP signals.

is extracted. The extracted BVP signals from all pairs are then evaluated based on a reliability metric. Finally, IBI is calculated from the selected most reliable BVP signal. Each step is detailed below.

A. Face patch sampling

To sample the face patches, we follow the random sampling manner in [8]. For each video frame, 66 facial landmarks are firstly detected using the algorithm in [11] (implemented by [12]). Based on the detected landmarks, a pair of two face patches is randomly and repeatedly sampled, as shown in Fig. 1. These patches are tracked between the video frames using the landmark tracking algorithm in [8].

B. BVP signal extraction

For each pair of face patches, one candidate BVP signal is extracted using a common iPPG framework [8], [13], [14]. Firstly, a temporal intensity trace of each patch is calculated using the averaged G channel intensity within each patch region. The pair of traces obtained at the pair of patches is then used as inputs for independent component analysis (ICA) [15]. ICA outputs are two independent components and the one corresponding to the BVP-related signal is selected using the algorithm in [8]. Then, detrending filter [16] and moving average filter are applied to the extracted BVP signal to remove trends and to reduce noise.

Since the outputs of ICA have the ambiguity of plus and minus, the extracted BVP signal may be reversed. Although this is not a significant problem for HR estimation based on the frequency-domain analysis, we need to estimate the correct sign (plus or minus) for IBI estimation that requires to accurately detect beat-to-beat peak positions. For this purpose, we calculate the cross correlations between the extracted BVP signal and the original two temporal intensity traces and regard the sign that provides a higher average correlation as the correct one.

C. Reliability-based BVP signal selection

After extracting the candidate BVP signals from all face patch pairs, we assess the reliability of them to select the most reliable BVP signal. We evaluate three reliability metrics, of which two are existing frequency-based metrics and one is our proposed metric based on the peak height variance. Each metric is detailed below.

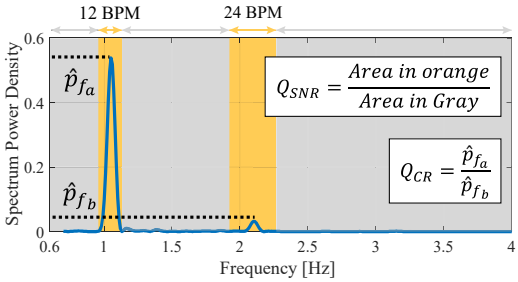


Fig. 2. Illustrative explanation of Q_{CR} and Q_{SNR}.

1) Confidence ratio (CR), Q_{CR}: Lam et al. proposed CR for HR estimation that is based on the spectral power density distribution of the extracted BVP signal [8]. CR evaluates the ratio between the power of the most dominant frequency and that of the second most dominant frequency as (see Fig. 2 for illustrative explanation),

$$Q_{CR} = \hat{P}_{f_a} / \hat{P}_{f_b}, \tag{1}$$

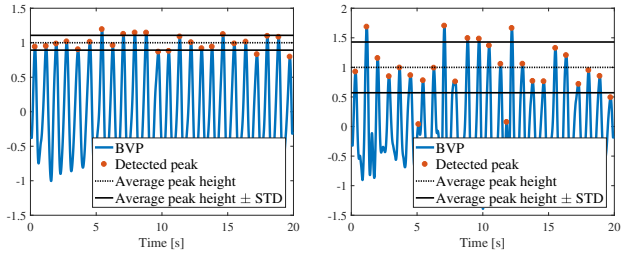
where $\hat{P}_{f_a}$ and $\hat{P}_{f_b}$ denote the spectral power at the most and the second most dominant frequency, f_a and f_b, respectively. CR evaluates the dominance of the heartbeats frequency and higher CR values represent better BVP signals.

2) Signal-to-noise ratio (SNR), Q_{SNR}: SNR is a widely used evaluation metric for the BVP signal (e.g., [7], [17]), which is calculated as,

$$Q_{SNR} = \frac{\int_{f_a-d}^{f_a+d} \hat{P}_f df + \int_{2f_a-2d}^{2f_a+2d} \hat{P}_f df}{\int_{\Omega} \hat{P}_f df - \left(\int_{f_a-d}^{f_a+d} \hat{P}_f df + \int_{2f_a-2d}^{2f_a+2d} \hat{P}_f df \right)}, \tag{2}$$

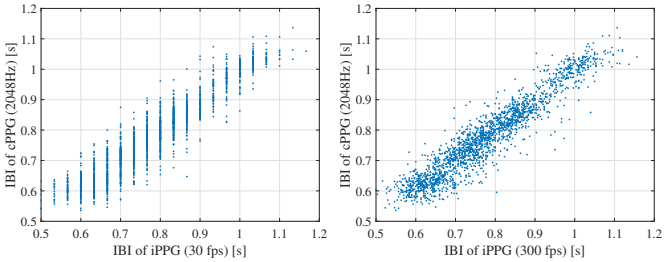
where $\hat{P}_f$ is the spectral power at the frequency f, Ω is the considered frequency range, which is typically set as [0.6Hz, 4Hz] according to human's possible HR, f_a is the estimated most dominant frequency, and d is a parameter that decides the frequency range containing the heartbeats derived frequency. According to [17], we used d = 0.1Hz corresponding to six beat per minute (BPM), as shown in Fig. 2. The second term in numerator considers the harmonic frequency. SNR evaluates the ratio of the spectral power of the heartbeats derived frequency over that of the other frequencies considered as noise. Higher SNR values represent better BVP signals.

3) Peak height variance (PHV), Q_{PHV}: We introduce a new reliability metric that directly evaluates the shape of



(a) BVP signal with low PHV (b) BVP signal with high PHV

Fig. 3. Examples of our Q_{PHV} metric for two BVP signals: (a) and (b). We consider that, if PHV is small as (a), the signal is reliable for IBI estimation. In contrast, if PHV is large as (b), the signal is regarded as not reliable.



(a) Correlation plot for 30 fps (b) Correlation plot for 300 fps

Fig. 4. The frame-rate effect on the IBI estimation. The figures show the correlation plots between the IBIs acquired by cPPG and the IBIs estimated by iPPG using our framework with Q_{PHV}. The figure (a) shows that the 30-fps result does not have enough temporal resolution, while the 300-fps result in (b) shows a better correlation with the reference cPPG sensor.

the BVP signal based on PHV. Figure 3 shows examples of our metric for two BVP signals. We consider that, if PHV is small as Fig. 3(a), the signal is reliable for IBI estimation. In contrast, if PHV is large as Fig. 3(b), the signal is regarded as not reliable. To calculate PHV, the mean of the BVP signal is firstly subtracted from the original signal. Then, peak detection is applied assuming that the minimum beat-to-beat distance is the 0.25 second, which corresponds to 240BPM, and peak height has a non-zero value. The peak detection algorithm is implemented using the MATLAB findpeaks function. After the peak detection, the BVP signal is normalized so that the peak height mean should be one. Finally, PHV (i.e., Q_{PHV}) is calculated using all detected peaks. Lower PHV values represent better BVP signals.

D. IBI calculation

IBI is calculated from the selected most reliable BVP signal. This step contains peak detection, IBI outlier removal, and IBI interpolation, as explained below.

1) Peak detection: The above-mentioned peak detection algorithm is applied to detect beat-to-beat peak positions. Based on the detected peaks, the IBI series is calculated as $IBI_{t_n} = t_n - t_{n-1}$, where t_n is the time of n-th detected peak.

2) Outlier removal: To robustly estimate IBI, we remove outliers that have an IBI value far from the median IBI value in a certain time window. Specifically, if $|IBI_{t_n} - IBI_{median}| > IBI_{median} \times 0.2$ is satisfied, IBI_{t_n} is removed as an outlier.

3) Interpolation: The IBI series $(IBI_{t_1}, IBI_{t_2}, \cdots, IBI_{t_N})$ is then interpolated with enough sampling rate for continuous IBI analysis. We use MATLAB piecewise cubic Hermite interpolating polynomial (PCHIP), which experimentally showed a lower mean IBI estimation error than commonly used spline interpolation.

III. Experimental Results

A. Data collection

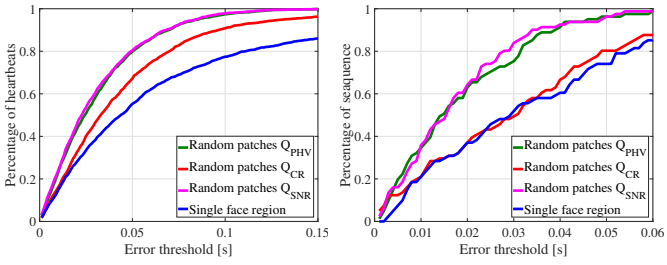
We captured a new high-fps face video dataset, which contains an exercise session to evaluate the IBI estimation accuracy including temporal IBI changes. The experimental protocol was approved by the research ethics committees of Tokyo Institute of Technology and Olympus Corporation. The informed consent was obtained from all subjects before the data collection. Nine subjects with both gender (1 females) and different age (20s - 60s) took part in the experiment. The subjects were asked to sit on a chair, which was placed at a distance of 1.5m from the camera [18]. Each video contains three sessions: relax, exercise, and relax sessions. In the exercise session, the subjects were asked to perform hand grip exercise. The video resolution is VGA (640×480) and the frame rate is 300 fps. The duration of each session is 60 seconds and the total video duration for each subject is 180 seconds. From each video, nine non-overlap 20 seconds sequences were extracted and used for experiments. To acquire reference BVP signals, a cPPG sensor (Procomp Infinity T7500M, Thought Technology Ltd., Canada) was attached to a subject's finger.

B. Frame-rate evaluation

Before the algorithm comparison, we analyze the frame-rate effect on the IBI estimation. Although most of publicly available datasets were recorded at a standard frame rate (e.g., 30 fps), we constructed the 300-fps dataset. Figure 4 shows the correlation plots (total 1873 samples from all sequences) between each beat-to-beat IBI acquired by the cPPG sensor (2048Hz) and the estimated IBI by our framework with the Q_{PHV} metric, where the 30-fps videos were synthesized by averaging every 10 frames of the 300-fps videos. We can clearly see that the 30-fps result does not have enough temporal resolution, while the 300-fps result shows a better correlation with the reference cPPG sensor.

C. Beat-to-beat IBI estimation

We first evaluate the absolute error of each beat-to-beat IBI. To calculate the absolute error, all 1873 IBIs from the reference cPPG sensor were compared with the interpolated IBIs from iPPG at the cPPG sensor's time stamps. Figure 5(a) shows the comparison of our framework with the random patch sampling (500 times) and a standard single face region-based framework [13], where a manually selected cheek region was used as a fixed region of interest and the temporal traces of the RGB channels were used as ICA inputs. The vertical axis represents a percentage of heartbeats whose error is less than the threshold in the horizontal axis. We can confirm that our framework significantly improves the IBI estimation accuracy. Among the three reliability metrics, Q_{SNR} and Q_{PHV} present better performance than Q_{CR}.



(a) Absolute error of IBI (b) Absolute error of RMSSD

Fig. 5. The results of (a) absolute error of IBI and (b) absolute error of RMSSD. The vertical axis represents a percentage of heartbeats/sequences whose IBI/RMSSD error is less than the threshold in the horizontal axis. Our framework with the Q_{SNR} or Q_{PHV} metric provides higher accuracy.

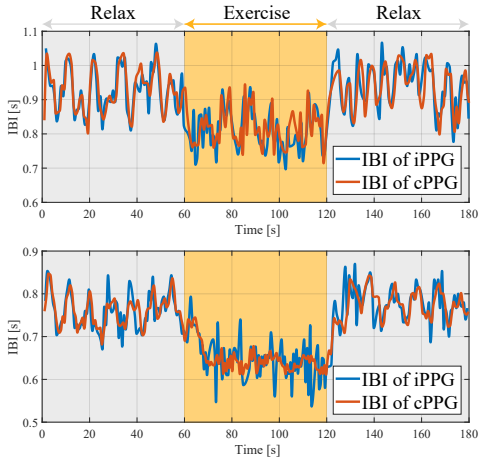


Fig. 6. Continuous IBI results on two subjects. The blue line is the one estimated using our framework with the Q_{PHV} metric, while the red line is the one from the reference cPPG sensor. Our framework can accurately estimate the IBI changes due to the exercise session.

Figure 6 shows the continuous IBI results for two subjects estimated using our framework with the Q_{PHV} metric. We can confirm that the IBI changes due to the exercise session can be accurately estimated using our framework.

D. HRV estimation

We next evaluate the absolute error of root mean square of the successive differences (RMSSD). RMSSD is one of the most common HRV parameters [19]. We calculated RMSSD of iPPG for each 20 seconds sequence only using the detected peaks after the outlier removal. RMSSD of iPPG was then compared with that of cPPG to calculate the absolute error. Figure 5(b) shows the percentage of sequences (total 81 sequences) whose error is less than the threshold in the horizontal axis. We can confirm that the absolute error of RMSSD is significantly reduced using our framework with the Q_{SNR} and our Q_{PHV} metrics.

We also evaluate the absolute error of low-frequency/high-frequency (LF/HF) ratio, which is another HRV parameter. Experimental results showed that our Q_{PHV} metric provides a lower average error (0.564) compared with the average errors when using Q_{CR} (0.821) and Q_{SNR} (0.640).

IV. Conclusion

In this paper, we have presented a framework for accurate IBI and HRV estimation from a facial video. Our framework is based on the reliability evaluation of the BVP signals extracted from randomly sampled face patches. Experimental results on our newly constructed 300-fps dataset demonstrate that our framework can accurately estimate continuous IBI and several HRV parameters.

References

[1] L. A. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. B. Oetomo, and W. Verkruysse, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - A pilot study," *Early Human Development*, vol. 89, no. 12, pp. 943–948, 2013.

[2] J. Kranjec, S. Beguš, G. Geršek, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control*, vol. 13, pp. 102–112, 2014.

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

[4] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, 2016.

[5] A. Sikdar, S. K. Behera, and D. P. Dogra, "Computer-vision-guided human pulse rate estimation: A review," *IEEE Reviews in Biomedical Engineering*, vol. 9, pp. 91–105, 2016.

[6] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali, and F. Mériaudeau, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control*, vol. 38, pp. 346–360, 2017.

[7] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[8] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 3640–3648, 2015.

[9] S. Tulyakov, X. A. Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Seve, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2396–2404, 2016.

[10] S. Kado, Y. Monno, K. Moriwaki, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Remote heart rate measurement from RGB-NIR video based on spatial and spectral face patch selection," *Proc. of Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5676–5680, 2018.

[11] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1867–1874, 2014.

[12] Y. Nirkin, I. Masi, A. T. Trân, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," *Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pp. 98–105, 2018.

[13] M. Z. Poh, D. 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. 10 762–10 774, 2010.

[14] ——, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE Trans. on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[15] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[16] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Trans. on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[17] D. McDuff, E. B. Blackford, and J. R. Estepp, "Fusing partial camera signals for noncontact pulse rate variability measurement," *IEEE Trans. on Biomedical Engineering*, vol. 65, no. 8, pp. 1725–1739, 2018.

[18] Y. Monno, H. Teranaka, K. Yoshizaki, M. Tanaka, and M. Okutom, "Single-sensor RGB-NIR imaging: High-quality system design and prototype implementation," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 497–507, 2019.

[19] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, no. 258, pp. 1–17, 2017.