Real-time heartbeat interval estimation from face video by multiple filters ensemble

Ryotaro SHIMA, SooIn KANG, and Taketoshi MORI
Graduate School of Advanced Engineering, Tokyo University of Science

Introduction / Background of the Study

The demand for telemedicine and remote health management for the elderly is increasing. There are high expectations for the use of small cameras as a method of casually monitoring health in daily life. Simple home health management methods include taking body temperature, observing the swelling and redness of the throat, and measuring the respiratory rate. More recently, measuring weight and body fat percentage, measuring blood pressure upon waking up, and measuring blood oxygen saturation have been adopted. Among these, daily heart rate measurements are becoming particularly important. Continuous heartbeat monitoring may lead to early detection and prediction of cardiac problems and other related diseases.

Common heart rate monitors or electrocardiographs are primarily contact methods for the measurement. In electrocardiographs, electrodes are attached to multiple points on the body. Most of heart rate monitors are designed as wristwatch-type or other body-mounted devices. And recently, heart rate can be measured non-contact and non-invasively by using a webcam. Various systems have been developed to extract the heart rate by capturing video images of the face with a smartphone or webcam [1-4]. If it becomes feasible to measure heart rate using these, the technique could be applied to various fields such as health management, stress management, sports training, and medical diagnosis. In particular, webcams, which are low-cost and widely used at present, could be utilized in health management for the elderly and in telemedicine services.

We would like to realize a system that continuously measures changes in heartbeat intervals over a short period of time in real time. In previous studies, the heart rate per minute is usually displayed by Fourier transform based on the green component from videos lasting several tens of seconds [5]. Difficulties are that the face video contains various noises such as body movements and lighting fluctuations. Moreover, the heart rate must be calculated at any time based on the peak-to-peak time of the heartbeat rather than the average heartbeat over several tens of second to be utilized in various applications. Additionally, the system should support a wide bandwidth of heartbeat periods from heavy intense periods to calm periods to accommodate a wide range of heart rate cycle bandwidths from fast to slow heartbeats.

In this study, based on the assumption that the time it takes to recover the resting heart rate after exercise reflects physical fitness and physical condition of the person, we aimed to develop a system that can measure exercise function based on the recovery of the heart rate after exercise by obtaining the heart rate every second in real time, non-invasively and without restraints, from facial video obtained with a web camera. The objective of this project is to realize a system that can measure exercise function based on the recovery of heart rate after exercise.

Fundamentals and Challenges of Heart Rate Measurement from Facial Video

Heart rate measurement by rPPG (remote photo-plethysmography), which is a low-cost and non-invasive method using a camera, is expected to be used in various fields. However, this method faces challenges due to sensitivity to ambient lighting and subject's movement. Moreover, it is difficult to imagine a situation in which the subject remains still for a long period of time, and the subject's movements may make it impossible to continue capturing the same skin area, resulting in discontinuous data for a short period of time. Therefore, it is necessary to obtain the heart rate from data of as short a time as possible. In particular, to obtain the heart rate with a small delay in real time, it is necessary to extract the peak of the oscillation and calculate the number of beats backward from the peak interval, rather than accumulating data for several tens of seconds and estimating the number of beats by filtering in the frequency domain [6]. In the method of obtaining the heart rate from the peak interval, a minimum of two peaks is all that is required. Thus, it is thought that the heart rate can be obtained from data of a short duration of at least two seconds. When the heart rate signal is extracted from a certain skin area on the face, it is necessary to keep tracking the area in order to capture the temporal changes in skin color. One approach is to detect facial landmarks such as the nose, eyes, mouth, and facial contours for tracking the same skin region consistently. While this method can reduce noise from body movements, it also has the problem of introducing noise due to differences in the detection position of landmarks in each frame, which must be dealt with in the same way as noise due to lighting fluctuations and subject movement. High heart rate data is susceptible to noise due to issues such as the short period of a single wave, and facial motion due to breathing is also significant. Therefore, the simple way is that a narrow bandpass filter must be applied to extract the peak of the heart rate. However, a narrow bandpass filter, as in contradiction, cannot cover a wide range of heartbeats, from low to high heartbeats.

System and Implementation of the Proposed Heart Rate Estimation Method

In this study, we used small web camera (Logitech Streamcam) and obtained facial landmark detection to identify the facial regions on both cheeks from which the heartbeat signal is easily extracted in a 720p resolution facial video at 30 fps, and extracted time-varying data of the average pixel value of green pixels from which the heartbeat signal tends to appear. In contrast, a moving average with a window of 4 frames and a wide bandpass filter from 0.8 Hz to 3.5 Hz were used to remove noise, and 13 narrow bandpass filters with different application ranges were used to extract heartbeats in various heart rate bands from low heart rate to high heart rate to facilitate extraction of peaks in each heart rate band. The bandpass filters ranged from 0Hz to 1Hz, 0.25Hz to 1.25Hz, ..., 2.75Hz to 3.75Hz, and 3Hz to 4Hz, employing 13 bandpass filters shifted by 0.25Hz increments for a 1Hz range. The method was implemented to calculate 13 heart rates from the past 5 seconds of data every second by finding the peaks in the data after applying these filters and determining the average peak interval after removing outliers from that peak interval (see the three figures below). A histogram of heart rates in increments of 15 beats per minute is created based on the maximum past 30 seconds of heart rate data for each filter, the largest cluster is estimated as the current heart rate band by K-means clustering, and the heart rate when the bandpass filter that can extract that heart rate band is used as the current heart rate (Fig. 1).

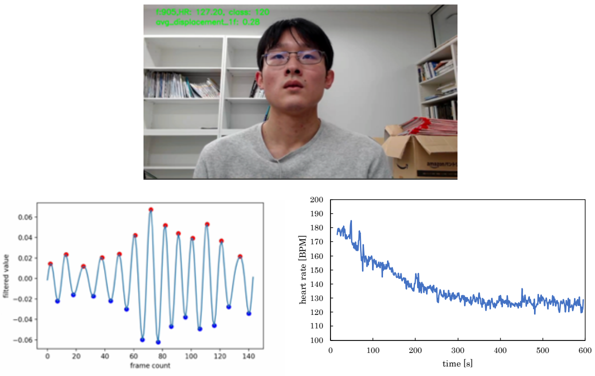


Fig 1. Screen shot of heart rate measurement (above) and output of beats every second (lower left), and heart rate every second after applying the filter (lower right)

Methods and Results of the Verification Experiment

To verify whether the realization system can accurately measure real-time heart rate, the heart rate was compared with the ECG heart rate measured by an electrocardiograph (Checkme ECG ADV, SAN-EI MEDISYS). For video recording, each subject was recorded in three sets: one at rest and two after exercise. Within a 10-minute time period, we performed five consecutive 60-second ECG and video recording measurements, defining this sequence as one set. To determine whether the heart rate changes after exercise could be captured, and whether the heart rate in various heart rate zones could be captured, we acquired the resting and post-exercise heart rates of two subjects. For the validation experiment, we compared the ECG heart rate displayed on the electrocardiograph during 60-second video with the heart rate calculated by our system in the face video for each second. We verified whether the system can accurately measure for two subjects with and without individual differences. Figure 2 shows an example of time variation after exercise for subjects X and Y. Figure 3 shows a scatter plot of the ECG heart rate and the method used in this study. The correlation coefficient between the facial animation and ECG heart rate was 0.753. The correlation coefficient of the post-exercise data alone was 0.93, and that of the resting data alone was 0.041. Outliers were more frequent at low heart rates of less than 90 beats per minute, and fewer at high heart rates of 130 beats per minute or more. We considered that the instability at low heartbeats was due to instability in the identification of the rough frequency region where heartbeats existed by K-means. The histogram of the area where the heart rate band was incorrectly identified showed that small clusters of noise were clustered together, creating clusters larger than the current heart rate cluster. We considered that this was caused by the width of the histogram bins being too wide, so we changed the bin width from 15 beats per minute to 10 beats per minute and conducted an experiment with the same data. The adjustment improved the correlation coefficient from 0.753 to 0.831.

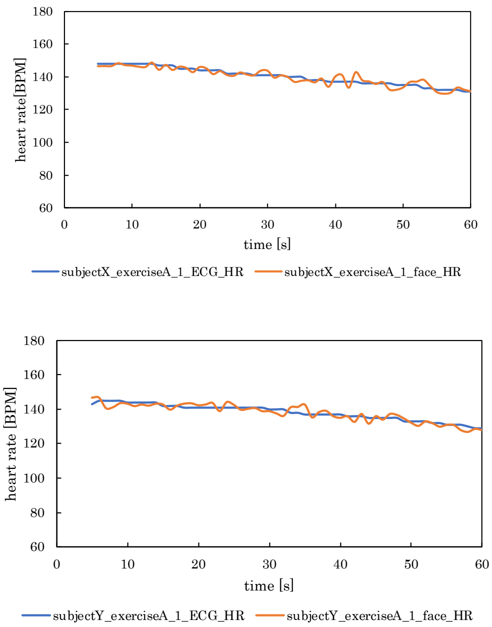


Fig 2. Comparison of real-time heart rate measurements of subjects X (above) and Y (below)

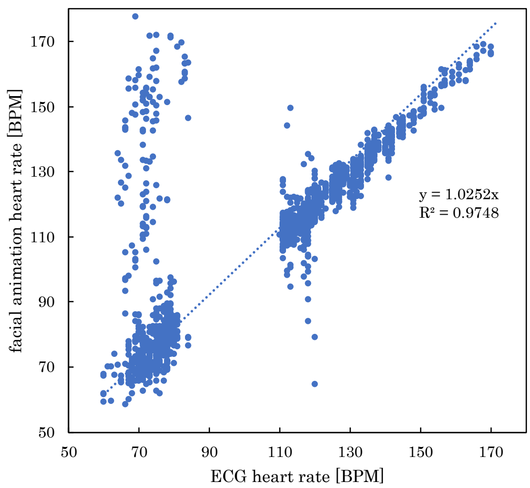
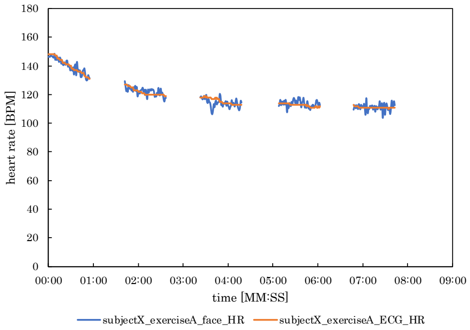


Fig 3. Scatter plots of ECG heart rate and facial animation

Our developed system successfully followed the change in the heart rate of each subject after exercise as it approached the resting state (see figure below). Subjects X and Y showed differences in heart rate recovery time. However, the exercise load of the two subjects was not controlled, and it could not be verified whether there was a difference in physical recovery ability.



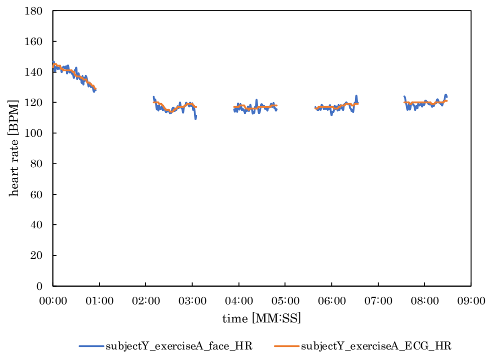


Fig 4. Tracking of changes in heart rate recovery time for subjects X (above) and Y (below)

Discussion

The system developed in this study still has an issue where even after changing the bin width of the histogram, the estimation accuracy for low heart rates remains low, resulting in the calculation of values that are clearly higher than the correct heart rate. To solve this issue, one idea is to first roughly distinguish between low and high heart rates by applying only a wide range of bandpass filters. Additionally, there is a possibility that the width of the bandpass filter was set to increase the estimation accuracy of high heart rates, so there is a chance for improvement by reviewing and tuning the system parameters.

This study was conducted based on the hypothesis that exercise function can be measured from the recovery time of heart rate after exercise. By using this system, it was possible to track the recovery of heart rate after exercise and confirm differences in heart rate recovery time. However, it has not been verified and is unclear whether it is actually possible to examine exercise function from the time it takes for the heart rate to recover and settle down to a certain extent after exercise. Therefore, it remains a future task to investigate whether there is a relationship between the heart rate recovery time after exercise and exercise function by aligning the exercise load of each subject.

Conclusion

In this study, we achieved real-time peak-to-peak heartbeat interval estimation by combining three methods: (1) a method to limit the heartbeat information acquisition area by extracting feature points from a face image and estimating the face shape, (2) a method to narrow the heartbeat interval range by identifying the cycle bands as a broad mass by clustering, and (3) a method to calculate the heartbeat interval sequentially by tracing the heartbeat change. In an experiment using a 30fps 720p web camera, a high correlation of 0.753 was obtained in comparison with a portable electrocardiograph. The correlation coefficient increased to 0.831 by changing the width of histogram bins from 15 beats per minute increments to 10 beats per minute increments.

Increased facial movement and changes in ambient light will result in more noise mixed into the heart rate signal; thus, a higher frame rate, more bit depth, and better resolution of camera image are expected to provide more accurate and frequent heart rate measurements.

Reference

[1] Sebastian Zaunseder, Alexander Trumpp, Daniel Wedekind, Hagen Malberg, "Cardiovascular assessment by imaging photoplethysmography – a review", Biomed. Eng.-Biomed. Tech., vol. 63, no. 5, pp. 617-634, 2018.

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

[3] Xun Chen, Juan Cheng, Rencheng Song, Yu Liu, Rabab Ward, Z. Jane Wang, "Video-Based Heart Rate Measurement: Recent Advances and Future Prospects", IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 10, pp. 3600-3615, 2019.

[4] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, Julien Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," Pattern Recognition Letters, vol. 124, pp. 82-90, 2019,
UBFC-RPPG Dataset
(https://sites.google.com/view/ybenezeth/ubfcrppg).

[5] Wim Verkruysse, Lars O Svaasand, J Stuart Nelson, "Remote plethysmographic imaging using ambient light", Opt Express, vol. 16, no. 26, pp. 21434-21445, 2008.

[6] Yuichiro Maki, Yusuke Monno, Kazunori Yoshizaki, Masayuki Tanaka, Masatoshi Okutomi, "Inter-Beat Interval Estimation from Facial Video Based on Reliability of BVP Signals", 2019