An exploration of a heart rate sensing garment solution based on rPPG technology

Meng Liu$^{1,2}$, Chuanjiang Wang$^{1,2,*}$, Ningning Wang$^{1,2}$, Xiujuan Sun$^{2}$

$^{1}$Robot Research Center, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China
$^{2}$College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China

*Corresponding author’s e-mail: cxjwang@sdust.edu.cn

Abstract. In order to solve the problem of heart rate detection during exercise, this article proposes a solution that combines ordinary clothing with non-contact heart rate measurement technology. The scheme integrates the camera into the shoulder of the jacket, which enables the camera to monitor the neck position in real time and convert the detected neck video image into the corresponding heart rate by the technique of optimized rPPG. In this regard, some experiments designed to verify the accuracy of heart rate detection in resting and exercise states, respectively. The experimental results were compared with a commercially available heart rate oximeter, and the comparative results verified the feasibility of the study.

1. Introduction

Real-time monitoring of heart rate has a major role to play in sports health and performance. By monitoring the heart rate, it can prevent excessive heart rate during exercise and also guide the intensity of exercise according to the heart rate. In professional training, heart rate has always been an important indicator to improve the level of exercise. Today, the products for sports heart rate detection on the market are divided into two main categories: electrode based ECG detection products, including heart rate bands [1] and heart rate sensing garments [2] and optical volume technology based detection devices, including heart rate watches (bracelet type products), heart rate armbands. Conventional electrode sensors are the most accurate in heart rate sensing. The principle is to use electrodes to measure the micro-voltage signal generated by the beating heart, and the signal is amplified to obtain the heart rate value. However, due to the independence of the sensor, heart rate bands and bracelets are prone to out of position or even fall-off in excessive exercise. More, many athletes feel uncomfortable with the painful rubbing sensation of the heart rate band on their skin during exercise. The heart rate sensor suit solves the pain of using a belt sensor and combines the high measurement accuracy of a heart rate belt with the relatively high price.

These traditional methods of heart rate detection are contact-based. In recent years, non-contact heart rate measurement has played an increasingly important role [3], and how to combine non-contact technology with smart clothing technology is a new field. This paper attempts to combine the two by integrating an ordinary camera into the shoulder of the garment, monitoring the person's neck in real time, and recovering the heart rate from the monitored image using the rPPG technique.
2. General Procedure for Video Heart Rate Detection

2.1. Timing signal acquisition

The blood flow in the blood vessels under the skin of the body changes periodically as the heart systolic-diastolic. This variation causes small differences in the absorption of the light spectrum by human skin, which are not visible to the naked eye but can be detected on video by image processing [4]. The detected video frame sequence is taken as a profile either horizontally or vertically by detecting a skin region (let’s say the neck region) of a part of the human body through a camera as shown in Fig.1. The trend of the pixel values over time can be seen on the profile, and from this trend, the heart rate can be extracted.

![Fig.1 Video slicing and timing of each channel](image)

A common temporal extraction method is to detect the skin regions contained in the video frame first, next select some or all the skin regions as regions of interest (ROI), and then synthesize a temporal signal in the time domain by averaging the skin pixels contained in the ROI in each frame [3]. In this paper, we adjust the distance between the camera and the neck as well as the focal length so that the video image is filled with human skin areas and directly obtains the skin tone timing signal by averaging the whole image. The timing diagram of the extracted video RGB channels after averaging in the airspace is shown in Fig.1.

2.2. Heart rate extraction method selection

Once the timing of the ROI is obtained, the signal sequence where the heart rate is located needs to be extracted by removing the interference and noise from the original timing signal through signal analysis. As technology and research continue to evolve, a number of methods have been used to recover heart rate signals through skin timing. This includes methods based on the skin reflection model [5], the methods of blind source separation include the method of ICA [6] and the method of PCA [7], the method of wavelet decomposition and reconstruction [8], and the method of EMD decomposition and reconstruction [9]. The method based on the skin reflection model requires the creation of a reflection model of the skin as well as the calculation of the various factors that influence it, blind source separation is performed by timing the three channels, but it is not possible to determine which component of the separation contains the largest heart rate. Further periodic detection is required, but this method will fail when there is similar periodic activity in the human body. The wavelet decomposition reconstruction allows multi-scale decomposition and time-frequency analysis, but it requires the selection of a certain wavelet base. The choice of the wavelet base has a great impact on the results of the entire wavelet analysis, once the wavelet base is determined. It will not be able to be replaced during the entire analysis, even though the wavelet base may be optimal globally, but in some localities, it may not be, so the base function of the wavelet analysis lacks adaptability. What's great about the frequency-domain analysis method with EMD is that for a segment of unknown signal, you can just start breaking it down without doing any prior analysis and research. This method will automatically divide it according to some solid modes in a hierarchical manner without the need for human setup and intervention. Based on the above considerations, this paper chooses the EMD decomposition and reconstruction method to decompose the original timing signal. Select the mean timing $S(t)$ of the green channel containing the maximum heart rate signal, according to the EMD decomposition principle:

$$S(t)=\sum_{i=1}^{n} \text{imf}_i(t) + r_n(t)$$  (1)
3. Optimized Heart Rate Detection Process

3.1. Model building and interference analysis

As can be seen from the results of the EMD decomposition in Fig. 2, the raw timing signal is mostly energy concentrated in the lower frequency band, as shown in the decomposition plot for \( \text{imf}_2(t) \) and in the residual component, which is mainly determined by the reflective and interference properties of the skin. Based on the modelling approach described in the literature [5], we create the following model:

\[
S(x, y) = (L_v + L_s)(P_m + P_s + P_f) + v_n
\]  
(2)

where \( S(x, y) \) is the pixel value of the \((x, y)\) position in the frame captured by the camera, \( L \) stands for light intensity, \( L_v \) is the part of light intensity that varies and \( L_s \) is the part of light intensity that is stable. \( P_m, P_s \) and \( P_f \) represent the coefficient of the effect of exercise on light, the coefficient of the effect of blood volume changes on light, and the specular reflection coefficient of the skin, respectively. In a stable environment, or changing light environment for a certain period of time, the light intensity can be approximated as stable and constant, i.e. \( L_v \approx 0 \). The sample quantization noise from the camera sensor can be eliminated by averaging the pixel timings over the entire frame, i.e., \( v_n \approx 0 \), where the more pixel timings involved in the calculation, the better. Based on the above considerations, equation (2) can be reduced to the following form:

\[
S(x, y) = L_s(P_m + P_s + P_f)
\]  
(3)

Equation (3) represents the process by which light intensity is sampled by the camera sensor through three different modes of modulation. where \( L_s \times P_f \) represents the specular reflection of the skin and is the DC-stabilized portion of the light reflection, which can be eliminated by de-biasing the signal, thus Eq. (2) can be rewritten as:

\[
S_v = L_sP_m + L_sP_s
\]  
(4)

where \( S_p \) represents the portion of the image that changes, \( L_s \times P_s \) represents the effect produced by heart rate, and \( L_s \times P_m \) represents the effect produced by movement. And \( P_m = \{ P_{m1} + P_{m2} + P_{m3} + \cdots \} \), \( P_{ml} \) represents modes of different movements. From our model it is clear that the influence of exercise is the main and important influence in the heart rate detection task, which coincides with the
rest of the literature. How to recover heart rate from exercise, especially cyclical exercise in the same frequency band with the heart rate, is a difficult task.

3.2. Time-series clustering to optimize ROI selection
The temporal sequence of pixel changes in an image containing a skin region is constructed in the time domain based on the pixel values at each point in the image, and the signal that best characterizes the heart rate is isolated from this. The traditional method of averaging over the entire skin area has limitations, firstly, different skin areas contain different melanin values, and specific areas of the skin such as moles, patches, hair, jewelry, etc. can be affected, especially when framing the neck area. We therefore thought of using a grid to compartmentalize skin regions, and then averaged within each sub-grid to construct a grid temporal sequence over the time domain. And the distance measure between the grid timings, the nearest neighbor distance diagram between the grid timings is shown in Fig. 3(a), and the different colors represent different distances. And the spatial distribution of grid temporal sequences is shown in Fig. 3(b). Based on the characteristics of the spatial distribution of grid timings, we choose the K-means clustering algorithm to classify the pixel timings and take the result of the classification as the ROI.

3.3. GSS filtering to remove transient interference
In addition to the interference of periodic movements, some sudden movements can also affect the extraction of the heart rate signal. For example, an unexpected head turn affects the neck area by pulling on it, and a swallowing motion affects the neck part. The effects caused by such movements will not be spread over the entire time domain process, and may even be in a tiny time period only, but such transient movements will affect the entire skin region, i.e., they cannot be filtered out by the previous step of temporal clustering. We refer to the method of GSS in the literature [10] for segmentation of the temporal sequence, and smoothly filter the entire temporal sequence by re-clustering the temporal fragments and replacing the distal fragments with centroid-like fragments.

In summary, we get a complete processing flow after optimization: after obtaining the grid timing sequence of the G-channel, we cluster it to remove the sequences that are far away from the class center. The remaining sequences are averaged to obtain the average timing of the G-channel, and then de-trended, standardized and band-pass filtered to remove the frequency band outside the heart rate range [0.5-3.5Hz,30-210BPM]. EMD decomposition was then used to time domain GSS filtering of the decomposed \( imf_i \), and the signal was reconstructed after filtering, i.e., the finally obtained heart rate signal.
4. Experiments and Results
This experiment designed how to integrate the camera into the jacket and complete the reading of the video of the neck. This experiment uses a lens + processing chip solution, and the processing chip is powered by a battery. The lens and chip are connected by a freely bendable material between the lens and the chip, and the lens orientation can be changed at will to select an area best suited for detection depending on the experimental object. The parameters of the video shot in the experiment is 18fps, each frame image size 1280 * 720 pixels, the experiment intercepted 30 seconds of video footage for analysis. The video footage collected by the camera is transmitted to the video processing chip for processing, and can be stored locally or transmitted to other devices via Wi-Fi. In this experiment, the video is transmitted to an Android phone via Wi-Fi, and the same phone is used to capture the signal from a heart rate oximeter via Bluetooth module, and compare the two signals. As the signal comparison and reference of this experiment, we chose a finger clip pulse oximeter for clinical diagnosis in the market, instrument model YK-83C, this instrument has the advantages of small size, low power consumption and simple operation, not only heart rate but also can be used to measure blood oxygen saturation. This experiment used only the function of the instrument to measure heart rate, the measurement range 30BPM-240BPM, measurement accuracy ±1BPM. Experimental device selection and camera integration scheme is shown in Fig. 5.

The experiment was conducted by 10 volunteers who knew the purpose of the experiment beforehand, the volunteers were 24-32 years of age and included 2 women and 8 men. The heart rate of the subjects was measured at rest and in exercise, and the results of the heart rate recovered from the camera were compared with the results of the oximeter.
We performed a Bland-Altman analysis of the experimental results. Bland-Altman analysis is a commonly used method to perform consistency analysis of two sets of data, and the results are shown in Fig. 6. From Figure (a), it can be seen that the data are relatively concentrated in the resting state, with an average error of 0.9 BPM and most of the data concentrated within the 95% confidence interval. In the motion detection experiment, we carried out some routine exercise, including fast walking, jogging and jumping in place, etc. The results of the motion state are shown in Figure (b), because the influence of motion is the main in our model, especially the cyclical motion has a large influence on the experimental results. This influence is also reflected in the experimental results, the data are more scattered, the average error increased to 4.1 BPM.

5. Conclusion

We explored a solution for integrating cameras into a shirt for real-time monitoring of heart rate, from the experimental results, it can be seen that the shoulder camera based solution for heart rate detection is feasible, wearing clothing with an integrated shoulder camera can detect the heart rate both at rest and in motion, and a comparison with professional heart rate measuring equipment proves that the accuracy of the detection results is also within the prediction range.

Acknowledgments

This work was supported by Shandong Province Key Research and Development Plan (No.2016GSF201197 & No. 2017CXGC0919).

References

[1] Adnane Mourad, Jiang Zhongwei, et al. Development of QRS detection algorithm designed for wearable cardiorespiratory system[J]. Computer Methods and Programs in Biomedicine. 2009, 93(1): 20-31.
[2] Lee Young-Dong, Chung Wan-Young. Wireless Sensor Network Based Wearable Smart Shirt for Ubiquitous Health and Activity Monitoring[J]. Sensors and Actuators B: Chemical. 2009, 140(2): 390-395.
[3] Chen Xun, Cheng Juan, Song, et al. Video-Based Heart Rate Measurement: Recent Advances and Future Prospects[J]. IEEE Transactions on Instrumentation and Measurement. 2019, 68(10): 3600-3615.
[4] Wu Hao-Yu, Michael Rubinstein, et al. Eulerian Video Magnification for Revealing Subtle Changes in the World[J]. ACM Transactions on Graphics. 2012, 31(4): 1-8.
[5] Wang Wenchun, Brinker Bert den, et al. Algorithmic Principles of Remote PPG[J]. IEEE transactions on bio-medical engineering. 2017, 64(7):1479-1491.
[6] Poh Ming-Zher, McDuff Daniel J, et al. Non-contact, automated cardiac pulse measurements
using video imaging and blind source separation[J]. Optics Express, 2010, 18(10):10762–10774.

[7] Magdalena Lewandowska, Jacek Ruminski, et al. Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity[C]. 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), Szczecin, 2011:405-410.

[8] Bousefsaf Frederic, Maaoui Choubeila. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate[J]. Biomedical Signal Processing and Control, 2013, 8(6):568-574.

[9] Zhang Qi, Zhou Yimin, et al. Heart Rate Extraction Based on Near-Infrared Camera: Towards Driver State Monitoring[J]. IEEE Access, 2018, 6:33076-33087.

[10] Chen Jie, Chang Zhuoqing, et al. RealSense = real heart rate: Illumination invariant heart rate estimation from videos[C]. 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2016: 1–6.