Video-based heart rate measurement via the anti-shake remote photoplethysmography (rPPG) pulse extraction

Fangyuan Lyu, Yuting Zhu, Shaolong Chen, Luping Wang*

Department of Electronic and Communication Engineering, Faculty of Electronic and Communication Engineering, Sun Yat-Sen University, Shenzhen, China

ABSTRACT

The prevalence of cardiovascular diseases in China is still on the rise, and it is estimated that 330 million people are suffering from cardiovascular diseases. In terms of physical health and mental health, a heart rate detection technology that is portable can be measured at any time, safe, comfortable, simple to operate, and low-cost is essential. Given the inevitable jitter of handheld mobile phones, based on the accuracy of remote photoplethysmography (rPPG) detection technology, we propose a moving window timing sampling to refine the original video frame signal. The heart rate value can be extracted by processing such as region of interest (ROI) selection, high-pass filtering, blind source separation algorithm, Fourier transforms, and peak detection. Compared with the heart rate value obtained without using the moving window timing sampling, we found that the effect of the moving window timing is about 10 seconds is the best. The root means as the square error (RMSE), mean absolute error (MAE), and standard deviation (STD) are the lowest, 6.6929, 5.1365, 6.6165 respectively. The errors compared to the sampling without moving piecewise function are 13.53, 10.79, 14.09, The errors were reduced by 50.5%, 52.4%, and 53.04% respectively.

Keywords: Heart rate, remote photoplethysmography (rPPG), moving window

1. INTRODUCTION

The guidelines of *Report on Cardiovascular Diseases in China 2020* pointed out that the prevalence of cardiovascular disease in China is still on the rise, and it is estimated that the number of cardiovascular disease patients is 330 million[1]. *China Health and Retirement Longitudinal Study* assessment the depressive symptoms of 6810 residents without cardiovascular disease were found, compared with those without any depressive symptoms, persistent depressive symptoms, cardiovascular risk and death risk. The increase is significantly correlated[2]. The elevated resting heart rate (RHR) (>90 bpm) increased the risk of cardiovascular disease among the 35-75 group[3]. Therefore, in human life and work, a heart rate detection technology that is portable can be measured at any time, safe, comfortable, simple to operate, and low-cost is indispensable.

At present, common heart rate detection technologies are divided into two categories, one is contact heart rate measurement, and the other is non-contact heart rate measurement. Contact heart rate measurement has electrode type electrocardiogram detection technology, photoplethysmography (PPG) detection technology; non-contact heart rate measurement methods include thermal infrared imaging detection technology[4], doppler radar detection technology, imaging ballistocardiography (iBCG) detection technology, rPPG detection technology[5, 6].

Electrode type electrocardiogram detection technology is to detect the heart rate value by sensing the heartbeat current of the human body through the electrode sheet directly in contact with the skin, which is inconvenient to carry and has poor comfort. PPG is a method to measure the change of blood volume in blood vessels by the photoelectric method. The change of light intensity reflects the contraction of the heart. When the light intensity is the largest, the arterial blood volume reaches the lowest and the heart relaxes. When the light intensity is the smallest, the arterial blood volume reaches the highest, and the heart contracts[7]. Diastole and contraction of the heart is a heartbeat, and an ascending and descending branch of the pulse wave represents a heartbeat[8].

The principle of thermal infrared imaging detection technology is based on the fact that the temperature of natural objects is higher than absolute zero, which will produce infrared radiation[4, 9]. Lu et al.[10] and Matsunaga et al.[11] used to

*wanglp27@mail.sysu.edu.cn

International Conference on Computer Application and Information Security (ICCAIS 2021), edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 122600L · © The Authors. Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2637576

filter the heart rate signal detected by Doppler radar. However, the long-term detection of radar will cause adverse effects on human health and is not suitable for daily use. The iBCG technology is a video-based non-contact technology that can detect the heart rate caused by the weak mechanical head movement caused by the heartbeat[12,13]. Due to the jitter of the head itself, the jitter is caused by the heartbeat. And the errors caused by facial expressions are still relatively large. It is not as robust as rPPG technology.

In 2011, Poh et al. proposed a new method called rPPG technology to calculate heart rate from facial video[14]. The first processed the video frame and divided it into R channel, G channel, and B channel, and used a method based on smoothing before filtering each channel[15], then normalized each signal, and Blind source analysis shows that three independent signal sources are linearly combined to form the RGB channel signal, to eliminate the interference of light changes on the pulse wave signal. In 2014, Li et al.[16] segmented the background area in the face video, using the average value of the G channel of the background area in each frame as the background signal, and then combined the video face. The background signal is subtracted from the G channel signal in the range as a pulse signal.

Jonathan and Leahy demonstrated for the first time the feasibility of detecting pulse waves through mobile phone cameras[17,18]. In the rPPG system based on the built-in camera of the mobile phone, the built-in camera of the mobile phone is used as the optical sensor, and the built-in LED light of the mobile phone is used as the light source. When the user covers both the camera and the light source with the fingertip, the camera can collect the rPPG signal of the fingertip. However, in addition to heart rate detection, the portrait video through the mobile phone camera has further functions such as attention judgment and so on. In recent years, many methods have focused on designing the effective region of interest extraction of input face videos[19,20] and using various deep learning networks[21,22] to perform physiological measurements.

Next, we will do anti-jitter processing and heart rate extraction of mobile phone video signal based on rPPG technology.

2. RESEARCH FRAMEWORK

As shown in Figure 1, the total technology roadmap shows that the acquisition of heart rate values is divided into the following parts: initial signal acquisition module, moving window timing sampling, initial signal preprocessing, blind source separation algorithm, heart rate calculation module, and then these modules will be processed Elaborate in detail.

2.1 Initial signal acquisition module

The initial signal acquisition module includes video acquisition time of *N*, video processing, face recognition, corner tracks, frame-by-frame interest area extraction, frame-by-frame interest area R, G, B channel average calculation, and removal of the first *t* seconds due to operating the mobile phone The shaking data is generated, and finally, we get the initial value signal.

The video acquisition duration is *N* means that the user captures a video with a duration of *N* and a frame rate of *F*, and uses the front camera of the mobile phone to take a picture of the face; the face is about 30 cm away from the mobile phone.

Video processing, to extract the required RGB three-channel data, an *N*-second video is decomposed into video frames whose total number is the product of *N* and *F*, and then a simple JPEG codec ffjpeg library is used to realize the conversion of video frames from NV21 to RGB24 code.

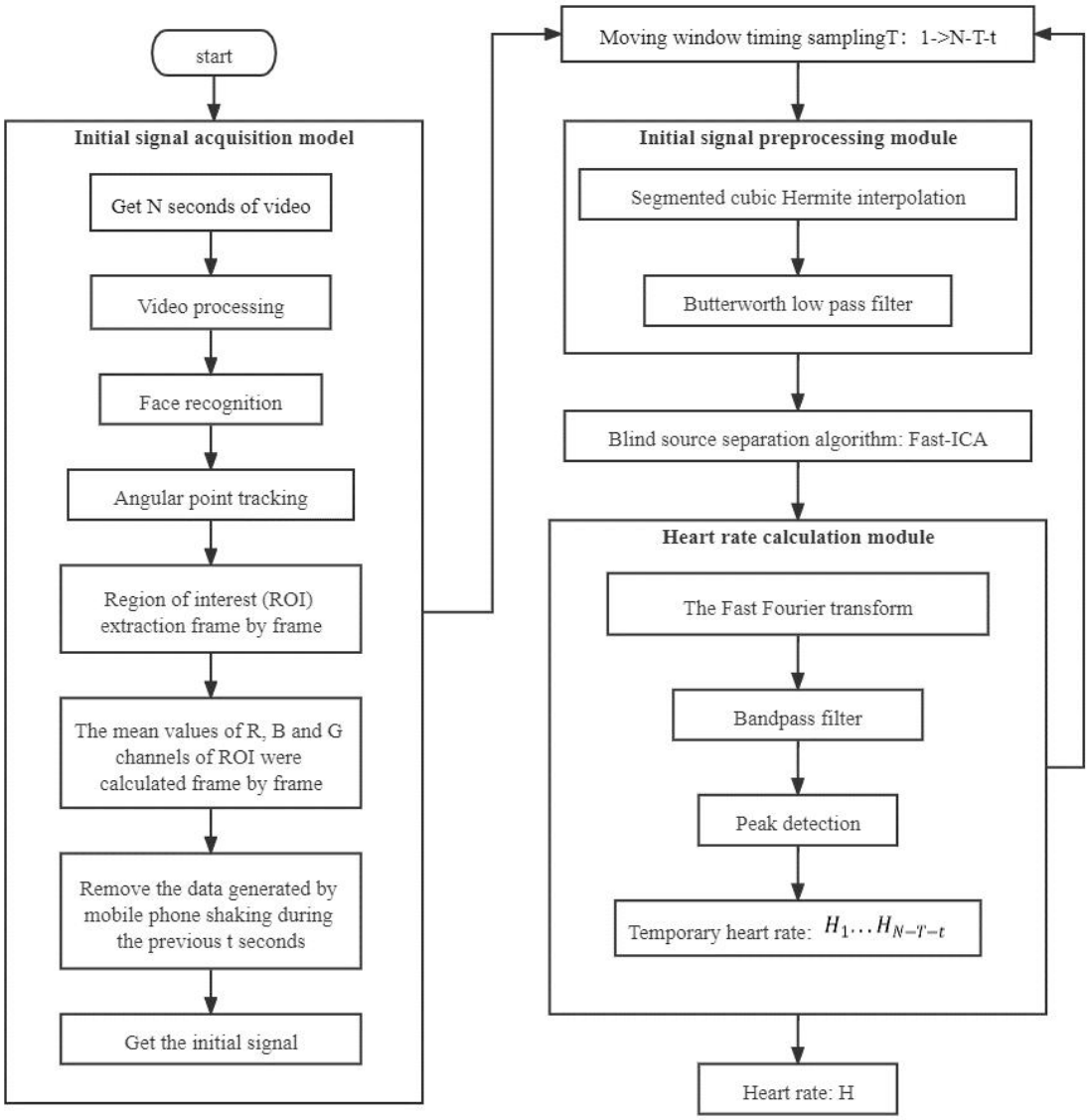

Figure 1. The total technology roadmap.

Face recognition uses face dynamic registration technology, which refers to face detection and face tracking. The face is tracked after the face is detected. It is to track the movement of the face instead of detecting it every frame. The face dynamic registration technology is mainly divided into two modules, the face detection module, and the face tracking module.

There are two main methods for ROI selection: the whole face area selection method and the partial area selection method. The local area selection has high computational complexity, there is a causal dilemma caused by the interdependence of skin detection and pulse waveform signal extraction, and the requirements for Spatio-temporal coherent local segmentation[23]. We choose the whole face area selection method. The whole face area selection method refers to selecting the entire face or the face area except the eyes. For example, Li[16] selects the face area except for the eyes. This method is the simplest ROI selection. The amount of calculation is relatively small.

The data of shaking caused by the operation of the mobile phone in the previous t seconds is removed because the error in the first two seconds is particularly large after calculation. The experimenter, with one arm tied to an Omron sphygmomanometer, and holding the phone in the other hand, when recording a video, you need to click the start button in the middle of the phone screen and find a comfortable angle that is 30 cm away from the phone and stay still at the

beginning. An adaptation process is required. Checking the video recorded by the experimenter found that this adaptation process can be controlled within 2 seconds.

2.2 Moving window timing sampling

To reduce the root mean square error (RMSE), mean absolute error (MAE), and standard deviation (STD), the initial signal is intercepted by the piecewise function of different periods. For example, a 20-second video can get an 18-second initial signal, and because the experimenter's hand-held mobile phone shakes more than a fixed camera. In this case, the longer the initial signal duration is, it is not conducive to the improvement of the accuracy of the heart rate value. The sampling time of the moving window sequence is too short, and the error of extracting the heart rate value is too large. This is that we sample T seconds through the moving window timing, and each time the moving window timing sampling is shifted to the right for 1 second, and these sampled signals are sent to further processing one by one, then we finally measure the heart rate value of *N*-*T*-*t*, that is, the temporary value of the heart rate$H_1 \ldots H_{N-T-t}$. Finally, the mean value of the temporary heart rate value is extracted as the heart rate value H.

2.3 Initial signal preprocessing module

The initial signal preprocessing module is divided into the initial signal segmented cubic Hermite interpolation and Butterworth low-pass filter.

Piecewise cubic Hermite interpolation. The signal is sampled with a sampling frequency of approximately twice the frame rate, and the segmented cubic Hermite interpolation method is used here.

Butterworth low-pass filter. After obtaining the PPG signal with the DC component removed, we need to digitally filter the PPG signal to remove low-frequency and high-frequency noise. 0.75-3.66 Hz. Because the frame rate of mobile phone recording is generally 25 or 30 frames per second, the sampling frequency is twice the frame rate. Through calculation and actual measurement, a sixth-order Butterworth low-pass filter with a normalized cut-off frequency of 0.04 is selected here. Because the Butterworth filter may cause extremely serious phase loss, we will adjust the phase of the signal after the Butterworth filter is processed. Realize zero-phase digital filtering by processing the input data forward and backward.

2.4 Blind source separation algorithm

In recent years, several core rPPG methods for extracting impulse signals from the video have been proposed. These include the following: blind source separation (BSS) (for example, principal component analysis)[24], based on independent component analysis (ICA)[25] and matrix completion[26]), which use different standards to combine. The temporal RGB trajectories are separated into uncorrelated or independent signals to retrieve the source of the pulses; CHROM[27], which linearly combines the chrominance signals by assuming a standardized skin color to white balance the image; PBV[28], which uses blood at different wavelengths The feature of volume change clearly distinguishes the color change caused by the pulse and the motion noise in the RGB measurement; 2SR[29], which measures the temporal rotation of the subspace of the skin pixel space used for pulse extraction.

Blind source separation algorithms are still the mainstream. Blind source separation refers to the analysis of unobserved original signals from multiple observed mixed signals. Among many ICA algorithms, the fixed point algorithm also called Fast-ICA is widely used in the field of signal processing because of its fast convergence speed and good separation effect. The algorithm can well estimate the original signals that are statistically independent and mixed by unknown factors from the observed signals.

2.5 Heart rate calculation module

The heart rate calculation module includes fast Fourier transform, filter, and peak detection.

The first step is to perform a fast Fourier transform on the signal, and the second step is to filter the frequency signal with a bandpass filter of 0.75-3.66 Hz, and find the value with the largest power spectral density in the frequency range of the signal spectrum signal. The frequency corresponding to the number of points, and then multiply the frequency by 60 to get the experimenter's heart rate value.

As for a normal person, the change of the heart rate value is not a sudden change process, so for the heart rate value determined in the adjacent time window, there will not be too much change, but there may be heart rate in the adjacent time window. A situation has changed too much. Considering this situation, we adopted moving window timing sampling

to shorten the sample time for each calculation and measured multiple times to reduce the influence of changes in the heart rate value on the experimental results.

2.6 Extract heart rate value

Average the temporary heart rate $H_1 \ldots H_{N-T-t}$ area and the final average of the temporary heart rate value is the heart rate value H.

3. DATA DESCRIPTION

As shown in the video capture demo in Figure 2, the experimenter wears an Omron medical sphygmomanometer in one hand and holds a mobile phone in the other hand, and tries to keep shooting as still as possible. In order to reduce the impact caused by the wide-angle distortion of the front camera, it has been tested that the face is about 30 cm away from the mobile phone, and the result is the best. The lens distortion is greater if the face is too close, and the pixel area occupied by the face in the video is reduced when the face is too far apart. Obtaining physiological indicators. Finally, it is worth noting that the mobile phone camera software should turn off the filtered beauty, turn off the high dynamic range imaging (High-Dynamic Range, HDR), turn off the white balance, and other functions that will process the original data.

Figure 2. Figure with the video capture demo.

Through 4 experimenters (3 males and 1 female, aged between 20-30 years old), a total of 13 sets of video data were obtained. The length of each video is about 20 seconds.

4. EXPERIMENT AND DISCUSSION

4.1 Initial signal acquisition module experiment

Through the experimenter shooting a video of about 20 seconds, we proceed to the next processing of the video.

In this experiment, the skin detection partition ratio is 1, and the face detection frame is rectangular. The small white dots in the picture on the right are feature points. It can be seen that the edges of the face are well detected, and the portrait information can be extracted accurately to help us in the subsequent processing and application.

The human jitter or the brain is always slightly tilted, and the two eyes are not on the same horizontal line. At this time, we calculate the tilt angle of the two eyes and use a rotation function to rotate the rectangular frame of the eyes to completely cover the eyes area. As shown in the area of interest shown in Figure 3, it can be seen that the area of interest only contains the part of the human face without mixing background information.

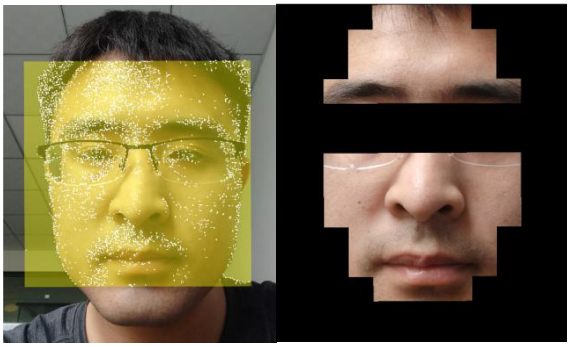

Figure 3. Figure with the area of interest. The right image is the result of region of interest extraction from the left image.

In the region of interest in Figure 3, the left image is the face recognition and feature value point tracking under the face dynamic registration technology just now, and the right image is the final obtained region of interest. If you look closely, you can see that the eye rectangle is not a horizontal rectangle. The eye rectangle has a slight, pixel-level tilt. It is very difficult for users to maintain an absolute level during use. In 20-second video shooting, tilt is inevitable. And we calculate the tilt angle and add the rotation matrix to further control the error and improve the accuracy.

The average value of the R, G, and B channels of the region of interest is calculated frame by frame, and the data that is shaken by the operation of the mobile phone in the first 2 seconds is removed, and finally, we obtain the initial value signal.

4.2 Moving window timing sampling experiment

We sample T seconds through the moving window time sequence, and each time the moving window time sequence sample is shifted to the right for 1 second, and these sampled signals are sent to further processing one by one, then we finally measure $18\text{-}T$ heart rate values, which are temporary heart rate values $H_1 \ldots H_{N-T-t}$, and finally, extract the mean value of the temporary heart rate value as the heart rate value H. The value of T is between 5 seconds and 16 seconds.

4.3 Initial signal preprocessing module experiment

Using the piecewise cubic Hermite interpolation is to divide the function domain interval into several intervals according to the piecewise interpolation method, and use the cubic Hermite interpolation on each interval.

A sixth-order Butterworth low-pass filter with a normalized cutoff frequency of 0.04 is selected. After filtering the data forward, the filtered sequence is inverted, and then it is passed through the filter again. Because the low-pass filter obtains the envelope of the signal, we will subtract this envelope from the signal to finally get a signal that reduces the effects of jitter and breathing, as shown in Figure 4.

4.4 Blind source separation algorithm experiment

As shown in Figure 5 before and after Fast-ICA processing, we can see the effect of the signal after the blind source separation algorithm is processed.

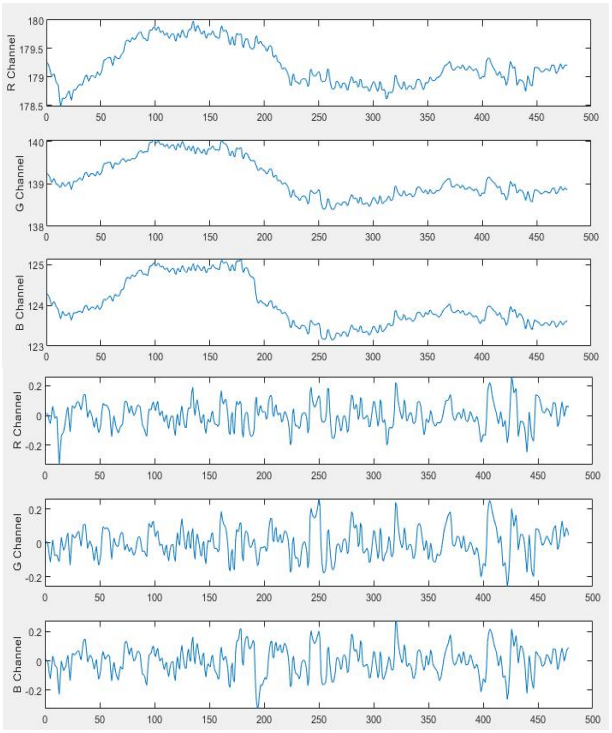

Figure 4. figure with before and after Butterworth filter processing.

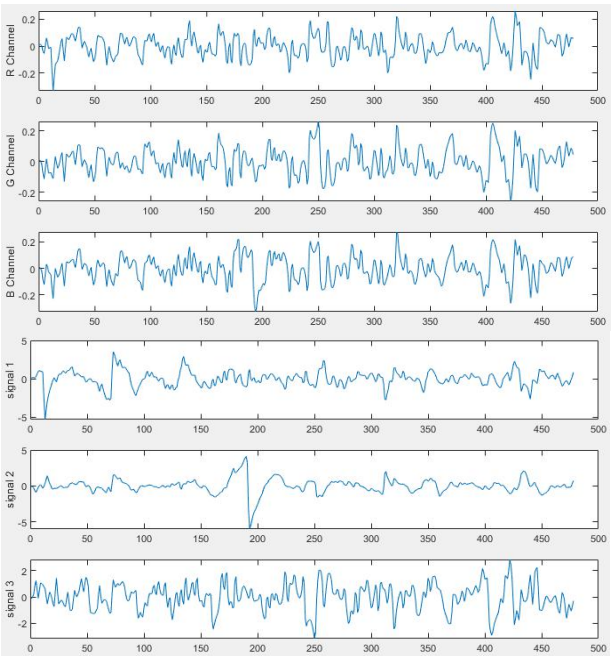

Figure 5. figure with before and after Fast-ICA processing.

4.5 Heart rate calculation module experiment

Through the heart rate calculation module, after the fast Fourier transform, the frequency signal is filtered with a 0.75-3.66Hz band-pass filter. After filtering the FFT signal, we can see the peak value. Finally, through peak detection, we extract the heart rate value.

4.6 Performance evaluation parameters

Root Mean Square (RMSE), Mean Absolute Error (MAE), and Standard Deviation (SD) are used to evaluate performance. Root Mean Square Error (RMSE): refers to the average square error between the output value and the true value. Mean Absolute Error (MAE): refers to the average error between the output value and the true value. Standard Deviation (SD): Refers to the arithmetic square root of the arithmetic mean of the squared deviation from the mean.

4.7. Analysis of experimental results

Table 1 is error analysis of experimental data. *N* is 20 seconds of video recorded, the result after the initial signal is obtained after removing the data of 2 seconds before *t* is 2 seconds, where the value of *T* is 5 to 16 seconds, you can see the movement at 10 seconds. The window timing sampling effect is the best. Table 1 shows that the root means square error (RMSE), average absolute error (MAE), and standard deviation (STD) are the lowest, which are 6.6929, 5.1365, 6, 6165, respectively, compared with sampling without using a moving piecewise function. The errors are 13.53, 10.79, and 14.09.

Table 1. Error analysis of experimental data.

Period (s)	5	6	7	8	9	10	11	12	13	14	15	16	Unused
RMSE	7.147	9.420	9.098	10.101	7.788	6.692	8.720	8.681	8.784	9.700	7.832	7.481	13.529
	0.538	0.696	0.672	0.747	0.567	0.495	0.645	0.642	0.649	0.717	0.579	0.553	1
MAE	5.381	6.249	6.171	7.475	6.139	5.136	6.412	6.568	6.361	6.690	6.026	6.109	10.793
	0.499	0.579	0.572	0.693	0.569	0.476	0.594	0.609	0.589	0.620	0.558	0.566	1
STD	7.045	9.075	8.905	9.340	7.382	6.616	8.248	8.410	8.772	9.913	8.058	7.840	14.087
	0.500	0.644	0.632	0.663	0.524	0.4696	0.586	0.597	0.623	0.704	0.572	0.557	1

Note: Errors are reduced by 50.5%, 52.4%, and 53.04%.

5. SUMMARY

In this paper, in order to improve the accuracy of rPPG detection technology for unrestricted scenes and scenes where the hand-held mobile phone extracts the heart rate, we propose a moving window timing sampling to refine the original video frame signal. And found that when the window is a 10-second moving window timing sampling, the effect is the best and the accuracy is the highest. It provides a basis for users to measure heart rate at any time and without restriction. Next, we will increase the number of experimenters and obtain more video samples to verify and adjust the algorithm. And enhance the algorithm's ability to de-jitter in the video frame signal.

ACKNOWLEDGEMENT

This work was supported in part by the Science and Technology Planning Project of Guangdong Science and Technology Department under Grant Guangdong Key Laboratory of Advanced IntelliSense Technology (2019B121203006).

REFERENCES

[1] “Summary of Cardiovascular Health and Disease Report 2020,” Chinese Journal of Circulation, 36(06), 521-545 (2021). (in Chinese)

[2] Li, H., Qian, F., Hou, C., Li, X., et al., “Longitudinal changes in depressive symptoms and risks of cardiovascular disease and all-cause mortality: A nationwide population-based cohort study,” Journals of Gerontology Series: A Biological Sciences and Medical Sciences, (2020).

[3] Wu, L. L., Zheng, C. Z., Liu, F. D., Qin, Y., Su, J., Cui, L., Du, W. C. and Zhou, J. Y., “The relationship between resting heart rate and high risk of cardiovascular disease,” Chinese Journal of Disease Control, (07), (2020). (in Chinese)

[4] Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J. and Garbey, M., "Interacting with human physiology," Computer Vision and Image Understanding, 108(1-2), 150-170 (2007).
[5] Li, Z. J., Wang, C., Zhu, H., et al., "Non-invasive continuous blood pressure measurement based on photoplethysmography," Chinese Journal of Biomedical Engineering, 31(04), 607-614 (2012). (in Chinese)
[6] Wang, W., Brinker, A. C. D., Stuijk, S. and de Haan, G., "Algorithmic principles of remote PPG," IEEE Transactions on Biomedical Engineering, 64(7), 1479-1491 (2017).
[7] Webster, J. G., [Design of Pulse Oximeters], CRC Press, (2002).
[8] Li, D. L., [Research on Non-Invasive Continuous Blood Pressure Measurement Method Based on Pulse Wave], Zhejiang University, Hangzhou, Doctor's Thesis (2008). (in Chinese)
[9] Kim, H. J., Kim, K. H., Hong, Y. S. and Choi, J. J., "Measurement of human heartbeat and respiration signals using phase detection radar," Review of Scientific Instruments, 78(10), 104703 (2007).
[10] Lu, G., Yang, F., Jing, X. and Wang, J., "Contact-free measurement of heartbeat signal via a doppler radar using adaptive filtering," Inter. Conf. on Image Analysis and Signal Processing, 89-92 (2010).
[11] Matsunaga, D., Izumi, S., Kawaguchi, H. and Yoshimoto, M., "Non-contact instantaneous heart rate monitoring using microwave Doppler sensor and time-frequency domain analysis," 2016 IEEE 16th Inter. Conf. on Bioinformatics and Bioengineering (BIBE), 172-175 (2016).
[12] Song, R., Li, J., Cheng, J., Li, C., Liu, Y. and Chen, X., "Motion robust imaging ballistocardiography through a two-step canonical correlation analysis," IEEE Transactions on Instrumentation and Measurement, 70, 1-10 (2021).
[13] Balakrishnan, G., Durand, F. and Guttag, J., "Detecting pulse from head motions in video," Proc. of 2013 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 3430-3437 (2013).
[14] Poh, M. Z., McDuff, D. J. and Picard, R. W., "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Transactions on Biomedical Engineering, 58(1), 7-11 (2011).
[15] Tarvainen, M. P., Ranta-aho, P. O. and Karjalainen, P. A., "An advanced detrending method with application to HRV analysis," IEEE Transactions on Biomedical Engineering, 49(2), 172-175 (2002).
[16] Li, X., Chen, J., Zhao, G. and Pietikainen, M., "Remote heart rate measurement from face videos under realistic situations," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 4264-4271 (2014).
[17] Jonathan, E. and Leahy, M., "Investigating a smartphone imaging unit for photoplethysmography," Physiological Measurement, 31(11), 79-83 (2010).
[18] Jonathan, E. and Leahy, M. J., "Cellular phone-based photoplethysmographic imaging," Journal of Biophotonics, 4(5), 293-296 (2011).
[19] Chen, W. and Mcduffff, D., "Deepphys: Video-based physiological measurement using convolutional attention networks," Proc. ECCV, (2018).
[20] Niu, X., Shan, S., Han, H. and Chen, X., "Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation," IEEE Trans. Image Processing, 29, 2409-2423 (2020).
[21] Špetlík, R., Franc, V., Cech, J. and Matas, J., "Visual heart rate estimation with convolutional neural network," Proc. BMVC, (2018).
[22] Yu, Z., Peng, W., Li, X., Hong, X. and Zhao, G., "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement," Proc. IEEE ICCV, (2019).
[23] Wang, W., den Brinker, A. C. and de Haan, G., "Full video pulse extraction," Biomedical Optics Express, 9(8), 3898-3914 (2018).
[24] Lewandowska, M., et al., "Measuring pulse rate with a webcam-A non-contact method for evaluating cardiac activity," Proc. Federated Conf. Comput. Sci. Inf. Syst., 405-410 (2011).
[25] Poh, M. Z., Mcduff, D. J. and Picard, R. W., "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Trans. Biomed Eng., 58, 7-11 (2011).
[26] Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J. F. and Sebe, N., "Selfadaptive matrix completion for heart rate estimation from face videos under realistic conditions," Proc. IEEE CVPR, (2016).
[27] de Haan, G. and Jeanne, V., "Robust pulse rate from chrominance-based rPPG," IEEE Trans. Biomed. Eng., 60, 2878-2886 (2013).
[28] de Haan, G. and van Leest, A., "Improved motion robustness of remotePPG by using the blood volume pulse signature," Physiol. Meas., 35, 1913-1922 (2014).
[29] Wang, W. et al., "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," IEEE Trans. Biomed. Eng., 63, 1974-1984 (2016).