

# Non-Contact Method of Heart Rate Measurement Based on Facial Tracking

Ruqiang Huang<sup>1</sup>, Weihua Su<sup>1\*</sup>, Shiyue Zhang<sup>1</sup>, Wei Qin<sup>2</sup>

<sup>1</sup>National Innovation Institute of Defense Technology, Academy of Military Science of Chinese PLA, Beijing, China

<sup>2</sup>Tianjin Artificial Intelligence Innovation Center, Tianjin, China

Email: \*directorsu@126.com

**How to cite this paper:** Huang, R.Q., Su, W.H., Zhang, S.Y. and Qin, W. (2019) Non-Contact Method of Heart Rate Measurement Based on Facial Tracking. *Journal of Computer and Communications*, 7, 17-28. <https://doi.org/10.4236/jcc.2019.75002>

**Received:** April 24, 2019

**Accepted:** May 24, 2019

**Published:** May 27, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Image photoplethysmography can realize low-cost and easy-to-operate non-contact heart rate detection from the facial video, and effectively overcome the limitations of traditional contact method in daily vital sign monitoring. However, it is hard to obtain more accurate heart rate detection values under the conditions of subject's facial movement, weak ambient light intensity and long detection distance, etc. In this article, a non-contact heart rate detection method based on face tracking is proposed, which can effectively improve the accuracy of non-contact heart rate detection method in practical application. The corner tracker algorithm is used to track the human face to reduce the motion artifact caused by the movement of the subject's face and enhance the use value of the signal. And the maximum ratio combining algorithm is used to weight the pixel space pulse wave signal in the facial region of interest to improve the pulse wave extraction accuracy. We analyzed the facial images collected under different experimental distances and action states. This proposed method significantly reduces the error rate compared with the independent component analysis method. After theoretical analysis and experimental verification, this method effectively reduces the error rate under different experimental variables and has good consistency with the heart rate value collected by the medical physiological vest. This method will help to improve the accuracy of non-contact heart rate detection in complex environments.

## Keywords

Heart Rate, Non-Contacting, Maximum Ratio Combining, Facial Video

---

## 1. Introduction

Heart rate is an important basis for the prevention and diagnosis of cardiovas-

cular disease, and it is also a key vital sign parameter for clinical monitoring [1]. Traditional Electrocardiogram (ECG) as the gold standard when it comes to the current heart rate measurement, is widely used in clinical diagnosis and physiological monitoring, but this method requires complicated professional operation and multiple electrodes to connect with the human body during use. Long-term daily monitoring is prone to skin discomfort, movement limited and other problems, especially in burns and scalds, allergic constitution, limb mutilation, neonate and other scenarios in the use of restrictions. In view of the shortcomings of traditional heart rate detection methods for contact subjects, researchers have proposed non-contact heart rate detection methods based on microwave radar [2], thermodynamic image [3], video image processing [4] and so on in recent years. Among them, the heart rate detection method based on video image processing has the advantages of simple equipment, stable system and fast operation, which has become a research hotspot and has received extensive attention.

The theoretical basis for non-contact heart rate detection method based on video images processing is image photoplethysmography (IPPG) [5]. Some literatures also refer to remote photoplethysmography (RPPG) [6] and photoplethysmography image (PPGI) [7]. Since Verkruyess [8] first proposed IPPG method, researchers have developed a variety of algorithms to improve heart rate detection accuracy, such as blind source separation method [9], chrominance method [10]. However, it is difficult for these algorithms to obtain a higher accuracy of heart rate estimation in the case of detection distance and face movement [11], Therefore, this paper proposes to use Kanade-Lucas-Tomasi (KLT) feature tracker [12] to track the detected face parts and improve the segmentation efficiency of region of interest extracted from pulse wave signals, enhance the signal value of the region of interest (ROI). The maximum ratio combination algorithm (MRC) is used to weight the pulse wave signals in the pixel space of the segmented area, so as to improve the facial pulse signal-to-noise ratio and obtain a higher precision heart rate. We set near and far groups of test distances. The experimental research was carried out under three scenarios of natural steady, free talk and head rotation, to verify the practical application effect of the algorithm proposed in this paper. The heart rate extracted by this method was compared with the parameter collected by medical physiological vest and independent component analysis (ICA) method [13]. Statistical test and error analysis were carried out based on the experimental data, and the results showed that the method presented in this paper could reduce errors in different experimental scenarios and improve the accuracy of heart rate detection.

## 2. Method

A periodic change in the volume of blood in the circulation of the body that is consistent with the pulse when the heart is pulsating. When exterior illumination intensity is stable, eliminate the factor that the skin, muscle, organization absorbs constant to smooth intensity, the heart is in systole and a pulsating cycle that diastole falls, reflex smooth intensity also produces periodic change as the

absorption change of peripheral blood volume. The essence of IPPG is to use this bio-optical principle to perceive and isolate heart rate signals from video images containing human faces, it has two basic steps: first, manually or automatically select the signal interest area of the face part in the video image. Then the image processing and signal analysis techniques are used to post-process the segmented ROI and extract the corresponding pulse waveform [5]. In practical applications, in the case of non-complete or non-autonomous movement of the subject, such as blinking eyes, licking lips, talking, it is not conducive to face segmentation. For this, the KLT method [14] is used to track the facial part in real time, which improves the extraction and separation efficiency of the face's interest area during the movement and enhances the signal use value.

Because of the different blood vessel distribution in the face part, the value of pulse wave signal in the effective region of interest is also different [11]. Therefore, the maximum ratio combination algorithm [15] is used to merge the pulse wave signals in different pixel regions of the region of interest in the face part, for improve the accuracy of heart rate detection. The method flow is divided into two steps: 1) preprocessing: decompose the video into image sequences, use Viola-Jones face detector [16] to identify face regions, then use KLT algorithm to track the face, segment the contour of the face, and finally extract the ROI. 2) Signal extraction: set up signal buffer and sampling window to carry out spatial transformation, filtering and spectral peak extraction for the signal strength of ROI, and finally the heart rate was proposed through MRC algorithm.

## 2.1. Face Tracking

The KLT algorithm is a corner feature point tracking algorithm proposed by Kanade and Lucas [14]. This method performs feature matching on two adjacent frames of the video image according to the second-order derivative gradient local search idea. The feature points in the current image frame are used as tracking points and the optical flow estimation is performed, and the position of the tracking point corresponding to the image of the next frame is calculated, thereby calculating the position change amount between the two frames of images. KLT has the advantages of high tracking efficiency and fast positioning speed in application scenarios such as constant brightness, continuous time and spatial consistency. Constant brightness means that the external illumination of the tracked part is stable; time continuous means that there is temporal correlation between adjacent frames, and the feature point motion is small, spatial consistency means that the pixel motion in the local area is consistent, and the proximity equation can be established to solve the motion of the center point pixel. Therefore, KLT has good applicability in theory in the non-contact heart rate detection method based on face information.

In order to eliminate the large motion artifacts caused by human physiological and natural movements and other factors that interfere with the extraction of physiological signals, we locate and extract the ROI with the forehead, eyes, nose and mouth and facial contour as feature points. Use Viola-Jones face detector to

recognize and detect faces, then using Harris corner detection, the image of the partial image is detected by the window of the partial image, and the pixel that moves in any direction causes the image gray to change significantly. The feature points are tracked and adjusted in real time according to the displacement vector to achieve the effect of face tracking region positioning. The KLT algorithm has a good tracking effect in the case of rotation, swing, near-far movement and partial occlusion of the face, which helps to accurately segment the face region and enhance the quality of physiological signal extraction. The effects of different face tracking scenes are shown in **Figure 1**.

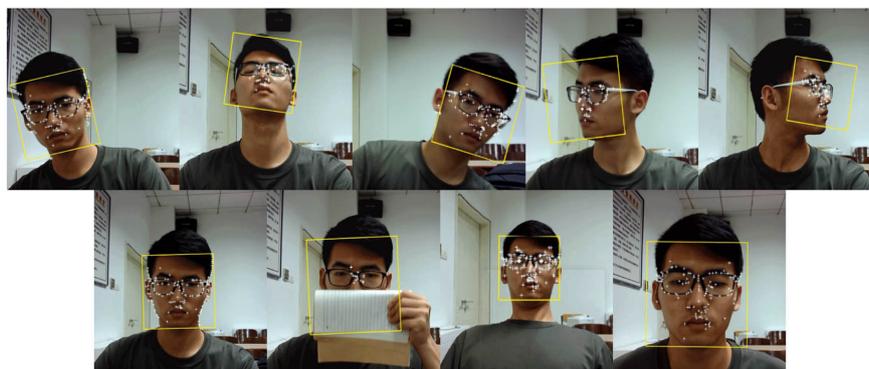
## 2.2. MRC Algorithm

MRC is the optimal choice in diversity merging technology, which can obtain the best performance and bring higher signal-to-noise ratio than selective combining and equal gain combining [15]. The core idea is to weight the pulse wave signal of each channel, that is, the pixel region, and increase the signal-to-noise ratio to obtain the optimal estimation value.

Each frame recorded by the camera contains a video image of the face that records the intensity change information of the ambient light reflected back by the face and is mapped into the image two-dimensional pixel space. The recorded video is in the form of intensity signal  $V(x, y, t)$  comprising of sequences of frames  $V(x, y, t = 1, 2, 3, \dots)$ . The change in light intensity produced by ambient light on the face can be broken down into two parts: 1) incident light intensity  $I(x, y, t)$ , e.g. ambient light or artificial light source produces light intensity. 2) facial reflection intensity  $R(x, y, t)$ , That is, the intensity of the reflected light generated by the light passing through the face, including the back-scattered light intensity of the transmitted skin after absorption by the blood and the surface reflected light intensity of the non-transmitted skin [11]:

$$V(x, y, t) = I(x, y, t)R(x, y, t) \quad (1)$$

Different regions of the skin show variation delay due to different blood flow time [12], for example, there is a delay of 80 ms in the pulse wave signal between the finger and the auricle, but the delay between the farthest two points of the face is less than 10ms, and the delay is too small to be considered in the case of



**Figure 1.** Multiple face tracking.

a camera with a sampling frequency of 30 - 60 Hz. We divide the segmented face ROI into small enough areas according to the pixel size and record it as set  $Q$ . The area of the elements in the  $Q$  set is locally correlated and the blood perfusion is consistent, and each area is recorded as  $i (i \in \{1, 2, \dots, n\})$ ,  $y_i(t)$  is pixel space average signal strength at time  $T$  on  $Q$ , then the model can be established by (1):

$$I_i(\alpha_i \cdot p(t) + b_i) + q_i(t) \quad (2)$$

Normally, the incident light intensity  $I(x, y)$  is constant,  $\alpha_i$  is physiological coefficient of performance, determined by factors such as skin color, blood volume, and oxygen saturation. For  $V(x, y, t)$ , calculate the average spatial signal strength of the pixel  $y_i(t)$  in the set  $Q$ . But  $y_i(t)$  contains different intensity pulse wave signals  $p(t)$  and different surface reflected light intensities and, using a Band-Pass Filter [0.5 Hz, 5 Hz] for time domain filtering to eliminate out-of-range surface reflections  $(I_i, b_i)$  and other signal noises, resulting in closer approximation of true pulse wave signals  $\hat{y}_i(t)$ , Equation (2) is further simplified as

$$\begin{cases} \hat{y}_1(t) = A_1 p(t) + w_1(t), \\ \hat{y}_2(t) = A_2 p(t) + w_2(t), \\ \vdots \\ \hat{y}_n(t) = A_n p(t) + w_n(t) \end{cases} \quad (3)$$

where  $i \in \{1, 2, \dots, n\}$  represents the set  $Q$  number of the corresponding ROI,  $A_i$  is the coefficient, determined by the physiological coefficient  $\alpha_i$  and the incident light intensity  $I_i$ , and  $w_i(t)$  represents the noise component such as camera quantization noise, motion artifact.  $\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_n(t)$  contain  $p(t)$  of different intensities received by different channels, and different levels of noise ( $A_i$  and  $w_i(t)$  are unknown), so the accurate pulse wave signal  $p(t)$  cannot be obtained. We use the MRC algorithm to perform weighted averaging operations on these different channel regions to obtain the pulse wave estimate  $\hat{p}_i(t)$ :

$$\hat{p}_i(t) = \sum_{i=1}^n G_i \hat{y}_i(t) \quad (4)$$

According to the MRC algorithm principle [15], in order to obtain the maximum overall measurement signal to noise ratio, the weight  $G_i$  of each channel should be proportional to the root mean square value of the channel signal component, and inversely proportional to the mean square error of the channel component. Here:

$$G_i = \frac{A_i}{\|w_i(t)\|^2} \quad (5)$$

Since the oscillation frequency of  $p(t)$  is substantially equal to the heart rate, the spectral power of the pulse wave signal  $p(t)$  fluctuates within the heart rate range, and the spectral energy  $w_i(t)$  of the noise is distributed be-

tween the bandwidth range [0.5 Hz, 5 Hz]. Based on the spectral structure of the signal, we can also define  $G_i$  as the ratio of the energy  $\hat{y}_i(t)$  in the pulse range to the noise within the bandwidth. Assuming that  $\hat{Y}(f)$  is the power spectral density of  $\hat{y}_i(t)$  between  $[0, T]$ , the optimal weight  $G_i$  is:

$$G_i(HR) = \frac{\int_{HR-b}^{HR+b} \hat{Y}_i(f) df}{\int_{B1}^{B2} \hat{Y}_i(f) df - \int_{HR-b}^{HR+b} \hat{Y}_i(f) df} \tag{6}$$

where  $[HR - b, HR + b]$  represents a small range of heart rate, which is the filter pass-band of the bandwidth filter. From this we can determine the best value of  $G_i$  by calculating the peak frequency in the spectrum. Since most of the large amplitude is from illumination changes or motion artifacts, we use threshold  $A_{th}$  to exclude excessive amplitudes in the region, according to the heart rate fluctuation range, the  $A_{th}$  equal 8. The final pulse wave estimates  $\hat{p}_i(t)$  in the  $T$  time window. The display is as follows:

$$\hat{p}(t) = \sum_{i=1}^n G_i \hat{y}_i(t) I(\hat{y}_{\max,i} - \hat{y}_{\min,i} < A_{th}) \tag{7}$$

$I(\cdot)$  is the indication function,  $\hat{y}_{\max,i} = \max_{t \in [0, T]} \hat{y}_i(t)$ ,  $\hat{y}_{\min,i} = \min_{t \in [0, T]} \hat{y}_i(t)$  is the maximum or minimum amplitude of  $\hat{y}_i(t)$  in the  $t$  time window.

### 3. Experiment Setup

#### 3.1. Experimental Premise

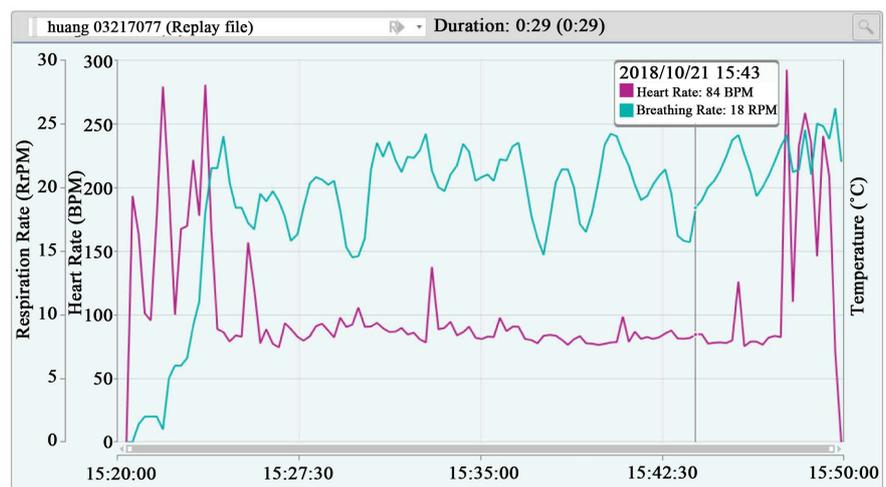
In the experiment, the CMOS webcam (Hikvision, ds-2cd8426fwd/f-i) was used as the video acquisition equipment, the camera work at a frame rate of 30 fps, with a resolution of  $1280 \times 720$  pixels and with 8 bits depth to record color video including human faces. Experimental computer calculates the frequency is 3.6 GHz. Equivital medical physiological monitoring vest (EQ02+ Life Monitor) was used as the contrast device, as heart rate gathering equipment, it has gained the certification of FDA and CE, the accuracy conforms to the standard of medical equipment, It can collect the bipolar leads of electrocardiogram signals of 256 Hz, and it can eliminate the psychological burden on the subjects caused by the traditional finger clip Oximeter and the electrode-patch ECG monitor, thus helping to improve the experimental effect, the vest and parameters are shown in **Figure 2** and **Figure 3**. This experiment based on Matlab programming and signal processing on Windows system.

#### 3.2. Experimental Posture

Natural light as the illumination condition in this experiment. Participants wore a physiological vest and sat facing the camera in a sitting position. At the distance of 1 m and 1.5 m respectively from the camera, the participants were tested with natural steady, free talking and head rotation, each experimental action was recorded as a group of video with a recording time of at least one minute, the experimental scenes are shown in **Figure 4**. In order to eliminate the interference



**Figure 2.** EQ02+ Life Monitor.



**Figure 3.** Physiological parameter.



**Figure 4.** Experimental scene.

factors during the experiment, participants are required to actively cooperate, naturally and evenly breathe, and perform the prescribed actions according to the requirements to avoid involuntary movement.

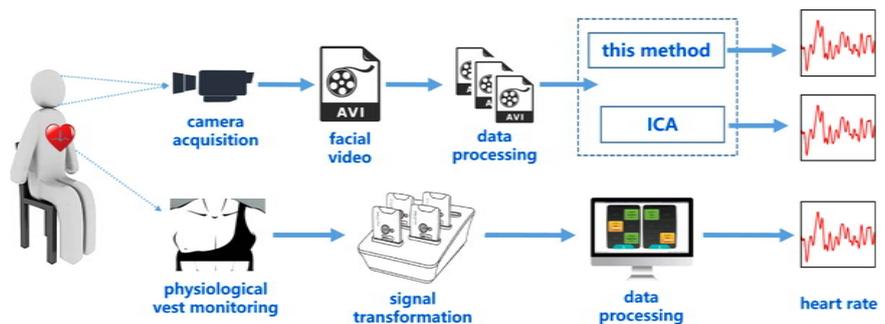
### 3.3. Experimental Procedure

The experimental flow arrangement is as shown in **Figure 5**, while facial video is recorded by camera, physiological parameters of subjects are collected by physiological vest, recorded video file was saved as the experimental sample by time stamping. Then ICA method and this method were used for processing and estimation respectively, and the heart rate estimation of the two methods was obtained. Finally, the experimental data is statistically analyzed.

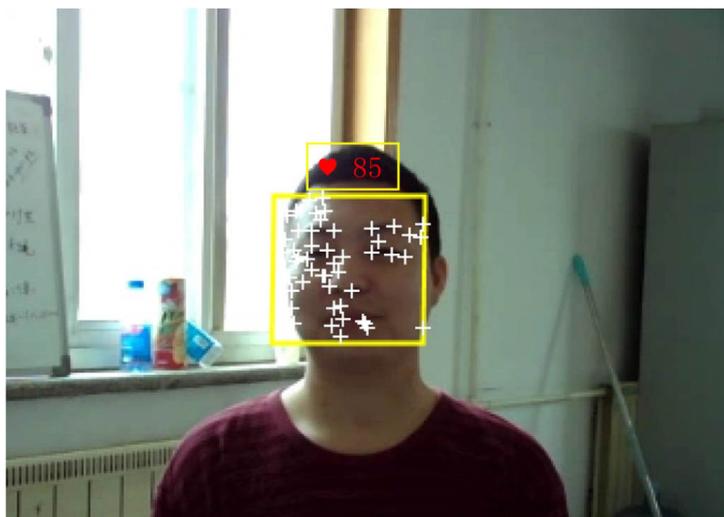
## 4. Results

### 4.1. Experimental Result

The experimental results obtained by this method are shown in the **Figures 6-8**, in this experiment, three subjects were tested, and the distance and motion were taken as experimental variables. 18 video images with a duration of 1 min were sorted out as experimental samples, the average heart rate within 10 s of the stabilization time was recorded as the effective value. **Figure 9** shows the comparison between the estimated heart rate of the same subject generated by the two methods and the truth value collected by the vest.



**Figure 5.** Experimental procedure.



**Figure 6.** Real time heart rate detection.

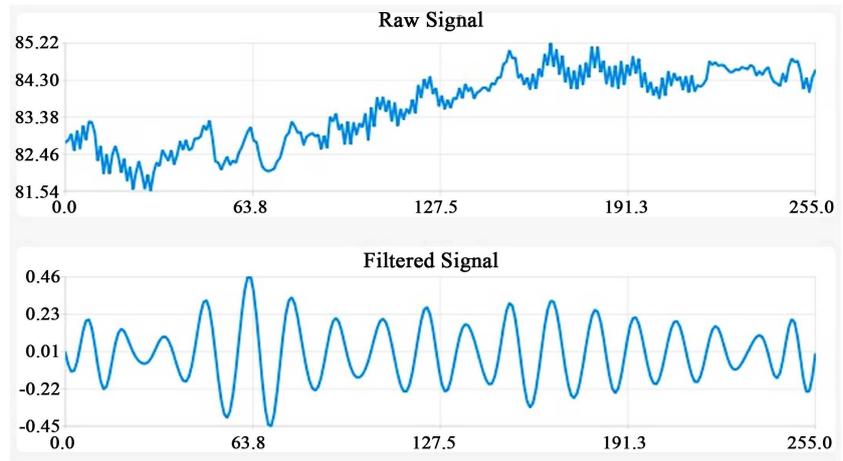


Figure 7. Filtering processing of the raw signal.

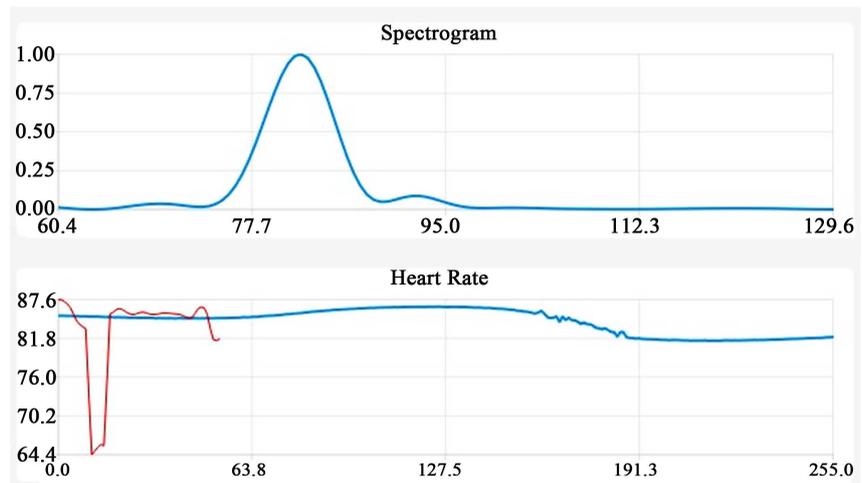


Figure 8. Heart rate extracted from the spectrogram peak interval.

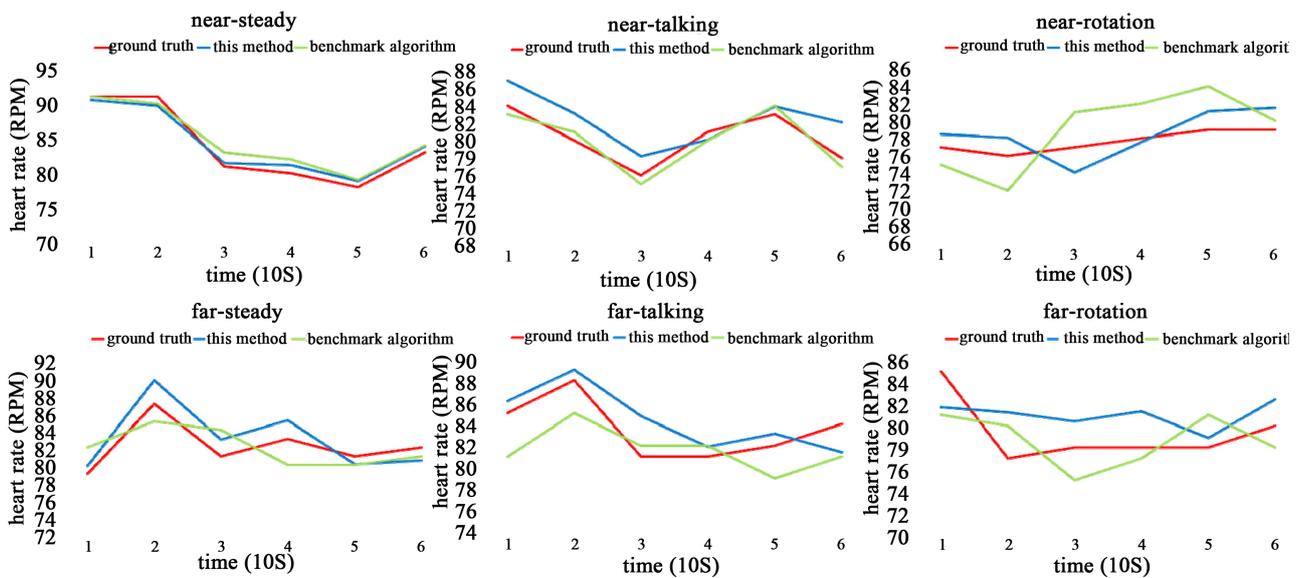


Figure 9. A comparison of heart rate between the two methods.

### 4.2. Consistency Analysis

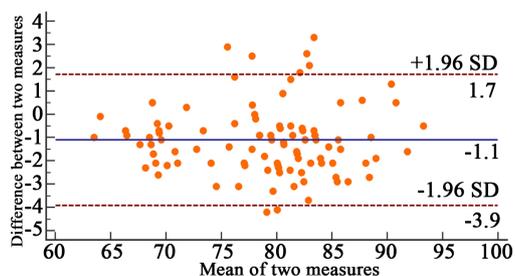
The heart rate is a continuous physiological signal, and the Bland-Altman method is used to analyze the random effects of the heart rate data obtained by this method and physiological vest. The analysis results show that the mean value of the paired heart rate data of 108 groups is  $-1.1$  bpm, the standard deviation of the difference is  $1.2$  bpm, the 95% limits of agreement were determined by  $[-3.9$  bpm,  $1.7$  bpm], the difference scatter plot is shown as **Figure 10**. It can be seen from the figure that most of the difference is within the consistency limit, and the absolute value does not exceed  $5$  bpm, and the average heart rate detected by the two methods is  $78.8$  bpm, so the difference is acceptable, while the results measured by the two methods have a good consistency.

### 4.3. Error Analysis

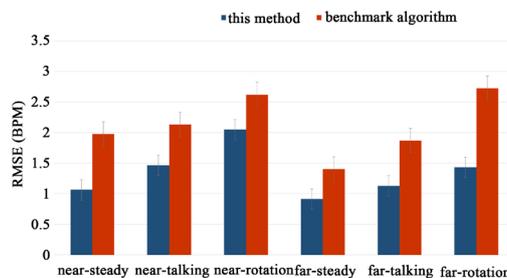
The heart rate obtained by the ICA method and the method is an observation value ( $X_{obs}$ ), and the physiological vest data is a contrast value ( $X_{model}$ ), then made error analysis respectively. The deviation between the observed value and the control value is measured by comparing the root mean square error value. With a mathematical model [11]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_{obs,i} - X_{model,i})^2} \tag{8}$$

After calculation, the RMSE comparison obtained is shown in **Figure 11**. A paired sample T-checks was performed on the relative error between the observed value and the control value to obtain a P value of  $0.009$  and less than  $0.05$ , indicating that the two detection methods have significant differences.



**Figure 10.** Bland-Altman plot: comparison of HR derived from this method and from physiological vest.



**Figure 11.** RMSE comparison.

#### 4.4. Evaluation

The experiment used different test distances and different action states as variables to detect heart rate by this method and ICA method respectively. By comparing the statistical analysis of the obtained data with the error, the heart rate detected by the method is in good agreement with the heart rate collected by the physiological vest, and the error rate under the near and far experimental distance is generally lower than that of the ICA method. A T-checks of the relative errors of the two methods shows that the two methods have significant differences. Non-contact heart rate detection using this method is more accurate than the classical independent component analysis method, and it is performance more robust in face movement and longer distance.

#### 5. Conclusion

Aiming at the IPPG method difficulties of weak pulse wave signal, motion artifact and low detection accuracy under the conditions of subject's face movement, weak ambient light intensity and long detection distance, this paper proposes a non-contact heart rate detection based on facial tracking. The method uses the KLT algorithm to accurately locate the human face, reduces the motion artifact generated by the face movement, enhances the use value of the ROI region signal, and uses the MRC algorithm to weight the pulse signal in the ROI to eliminate signal interference in different regions, to obtain a higher accuracy heart rate estimates. The experimental results show that under natural light conditions, the method in this paper processes and analyzes videos containing human face under different test distances and experimental actions. The extracted heart rate and the collected control value of physiological vest have a good consistency, and are more accurate than the detection value of ICA method. In the case of multi-target extraction, change of illumination intensity and short sampling time, how to realize heart rate detection in real scenes is the development direction of non-contact physiological parameter detection technology in the future. With more advanced algorithms, multi-parameter, high-accuracy and robust non-contact real-time physiological parameter monitoring technology will have great development prospects in application platforms such as neonatal intensive care unit and medical service robot.

#### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

#### References

- [1] Wang, W., Brinker, B.D., Stuijk, S. and Haan, G.D. (2016) Algorithmic Principles of Remote-PPG. *IEEE Transactions on Biomedical Engineering*, **99**, 1479-1491. <https://doi.org/10.1109/TBME.2016.2609282>
- [2] Lu, G., Yang, F., Jing, X., Yu, X., Zhang, H., Xue, H., *et al.* (2011) Contact-Free

Monitoring of Human Vital Signs via a Microwave Sensor. *5th International Conference on Bioinformatics and Biomedical Engineering*, Wuhan, 10-12 May 2011, 1-3. <https://doi.org/10.1109/icbbe.2011.5781497>

- [3] Brueser, C., Hoog Antink, C., Wartzek, T., Walter, M. and Leonhardt, S. (2015) Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques. *IEEE Reviews in Biomedical Engineering*, **8**, 30-43. <https://doi.org/10.1109/RBME.2015.2414661>
- [4] Poh, M.Z., Mcduff, D.J. and Picard, R.W. (2010) Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Transactions on Biomedical Engineering*, **58**, 7-11. <https://doi.org/10.1109/TBME.2010.2086456>
- [5] Sun, Y. and Thakor, N. (2015) Photoplethysmography Revisited: From Contact to Noncontact, from Point to Imaging. *IEEE Transactions on Bio-Medical Engineering*, **63**, 463-477. <https://doi.org/10.1109/TBME.2015.2476337>
- [6] Sun, Y., Papin, C., Azorin-Peris, V., Kalawsky, R., Greenwald, S. and Hu, S. (2012) Use of Ambient Light in Remote Photoplethysmographic Systems: Comparison between a High-Performance Camera and a Low-Cost Webcam. *Journal of Biomedical Optics*, **17**, Article ID: 037005. <https://doi.org/10.1117/1.JBO.17.3.037005>
- [7] Moço, A.V., Stuijk, S. and Haan, G.D. (2016) Motion Robust PPG-Imaging through Color Channel Mapping. *Biomedical Optics Express*, **7**, 1737-1754. <https://doi.org/10.1364/BOE.7.001737>
- [8] Verkruyse, W., Svaasand, L.O. and Nelson, J.S. (2008) Remote Plethysmographic Imaging Using Ambient Light. *Optics Express*, **16**, 21434-21445. <https://doi.org/10.1364/OE.16.021434>
- [9] Poh, M.Z., Mcduff, D.J. and Picard, R.W. (2010) Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. *Optics Express*, **18**, 10762-10774. <https://doi.org/10.1364/OE.18.010762>
- [10] De Haan, G. and Jeanne, V. (2013) Robust Pulse Rate from Chrominance-Based RPPG. *IEEE Transactions on Biomedical Engineering*, **60**, 2878-2886. <https://doi.org/10.1109/TBME.2013.2266196>
- [11] Kumar, M., Veeraraghavan, A. and Sabharwal, A. (2015) Distance-PPG: Robust Non-Contact Vital Signs Monitoring Using a Camera. *Biomedical Optics Express*, **6**, 1565-1588. <https://doi.org/10.1364/BOE.6.001565>
- [12] Mstafa, R.J. and Elleithy, K.M. (2016) A Video Steganography Algorithm Based on Kanade-Lucas-Tomasi Tracking Algorithm and Error Correcting Codes. *Multimedia Tools and Applications*, **75**, 10311-10333. <https://doi.org/10.1007/s11042-015-3060-0>
- [13] Macwan, R., Benezeth, Y. and Mansouri, A. (2018) Remote Photoplethysmography with Constrained ICA Using Periodicity and Chrominance Constraints. *BioMedical Engineering OnLine*, **17**, 22. <https://doi.org/10.1186/s12938-018-0450-3>
- [14] Lee, H.K., Choi, K.W., Kong, D. and Won, J. (2013) Improved Kanade-Lucas-Tomasi Tracker for Images with Scale Changes. 2013 *IEEE International Conference on Consumer Electronics*, Las Vegas, 11-14 January 2013, 33-34.
- [15] He, F., Man, H. and Wang, W. (2011) Maximal Ratio Diversity Combining Enhanced Security. *IEEE Communications Letters*, **15**, 509-511. <https://doi.org/10.1109/LCOMM.2011.030911.102343>
- [16] Viola, P., Jones, M. and Snow, D. (2003) Detecting Pedestrians Using Patterns of Motion and Appearance. *Proceedings Ninth IEEE International Conference on Computer Vision*, Nice, 13-16 October 2003, Vol. 2, 734-741. <https://doi.org/10.1109/ICCV.2003.1238422>