Multimedia Tools and Applications (2019) 78:26747–26757
https://doi.org/10.1007/s11042-019-07849-x

# Motion-tolerant heart rate estimation from face videos using derivative filter

Zhao Yang[1,2] · Xuezhi Yang[1,2] · Xiu Wu[1,2]

Received: 21 September 2018 / Revised: 14 April 2019 / Accepted: 31 May 2019 /
Published online: 13 June 2019


**Abstract**
Imaging photoplethysmography (IPPG) technique allows us to extract blood volume pulse (BVP) signals from face videos for measuring heart rate (HR), which is useful in applications such as neonatal monitoring, telemedicine and affective computing. Because the BVP signal is small, the HR estimation results are sensitive to face motion disturbance caused by spontaneous head movements and facial expressions of subjects. In this paper, we design a novel filtering method for refining the RGB signals with motion artifacts. Based on the observation that subtle color changes of face skin are smoother than large face motions at temporal scale, we use the three-order derivative of Gaussian filter to select subtle color changes under large motions. Our method is validated on both our self-collected dataset and public dataset MAHNOB-HCI containing face videos with head movements and facial expressions. By employing the proposed filtering method to pre-process the RGB signals before BVP signal extraction, a range of IPPG methods are improved to generate robust HR estimation results under realistic situations.

**Keywords** Photoplethysmography · Imaging · Derivative filter · Motion interference · Heart rate estimation

## 1 Introduction

Heart rate (HR) is an important physical sign for clinical diagnosis and health monitoring. The HR information can be widely used in many application fields, such as telemedicine [26], fitness exercise [18] and affective computing [12]. Traditional methods for HR measurement are based on contact sensors, such as electrocardiogram (ECG) [5] and photoplethysmogram

✉ Xuezhi Yang
xzyang@hfut.edu.cn

1 School of Computer and Information, Hefei University of Technology, No.193, Tunxi Road Baohe District, Hefei 230009, China

2 Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, China

(PPG) [17], which are not convenient or comfortable for subjects. To address these issues, imaging photoplethysmography (IPPG) technique is presented for contactless HR estimation [17]. Inspired by the contact reflective PPG, IPPG uses a camera as a photodetector array to record the skin color variation which reflects the blood volume pulse (BVP) in the microvascular tissue bed beneath the skin. The BVP signal can further be processed to calculate corresponding HR value.

In recent years, several IPPG methods for HR estimation from face videos have been presented [2, 3, 14, 21, 23]. These approaches use multi-channel information of camera sensors to extract pure BVP signals for HR inference. Verkruysse et al. [21] pointed out that the green channel signal from face videos contains the strongest blood pulsating signal compared to other channels. Poh et al. [14] used independent component analysis (ICA) method to separate the BVP signal from the RGB channel signals. The CHROM [2] method designed a chrominance feature for HR estimation by combining the RGB channels linearly under skin-tone standardization. The Pulse Blood Vector (PBV) method [3] weighted the RGB channels to recover the pulse wave based on the different absorption spectra of arterial blood and bloodless skin. The plane orthogonal to the skin (POS) [23] algorithm extracted the BVP signal in the presence of large head motions by using the plane orthogonal to the skin color space.

However, the subjects in realistic environments often have spontaneous head movements and facial expressions which severely degrade the performances of current HR measurement methods. The BVP signal is prone to noise or motion artifacts, so two major filtering methods for motion disturbance reduction are proposed to pre-process the RGB signals before BVP signal extraction. One is the bandpass filtering (BPF), which is most commonly used to eliminate the frequency components of noise beyond the range of HR frequency. A drawback of BPF is that it cannot handle the case when the frequency of motion disturbance is within the HR frequency band. The other is the amplitude selective filtering (ASF) [24], which distinguishes the frequency components of motion noise and BVP signal. It is based on the observation that the spectral amplitude of motion is much larger than the spectral amplitude of pulse. However, this observation fact does not always exist when the subjects produce a lot of facial expressions. In addition to BPF and ASF, it is shown that the Least Mean Square (LMS) adaptive filter can also efficiently reduce motion artifacts in the BVP signal [10]. A limitation of the LMS filter is that it is sensitive to the scaling of input signals.

This paper presents a novel filtering method for refining the RGB signals with motion artifacts before BVP signal extraction. Based on the observation that subtle color changes in face skin are smoother than large face motions at temporal scale, we use the three-order derivative of Gaussian filter to select subtle color changes under large motions. By leveraging the proposed filtering method to pre-process the RGB signals, a range of IPPG methods are improved to generate robust HR estimation results under realistic situations.

## 2 Problem formulation

The HR measurement from face videos depends on an accurate BVP signal $s(t)$ estimated from the recorded video $V(x, y, t)$, where $x$, $y$ denote the pixel coordinates and $t$ is the frame number. The signal $s(t)$ is proportional to the blood volume change in the arteries and capillaries caused by cardiac activity, which corresponds to the synchronous small changes in skin color [17].

Previous works [8, 25] pointed out that the reflected light $V(x, y, t)$ recorded by the video camera sensors can be divided into three parts: (i) the surface reflected light $R(x, y, t)$ which is directly reflected from the skin surface, (ii) the subsurface reflected light $A(x, y, t)$ which penetrates the skin and then is reflected after absorbance by various tissues, and (iii) the pulsatile reflected light $B(x, y, t)$ which penetrates deeper to probe the vasculature and is reflected after the modulation of pulsation:

$$V(x,y,t) = R(x,y,t) + A(x,y,t) + B(x,y,t) \tag{1}$$

We assume that the illumination is invariable such that $R(x, y, t)$ and $A(x, y, t)$ are constant along the time while $B(x, y, t)$ fluctuates quasi-periodically due to the absorption by oxygenated arterial blood. However, the facial motion such as face shaking and expression changes will affect these parts of the reflected light $V(x, y, t)$. The motion-induced terms are introduced into $R(x, y, t)$, $A(x, y, t)$ and $B(x, y, t)$ as follows:

$$V(x,y,t) = (R_c(x,y) + R_m(x,y,t)) + (A_c(x,y) + A_m(x,y,t)) + \left(B_p(x,y,t) + B_m(x,y,t)\right) \tag{2}$$

where $R_c(x, y)$, $A_c(x, y)$ denote the time-constant terms, $R_m(x, y, t)$, $A_m(x, y, t)$, $B_m(x, y, t)$ are the motion-induced terms, and $B_p(x, y, t)$ is the pulse-induced term.

We combine the time-constant terms and motion-induced terms respectively to simplify the eq. (2) as:

$$V(x,y,t) = C(x,y) + B_p(x,y,t) + M(x,y,t) \tag{3}$$

where $C(x, y) = R_c(x, y) + A_c(x, y)$ and $M(x, y, t) = R_m(x, y, t) + A_m(x, y, t) + B_m(x, y, t)$.

By assuming that the time-constant term and the motion-induced term are outside the range of human pulse-rate such as [40, 240] bpm (beat per minute), a bandpass filter is typically applied to extract the BVP signal $s(x, y, t)$:

$$s(x,y,t) = V(x,y,t) \otimes BPF(t) \tag{4}$$

where $BPF(t)$ denotes the bandpass filter and $\otimes$ denotes the convolution operator.

Similar to the bandpass filter, the amplitude selective filter [24] is presented to obtain the clean BVP signal based on the difference of spectral amplitudes between pulse signals and motions. It can be used as a pre-process for BVP signal extraction by weighting in the frequency domain, which is equivalent to the convolution operation in the time domain:

$$s(x,y,t) = V(x,y,t) \otimes ASF(t) \tag{5}$$

where $ASF(t)$ denotes the amplitude-selective filter and $\otimes$ is the convolution operator.

## 3 Methods

In this section, we first analyze the criteria for designing a derivative filter, and then describe the proposed algorithm for motion-tolerant HR estimation from face videos.

### 3.1 Analysis

Compared to the motion interference, the facial color variation is a subtle change which depicts smoother trajectories at temporal scale. This observation fact can be characterized by the third temporal derivative of input video signal $V(x, y, t)$, which has been applied in a wide range of fields for assessing steepness or smoothness of time series data [4, 13, 15]. The third derivative value becomes low during smooth changes but high during steep changes, so we can use it to separate the pulse signal and motion noise. First, the third derivative of $V(x, y, t)$is calculated to cancel the time-constant term while to remain the pulse-induced and motion-induced parts. Then, based on the difference of smoothness, we use a Gaussian filter $G_\sigma(t)$ to select the pulse-induced part as the BVP signal $s(x, y, t)$:

$$s(x,y,t) = \frac{\partial^3 V(x,y,t)}{\partial t^3} \otimes G_\sigma(t) \tag{6}$$

where $G_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}}$ is the one-dimensional Gaussian distribution function and $\sigma$ denotes the standard deviation of Gaussian.

The Gaussian is the only filter that does not introduce spurious resolution [6] and due to the linearity of the operators [7] the eq. (6) can be rewritten as:

$$\begin{aligned} s(x,y,t) &= \frac{\partial^3}{\partial t^3}[V(x,y,t) \otimes G_\sigma(t)] \\ &= \frac{\partial^3}{\partial t^3} \int V(x,y,\tau) G_\sigma(t-\tau) d\tau \\ &= \int V(x,y,\tau) \frac{\partial^3}{\partial t^3} G_\sigma(t-\tau) d\tau \\ &= V(x,y,t) \otimes \frac{\partial^3 G_\sigma(t)}{\partial t^3} \\ &= V(x,y,t) \otimes DF(t) \end{aligned} \tag{7}$$

We call $DF(t) = \frac{\partial^3 G_\sigma(t)}{\partial t^3}$ in eq. (7) the derivative filter (DF) in this paper, short for the three-order derivative of Gaussian filter. The standard deviation $\sigma$ of Gaussian makes us detect the three-order derivative value with the desired frequency component [9, 11]. For a given temporal frequency of interest $f$, we denote the observation scale as $\sigma = \frac{r}{4f\sqrt{2}}$, where $r$ is the frame rate of the input video and the temporal window width of Gaussian filter is $\frac{r}{4f}$.

### 3.2 Algorithm

The flow chart of our approach for HR estimation from face videos is shown as Fig. 1. It includes three main phases: Phase 1 is devoted to deal with face detection and tracking, Phase 2 first applies the derivative filter to eliminate the face motion interference on raw RGB signals and then extracts the BVP signal by a traditional IPPG method, and Phase 3 analyzes the estimated BVP signal based on Fast Fourier Transform (FFT) to select the maximum peak as the HR frequency. The maximum peak is finally multiplied by 60 to produce the average HR measurement.

#### 3.2.1 Face detection and tracking

In order to reduce the impact of facial rigid motion and background clutter, the face should be detected and tracked satisfactorily. The Viola-Jones face detector [22] is first applied to detect the facial region of interest (ROI) which is drawn as a rectangle in the starting frame. Then the Discriminative Response Map Fitting (DRMF) algorithm [1] is used to locate the coordinates of 66 facial landmarks in the ROI

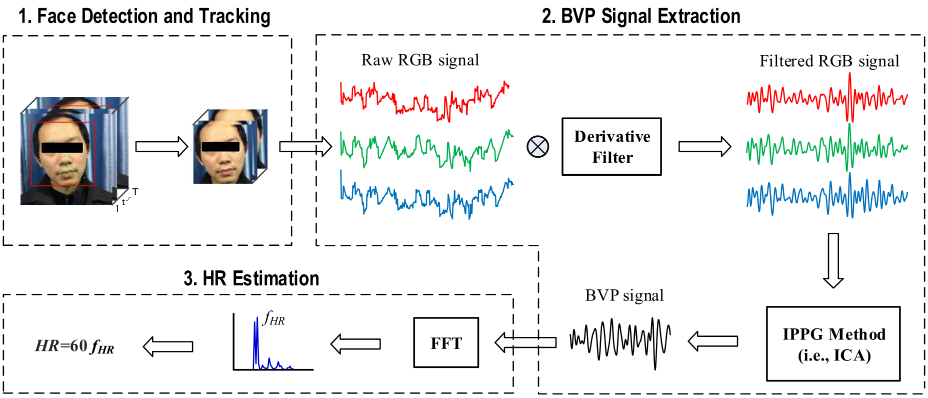


**Fig. 1** Flow chart of the proposed approach for HR estimation

rectangle, which describe the eyes, the nose, the mouth and the facial shape. Once located, these landmarks are tracked using the Kanede-Lucas-Tomasi (KLT) algorithm [19] to estimate the 2D geometric transformation for the ROI rectangle between the current and the next frame.

### 3.2.2 BVP signal extraction

We calculate the average value of all pixel in the facial ROI to obtain the raw RGB signals:

$$V_i(t) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} V_i(x,y,t)}{M \times N} \tag{8}$$

where $V_i(t)$ denotes the color variation signal in the i-th channel and $M \times N$ is the size of facial ROI. By spatially averaging, the incoherent changes due to camera quantization noise will cancel out over all the pixels within the ROI, and the coherent changes due to blood volume pulse will accumulate to form a quasi-periodic signal.

Then the proposed derivative filtering is applied to the traditional IPPG methods as a preprocessing step, which extracts the BVP signal $s(t)$ based on the raw RGB signals:

$$s(t) = IPPG(V_i(t) \otimes DF(t)) \tag{9}$$

where $IPPG(\cdot)$ denotes the traditional IPPG operator, such as ICA [14].

**Table 1** The $|M_e|$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the self-collected dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 5.854 | 5.273 | 4.842 | 4.769 | 4.273 |
| ASF | 4.873 | 4.787 | 3.286 | 3.275 | 3.187 |
| LMS | 4.562 | 4.358 | 3.235 | 3.232 | 3.163 |
| BPF | 3.524 | 3.471 | 3.157 | 3.156 | 3.144 |
| **DF** | **2.281** | **2.178** | **2.147** | **2.134** | **2.072** |

**Table 2** The $SD_e$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the self-collected dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 4.352 | 4.252 | 3.848 | 3.475 | 3.335 |
| ASF | 4.336 | 4.326 | 3.755 | 3.714 | 3.252 |
| LMS | 4.258 | 4.214 | 3.650 | 3.647 | 3.241 |
| BPF | 3.985 | 3.975 | 3.624 | 3.623 | 3.123 |
| **DF** | **2.772** | **2.735** | **2.714** | **2.663** | **2.612** |

### 3.2.3 HR estimation

In this phase the BVP signal is transformed into the frequency domain using FFT. We select the frequency with the highest peak in the Fourier domain as the HR frequency $f_{HR}$. The HR in bpm is then computed as $HR = 60f_{HR}$.

## 4 Experiments

The proposed method is evaluated on two datasets: a self-collected dataset and a publicly available MAHNOB-HCI dataset [16]. Experimental results are generated using a non-optimized MATLAB code on a desktop computer with Inter(R) Core(TM) i7-3770 CPU and 8 GB RAM.

### 4.1 Experimental setup

We first use a self-collected dataset to verify the validity of our method. For this dataset, eight videos are recorded for each of 20 subjects (8 female and 12 male) aged from 18 to 60 years by a Logitech HD 1080p camera in laboratory lighted with fluorescent lamp. All videos are recorded in RGB color format at 30 frames per second with duration of 15 s and saved in AVI format with resolution of 640 × 480. During the video recording, the camera fixed on a laptop is placed in parallel to the face of subject at the distance of 0.5 m. An electrocardiograph (Heal Force PC-80B) with three sensors attached on participants' body is used to record the ground truth HR. The videos are recorded under realistic situation where the subjects are free to speak or move heads. In some videos, the subjects may also have expression changes due to laughing or smiling.

To further verify the robustness, we test our proposed method on a publicly available MAHNOB-HCI dataset. The MAHNOB-HCI dataset contains 600 facial videos with both head movements and facial expressions, which are captured in 24-bit color at 61 frames per

**Table 3** The $RMSE$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the self-collected dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 6.673 | 6.623 | 5.834 | 5.673 | 5.623 |
| ASF | 6.331 | 6.329 | 4.776 | 4.745 | 4.436 |
| LMS | 5.589 | 5.562 | 4.607 | 4.598 | 4.223 |
| BPF | 4.683 | 4.677 | 3.712 | 3.631 | 3.530 |
| **DF** | **2.905** | **2.856** | **2.665** | **2.592** | **2.590** |

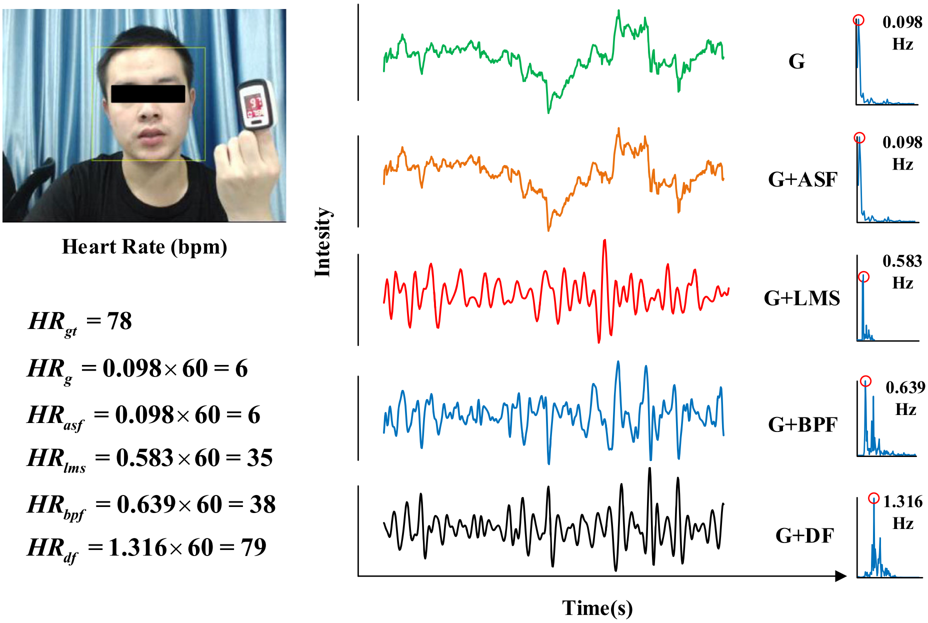


**Fig. 2** Result comparison between ASF, LMS, BPF and DF combined with the single channel method G [21]

second with a resolution of 780 × 580 from 30 human subjects. Following [20], we choose the first 27 human subjects (12 males and 15 females) and use a 30 s clip (from frames 306 to 2135) of 527 facial videos for a fair comparison. The second channel (EXG2, upper left corner of the chest) of the corresponding ECG signal is used to obtain the ground truth HR.

We compare four filtering methods: bandpass filter (BPF), amplitude-selective filter (ASF), Least Mean Square (LMS) adaptive filter and derivative filter (DF), as the pre-processing step in five existing IPPG algorithms, i.e., G [21], ICA [14], CHROM [2], PBV [3], and POS [23]. These IPPG algorithms are implemented as closely as possible and the parameters are fixed according to the original papers. A few parameters are defined for the benchmarked filters: the frequency-band $[f_l, f_h]$ for BPF, the amplitude-band $[a_l, a_h]$ for ASF and the desired frequency $f$ for DF. The default parameters are set as: $[f_l, f_h] = [0.7, 4]$Hz, $[a_l, a_h] = [0.0001, 0.002]$ and $f = (f_h - f_l)/2 = 1.65$Hz.

## 4.2 Results and discussion

To evaluate the proposed method, we select three commonly used metrics in the past work on remote HR estimation and analysis [20]. The estimation error $HR_e = HR_v - HR_{gt}$ is defined as the

**Table 4** The $|M_e|$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the MAHNOB-HCI dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 20.943 | 15.956 | 14.428 | 13.769 | 12.384 |
| ASF | 19.525 | 14.033 | 13.367 | 12.805 | 11.168 |
| LMS | 15.392 | 10.414 | 9.406 | 8.068 | 7.221 |
| BPF | 10.856 | 8.950 | 8.165 | 7.937 | 7.156 |
| **DF** | **6.843** | **6.672** | **5.881** | **5.753** | **5.694** |

**Table 5** The $SD_e$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the MAHNOB-HCI dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 40.291 | 38.869 | 36.532 | 33.435 | 31.880 |
| ASF | 39.572 | 37.438 | 35.342 | 32.654 | 30.795 |
| LMS | 34.863 | 32.944 | 30.335 | 27.575 | 25.866 |
| BPF | 28.786 | 24.345 | 22.424 | 19.467 | 17.731 |
| **DF** | **14.537** | **13.879** | **13.743** | **12.645** | **12.528** |

difference between heart rate $HR_v$ estimated from video and the ground truth heart rate $HR_{gt}$. The first and second metrics are the absolute mean and standard deviation of $HR_e$ denoted as $|M_e|$ and $SD_e$. The third evaluation metric is the Root Mean Squared Error (RMSE) denotes as $RMSE$. Based on these metrics, the performance comparison of five IPPG algorithms with different pre-processing filters on the self-collected dataset is shown in Tables 1, 2 and 3.

In the videos with face motions, the rigid motion interference can be eliminated by general ROI detection and tracking, but the non-rigid motions, such as speaking, laughing and facial expressions, are relatively difficult to handle. The goal of pre-processing filters is to deal with the non-rigid motion interference.

The results in Tables 1, 2 and 3 demonstrate that all pre-processing filters improve the baseline result of all IPPG algorithms to different extents. ASF has the minimum improvement from the baseline in metrics due to the fact that the spectrum of BVP signal is overwhelmed by the large non-rigid motions. LMS makes a slight improvement than ASF, but it is still unable to handle the non-rigid motions. BPF eliminates the frequency components of noise beyond the range of HR frequency, but the motion signals within the HR frequency band are remained to deteriorate the performance. DF yields the largest improvement from the baseline and makes the difference of performance between different IPPG algorithms disappear. It shows that DF can select subtle color changes in RGB channels while remove most motion noise.

Even using DF with the single channel method G [21], we can have good HR estimation results from face videos under motion disturbance. A sample result on the self-collected dataset is shown in Fig. 2, which demonstrates that DF is superior to BPF and ASF when combing with G [21]. After conducting DF, the waveform of BVP signal presents a prominent periodicity with less motion artifacts. The estimated HR with DF (79 bpm) is approximately equal to the ground truth HR of oximeter (78 bpm).

To demonstrate the validity and robustness, we also compare our method with the state-of-the-art methods on the public dataset MAHNOB-HCI. The experimental results on this more challenging dataset are shown in Tables 4, 5 and 6, which indicate that our method significantly outperforms the state-of-the-art by achieving the best performance.

**Table 6** The $RMSE$ (bpm) for evaluating the performance of each IPPG algorithm by using different pre-processing filters on the MAHNOB-HCI dataset (best performance in bold)

| Pre-processing | G | ICA | CHROM | PBV | POS |
|---|---|---|---|---|---|
| None (baseline) | 50.476 | 37.993 | 35.346 | 32.563 | 30.971 |
| ASF | 40.635 | 36.347 | 34.352 | 31.855 | 29.643 |
| LMS | 35.459 | 33.853 | 32.165 | 30.544 | 28.768 |
| BPF | 29.686 | 25.912 | 23.533 | 20.358 | 18.842 |
| **DF** | **15.423** | **14.701** | **14.097** | **13.649** | **13.516** |

## 5 Conclusion

In this paper, we presented a novel derivative filtering method to pre-process the RGB signals before BVP signal extraction for HR estimation. Based on the observation that the facial color variation depicts smoother trajectories than large face motions at temporal scale, we use the third temporal derivative of input video signal to separate the pulse signal and motion noise. The test databases contain challenging videos with face motions and experimental results indicate that a range of IPPG methods are improved to generate robust HR estimation results by employing the proposed DF method. Future work will include improving the robustness under varying illumination and using the modern deep learning technique for non-contact HR measurement.

**Acknowledgements** We acknowledge funding support from: Training Programme Foundation for Application of Scientific and Technological Achievements of Hefei University of Technology (JZ2018YYPY0289) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (JZ2018HGBZ0186).

## References

1. Asthana A, Zafeiriou S, Cheng S, Pantic M (2013) Robust Discriminative Response Map Fitting with Constrained Local Models. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, pp 3444–3451. https://doi.org/10.1109/CVPR.2013.442
2. de Haan G, Jeanne V (2013) Robust pulse rate from chrominance-based rPPG. IEEE Trans Biomed Eng 60(10):2878–2886. https://doi.org/10.1109/TBME.2013.2266196
3. de Haan G, van Leest A (2014) Improved motion robustness of remote-PPG by using the blood volume pulse signature. Physiol Meas 35(9):1913. https://doi.org/10.1088/0967-3334/35/9/1913
4. Flash T, Hogan N (1985) The coordination of arm movements: An experimentally confirmed mathematical model. J Neurosci 5(7):1688–1703. https://doi.org/10.1523/JNEUROSCI.05-07-01688.1985
5. Guven G, Gurkan H, Guz U (2018) Biometric identification using fingertip electrocardiogram signals. SIViP 12:933. https://doi.org/10.1007/s11760-018-1238-4
6. Koenderink JJ (1984) The structure of images. Biol Cybern 50(5):363–370. https://doi.org/10.1007/BF00336961
7. Koenderink JJ, van Doorn AJ (1992) Generic neighborhood operators. IEEE Trans Pattern Anal Mach Intell 14(6):597–605. https://doi.org/10.1109/34.141551
8. Kumar M, Veeraraghavan A, Sabharwal A (2015) DistancePPG: Robust non-contact vital signs monitoring using a camera. Biomedical Optics Express 6(5):1565–1588. https://doi.org/10.1364/BOE.6.001565
9. Lindeberg T (2013) Scale-space theory in computer vision. Springer Science & Business Media, Kluwer Academic Publishers, Norwell
10. Liu XN, Yang XZ, Jin J, Li JS (2018) Self-adaptive signals separation for non-contact heart rate estimation from facial video in realistic environments. Physiol Meas 39(6):06NT01
11. Mikolajczyk K, Schmid C (2001) Indexing based on scale invariant interest points, vol 1. Proceedings Eighth IEEE International Conference on Computer Vision (ICCV), Vancouver, pp 525–531. https://doi.org/10.1109/ICCV.2001.937561
12. Monkaresi H, Bosch N, Calvo R et al (2017) Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans Affect Comput 8(1):15–28. https://doi.org/10.1109/TAFFC.2016.2515084
13. Piazzi A, Visioli A (2000) Global minimum-jerk trajectory planning of robot manipulators. IEEE Trans Ind Electron 47(1):140–149. https://doi.org/10.1109/41.824136
14. Poh MZ, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10762–10774. https://doi.org/10.1364/OE.18.010762
15. Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, Stein J, Hogan N (2002) Movement smoothness changes during stroke recovery. J Neurosci 22(18):8297–8304. https://doi.org/10.1523/JNEUROSCI.22-18-08297.2002

16. Soleymani M, Lichtenauer J, Pun T, Pantic M (2012) A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput 3(1):42–55. https://doi.org/10.1109/TAFFC.2011.25
17. Sun Y, Thakor N (2016) Photoplethysmography revisited: from contact to noncontact from point to imaging. IEEE Trans Biomed Eng 63(3):463–477
18. Temko A (2017) Accurate wearable heart rate monitoring during physical exercises using PPG. IEEE Trans Biomed Eng 64(9):2016–2024. https://doi.org/10.1109/TBME.2017.2676243
19. Tomasi C, Kanade T (1991) Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University
20. Tulyakov S, Alameda-Pineda X, Ricci E, Yin L, Cohn JF, Sebe N (2016) Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp 2396–2404. https://doi.org/10.1109/CVPR.2016.263
21. Verkruysse W, Svaasand LO, Nelson JS (2008) Remote plethysmographic imaging using ambient light. Opt Express 16(26):21434–21445. https://doi.org/10.1364/OE.16.021434
22. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, pp 511–518. https://doi.org/10.1109/CVPR.2001.990517
23. Wang W, den Brinker AC, Stuijk S, de Haan G (2017) Algorithmic Principles of Remote-PPG. IEEE Trans Biomed Eng 64(7):1479–1491. https://doi.org/10.1109/TBME.2016.2609282
24. Wang W, den Brinker AC, Stuijk S, de Haan G (2017) Amplitude-selective filtering for remote-PPG. Biomedical Optics Express 8(3):1965–1980. https://doi.org/10.1364/BOE.8.001965
25. Xu S, Sun L, Rohde GK (2014) Robust efficient estimation of heart rate pulse from video. Biomedical Optics Express 5(4):1124–1135. https://doi.org/10.1364/BOE.5.001124
26. Zhao F, Li M, Qian Y, Tsien JZ (2013) Remote measurements of heart and respiration rates for telemedicine. PLoS One 8(10):e71384. https://doi.org/10.1371/journal.pone.0071384




**Zhao Yang** received his B.S. degree in communication engineering and M.S. degree in electronics and communication engineering from Hefei University of Technology, Hefei, China, in 2010 and 2013, respectively. He is currently working toward the Ph.D. degree at Hefei University of Technology, Hefei, China. His current research interests include image processing, computer vision, and biomedical signal processing.

**Xuezhi Yang** received the B. Eng. degree from Anhui University, China, in 1992, the M. Eng. degree from Hefei University of Technology, China, in 1995, and the Ph.D. degree from The University of Hong Kong, China, in 2003. From 2006 to 2007, he was a postdoctoral research fellow in the Department of Systems Design Engineering at the University of Waterloo, Canada. He is currently a Professor in the School of Computer and Information at Hefei University of Technology. His research interests include digital image processing, pattern recognition and computer vision.

**Xiu Wu** received her B.S. degree in electronic and information engineering from Anhui Jianzhu University, Hefei, China, in 2007 and M.S. degree in signal and information processing from University of Electronic Science and Technology of China, Chengdu, China, in 2013. She is currently working toward the Ph.D. degree at Hefei University of Technology, Hefei, China. Her current research interests include image processing, pattern recognition, computer vision, and biomedical signal processing.