IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 2, 2018, accepted April 9, 2018, date of publication April 18, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2828133

Motion Resistant Image-Photoplethysmography Based on Spectral Peak Tracking Algorithm

BING-FEI WU[1], (Fellow, IEEE), PO-WEI HUANG[2], (Student Member, IEEE), CHUN-HSIEN LIN[2], (Student Member, IEEE), MENG-LIANG CHUNG[1], (Member, IEEE), TSONG-YANG TSOU[2], AND YU-LIANG WU[1]

[1]Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan
[2]Institute of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan

Corresponding author: Po-Wei Huang (abc1199281@cssp.cn.nctu.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 106-2622-E-009-009-CC2.

ABSTRACT Medical fields have seen increasing attention being given to image based heart rate measurement in recent years. One of the major limitations is motion artifacts of subject's head. Although there have been many studies focusing on signal extraction using different parameters and models, the development of frequency domain analysis is emerging slowly and moving in many directions. In the field of contact photoplethysmography (PPG), recent studies employed the acceleration signals to assist their spectral peak tracking algorithms. Inspired by the development of contact PPG, we are proposing a motion resistant spectral peak tracking (MRSPT) framework which eliminates the motion artifacts by integrating facial motion signals. The effectiveness of MRSPT coupled with the optimal image-based PPG (iPPG) signal has been tested against the state-of-the-art spectral peak tracking algorithms, multi-channel spectral matrix decomposition (MC-SMD), and the maximum peak selection coupled with optimal iPPG signal (Optimal MPS). Compared with MC-SMD and Optimal MPS, MRSPT uplifts the success rate-10 (success rate-5), the probability in which the absolute error is below ten (five) beats per mins, from 54.7% (36.3%) with MC-SMD and 73.0% (61.3%) with Optimal MPS to 90.7% (75.7%) with MRSPT in motion scenarios where subject moves arbitrarily with different distance or lighting. MRSPT also enhances the success rate-10 (success rate-5) from 40.7% (26.3%) with MC-SMD and 57.4% (45.7%) with Optimal MPS to 73.4% (58.4%) with MRSPT in all seven motion conditions including driving and running. Averagely, the success rate-five of Optimal MRSPT surpass the success rate-10 of both Optimal MPS and MC-SMD.

INDEX TERMS Biomedical signal processing, biomedical monitoring, heart rate, image sequence analysis, photoplethysmography (PPG), spectral peak tracking.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are prime causes of death globally. An estimated 17.7 million people died from CVDs in 2015, representing 31% of all deaths globally [1]. Among all of the cardiac activity indicators in the diagnosis of CVDs, heart rate (HR) is one of the most significant indicators used by medical professionals. Long-term HR monitoring can not only help physicians prescribe CVDs treatments but also detect potentially life-threating heart rhythm malfunctions. Conventionally, HR is measured via electrocardiography (ECG) or photoplethysmography (PPG). In order to measure ECG or PPG, patients are required to wear adhesive gel patches, finger clips, and/or chest straps which can lead to skin irritation and discomfort. In contrast, non-contact HR measurement technique serves as a great alternative when direct contact with subject's skin should be avoided (e.g., subjects with skin damage, neonates). Driven by the advantages of non-contact HR measurement, medical fields have seen increasing attention being given to non-contact HR measurement techniques in the recent years, including microwave Doppler radar [2], thermal imaging [3], and image-based PPG (iPPG) [4].

Although non-contact HR measurement technique provides many benefits, HR measurement techniques using Doppler radar or thermal imaging shares a common drawback—the requirement of specialized and expensive equipment. In contrast, by simply using a regular RGB camera, iPPG can detect human cardiac pulses from variations

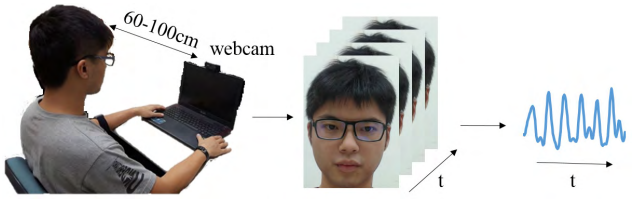


FIGURE 1. The common operating principle of iPPG.

of a person's skin color. The principle of iPPG is illustrated in Fig.1. A digital camera is directed at the subject's face. The ambient light is the light source. The webcam captures the skin tone variations caused by the blood-volume pulsation (BVP) and arterial transmural pressure during the cardiac cycles. In 2007, Allen described the principle of PPG operation and reviewed its typical clinical applications in physiological measurement [5]. In 2015, Kamshilin *et al.* [6] reported some thought-provoking observations and then proposed a new model for the essence of iPPG. The oscillations of transmural pressure can lead to deformations to adjacent tissues and then result in variations of reflected light intensity. Based on the iPPG algorithms, many advanced physiological monitors have been developed, including respiratory rate [7], blood pressure [8], peripheral oxygen saturation (SpO2) [9], [10], and heart rate variability (HRV) [11], etc. Applications in vision-based intelligent systems have also been invented including neonates monitoring [12], mental-stress detection [13], [14], subject detection [15], [16], and atrial fibrillation [17], etc.

Over the past several years, various core iPPG algorithms have been presented. In 2008, Verkruysse et al. used the green channel of a regular camera pointed at human skin to estimate HR [18]. Poh *et al.* [19] proposed a linear combination of all three channels to extract three independent signals with independent component analysis (ICA). To reduce the computational complexity of ICA, Lewandowska *et al.* [20] constructed a linear combination with a different technique, i.e., principle component analysis (PCA). Both of these techniques belong to the blind source separation category (BSS). In 2016, Cheng *et al.* [21] extended the BSS to Joint-BSS after taking the background images into consideration, and then proposed an illuminance variation-resistant algorithm based on independent vector analysis (IVA) and ensemble empirical mode decomposition (EMD). In 2017, Xu *et al.* [22] improved their illuminance variation-resistant framework by using partial least squares (PLS) and multivariate EMD. Qi *et al.* [23] proposed another type of J-BSS, connectivity multiset canonical correlation analysis (C-MCCA), and showed that C-MCCA outperforms ICA.

However, BSS based methods may suffer from the risk of treating other periodic signals such as motion or illuminance variation as HR signal [24], [25]. To eliminate this problem, several model based methods have been proposed. Haan and Jeanne proposed a chrominance-based iPPG (CHROM) by referencing a standardized skin tone to white-balance the camera [24]. Based on this study, Wang et al. exploited spatial redundancy to detect iPPG [25]. Haan, as well as Wenjin Wang et al., also developed two new ways to extract iPPG, spatial subspace rotation (2SR) [26] and plane orthogonal to skin (POS) [27]. With subject-dependent skin-color space analysis, 2SR can extract iPPG even in complicated illuminance conditions [26]. However, 2SR, an entirely data-driven algorithm, may bring about unreliable measurement owing to noisy or poorly-chosen skin-mask. POS resembles CHROM but reducing different main expected distortions in order [27]. The article [27] also reviews several iPPG recovery methods. Assuming ambient illuminance spectrum is constant, both POS and CHROM project the signal onto a fixed plane to reduce noise. To deal with varying spectrum and motion artifacts, Feng *et al.* [28] investigated the effect of region of interest (ROI) tracking for six different types of motion and proposed adaptive green red difference method (GRD). Nonetheless, almost all of the above studies have focused on the iPPG recovery stage as shown in Fig. 2. In reality, robustness and quality of iPPG can be enhanced by integrating frequency analysis. Sun *et al.* [29] employed a joint time-frequency analysis method which illustrated the feasibility to measure HR on relative low frames per second (20 fps). Cheng *et al.* [21] have developed frequency analysis method using EMD to overcome distortion caused by illuminance variation. Nevertheless, these frequency analysis methods did not focus on the issue of motion artifacts. Wang *et al.* [30] proposed a sub-band based method to deal with motion artifacts. Our previous work [31] studied the origins of residual noise after iPPG recovery algorithms and proposed a wavelet based method to cope with short duration motion noise.

To further enhance the performance of frequency analysis, we take reference from the development of contact PPG. In the field of contact PPG signal, frequency analysis methods like EMD, joint time-frequency analysis, or wavelet analysis are mainly proposed for clinical scenarios with small motions [32]. For strong motion scenarios, recent studies have been focusing on the framework of spectral peak tracking to further enhance the accuracy [32], [33]. In 2015, Zhang *et al.* [33] has developed a TROIKA framework to cope with motion artifacts. In 2016, Xiong *et al.* [32] employed the sparse properties and then proposed the state-of-the-art spectral peak tracking algorithm, multi-channel spectral matrix decomposition (MC-SMD). However, the requirement of iterations for each window can be time-consuming.

In this paper, we are proposing a motion resistant spectral peak tracking (MRSPT) method. Referring to related work as mentioned above, we extend our previous work, which studied the origins about residual noise [31], to further identify the motion artifacts on the spectrum of distorted iPPG signals. These signals include three state-of-the-art iPPG recovery algorithms (CHROM, GRD, POS). On the basis of these studies, we have developed MRSPT which can cope with strong and irregular motion artifacts. Because skin is a kind of Lambertian source, the motion of facial position

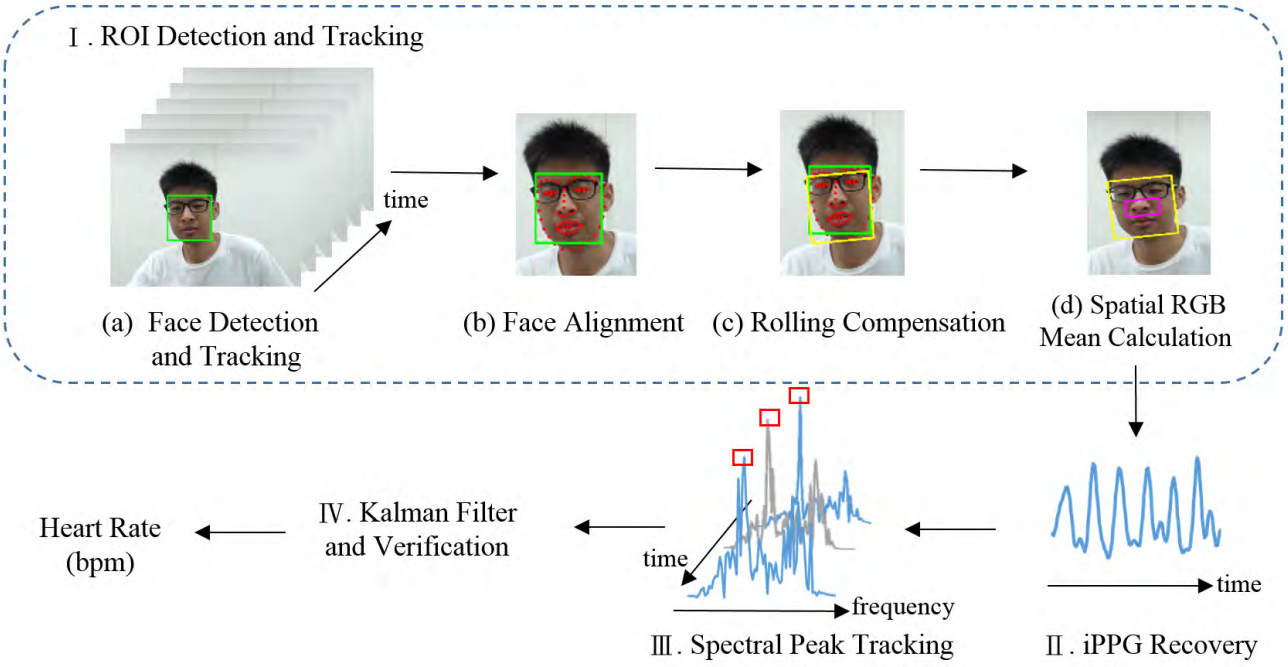


FIGURE 2. The flowchart of major steps involved in the proposed system.

is related to the motion artifacts. Hence, MRSPT exploits the x,y-coordinates of facial position to alleviate the motion artifacts. To the best of our knowledge, this is the first successful subject independent spectral peak tracking method applied in the field of iPPG.

To verify the effectiveness of MRSPT, we expand and modify MC-SMD from contact PPG to iPPG as the benchmark algorithm. In addition, MRSPT is also benchmarked against three state-of-the-art iPPG algorithms coupled with maximum peak selection (MPS). Moreover, we also test MRSPT with realistic conditions like driving conditions where motion artifacts occur irregularly. The feasibility of iPPG on driving conditions is rarely discussed before. The experiments show promising results and confirm the effectiveness of proposed MRSPT framework in comparison with MC-SMD and MPS.

In summary, the purpose of this paper is twofold: (a) to preliminarily identify and classify the spectrum distorted by motion artifacts. (b) to alleviate the motion artifacts and evaluate HR more accurately via spectral peak tracking system.

The organization of the paper is as follows. Section II begins with an overview of the entire system and introduces spectral peak tracking. Section III highlights the experimental setup and assessment details. Section IV discusses experimental results. Section V concludes the observations and sets forth the future work.

II. METHODOLOGY

As depicted in Fig. 2, the proposed system consists of four stages, Stage I—ROI Detection and Tracking, Stage II—iPPG recovery, Stage III—Spectral Peak Tracking, and Stage IV—Kalman Filter and Verification.

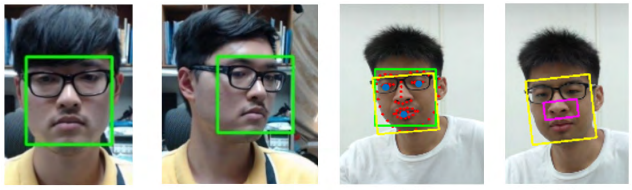


FIGURE 3. The green rectangle represents the results of face detection with (a) glasses and bangs, (b) yawing, and (c) rolling. The red dots indicate facial landmarks, and the blue dots denote the position of eyes and mouth. The yellow rectangles in (c) and (d) point out the rolling compensation. And the pink rectangle in (d) implies the ROI.

A. STAGE I—ROI DETECTION AND TRACKING

Litong noted that successful ROI tracking can compensate for several types of noise induced by subject's motions (i.e. yawing, pitching, rolling, and surging) [28]. Four major steps involved in ROI detection and tracking are described in Stage I of Fig. 2. First, we employ Dlib-ml face detection algorithm [34]. The reason for using Dlib-ml is its ability to detect face with glasses (Fig. 3a), bangs (Fig. 3a), yawing (Fig. 3b), and rolling (Fig. 3c).

Even so, face detection may fail for certain frames in some motion polluted scenarios like driving or exercising. In these cases, template matching method is applied to further help track the faces. Secondly, facial landmarks can be carried out via an ensemble of regression trees based on [35] (red dots in Fig. 3b). After averaging the feature points around eyes and mouth, we can acquire the center points of eyes and mouth (blue dots in Fig. 3c). Thirdly, to compensate for the effect of head's rolling, the detected facial rectangle will be rotated

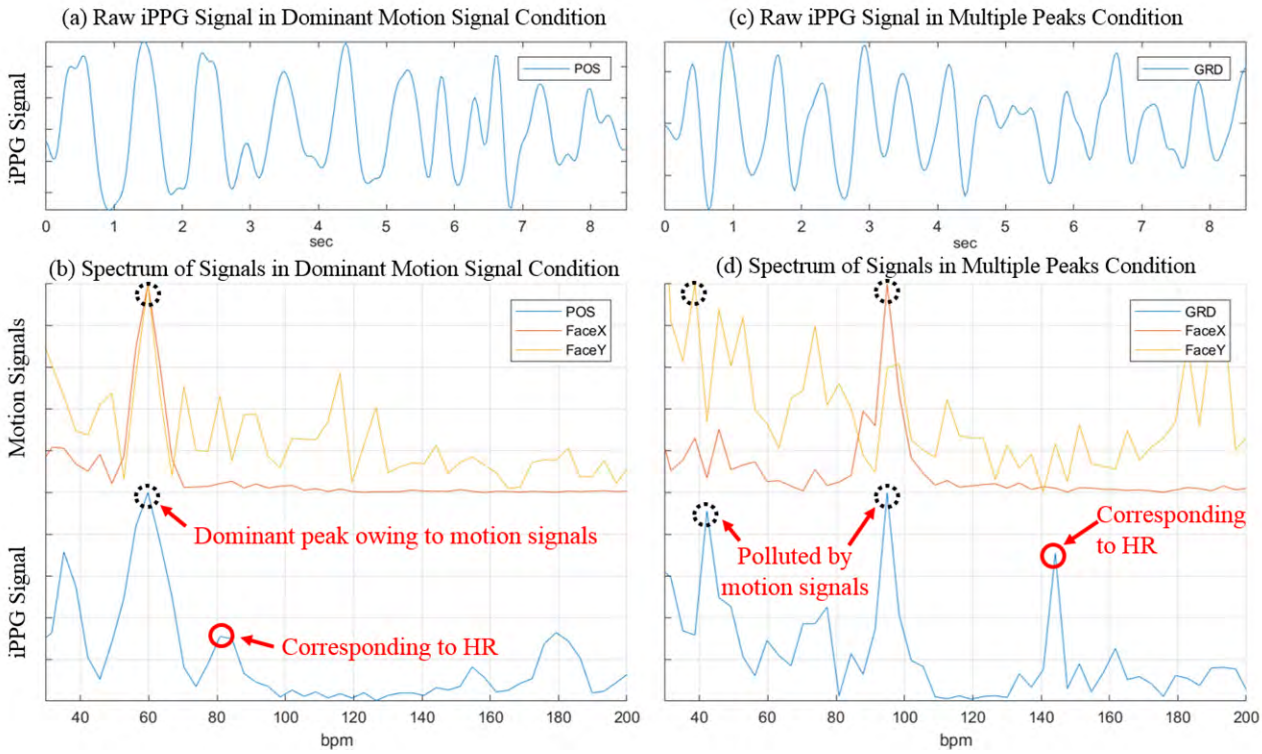


FIGURE 4. Raw iPPG/motion signals (a) and spectrum (b) in dominant motion signal condition occurs in Intentional Motion Scenario. Raw iPPG/motion signals (c) and spectrum (d) in multiple peaks condition occurs in Fitness Scenario. The iPPG signals are derived from POS and GRD algorithm respectively. FaceX and FaceY are frequency domain of facial position signals recorded each frame. The red circles indicate the true heart rate frequency. The black dotted circles indicate the peaks which are induced by motion signals (FaceX and FaceY). If we regard the highest bin as HR, these phenomena usually lead to measurement errors.

to parallel the line connecting the positions of eyes (yellow rectangle in Fig. 3c). To determine the size of ROI, the vertical length from eyes to mouth is defined as

$$L = [y_m - (y_{le} + Y_{re}/2)], \tag{1}$$

where y is the vertical position in rotated image; the subscripts m, le, and re represent the abbreviation of mouth, left eye and right eye respectively. Next, the size of ROI centered at middle point of face is empirically set as $1.3 \times L$ multiplied by $0.6 \times L$; the result is drawn in Fig. 3d. The reasons for selecting cheeks as ROI are twofold: (a) less influence from glasses, bangs and expression. (b) better Signal-to-Noise Ratio in comparison to forehead and mouth [36]. Lastly, the raw iPPG signal is acquired via taking the average of pixel values within the ROI as:

$$iPPG_{raw}(t) = \frac{Sum\ of\ pixel\ values}{width \times height} = \frac{\sum_{x,y \in ROI} P(x, y, t)}{(1.3 \times L) \times (0.6 \times L)}. \tag{2}$$

B. STAGE II–iPPG RECOVERY AND CASE STUDY ANALYSIS

The reason why pulse signal can be acquired by a webcam focused to skin is that a small fraction (<5%) [24] of incident light is absorbed by the microvascular network which varies with blood volume pulse. In view of this, several studies have demonstrated reliable iPPG recovery algorithms (i.e., CHROM, POS, and GRD). According to American National Standard for Cardiac Monitor [37], the measurement range of clinically approved HR monitor is between 30-200 beats per minute (bpm). A finite impulse response (FIR) band pass filter with 30 and 200 (bpm) cut-off frequency is applied to iPPG signal. HR can be preliminarily evaluated from the iPPG signal which has the highest bin when adopting Fast Fourier Transform (FFT). Because these techniques are basically linear combination of different channels, some residual noises might still emerge occasionally. Our previous work has investigated the reasons why these noises occur in some extreme situations [31]. The situations are listed and described as follows.

1) DOMINANT MOTION SIGNAL

Some face movements such as heaving and swaying cannot be eliminated by ROI tracking [24]. The motion artifacts bring about dominant motional peaks during some scenarios like running and driving as shown in Fig. 4 (a) and (b). Because the motion signal dominates the spectrum, simply considering maximum peak as HR usually leads to severe measurement error.

2) MULTIPLE PEAKS CONDITION

As illustrated in Fig. 4 (c) and (d), the iPPG signal is polluted by residual motion noises from several frequencies, bringing

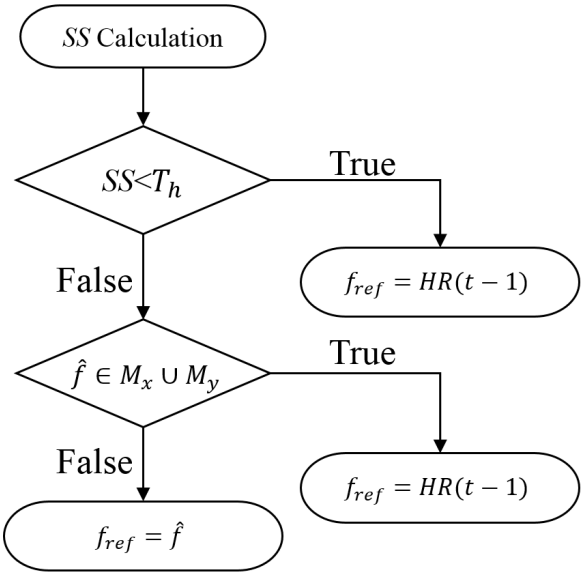


FIGURE 5. The flowchart of deriving reference frequency.

about multiple peak condition. Other causes like varying illuminance and motion blur may lead to similar artifact.

C. STAGE III–MOTION RESISTANT SPECTRAL PEAK TRACKING

To address the extreme conditions mentioned above, a motion resistant spectral peak tracking (MRSPT) algorithm is proposed with three steps: spectral stability calculation, reference frequency determination, and HR preliminary estimation. First, inspired from the spirit of signal-to-noise ratio (SNR), we introduce an index, spectral stability (*SS*), which is defined as follows

$$SS(t) = 10 \times \log\left(\frac{S^2(\hat{f}, t)}{\sum_{f=30}^{200}[S^2(f, t) - S^2(\hat{f}, t)]}\right), \quad (3)$$

where $\hat{f} = \arg max_f(S(f, t))$, f is frequency in beats per minute, S represents the frequency domain of iPPG at time t. The difference between SNR and *SS* is that $S(\hat{f}, t)$ is not guaranteed to be the real HR signal (e.g., dominant motion signal).

In order to distinguish the spectral peak between real signal and motion noises, two motion frequency sets (M_x, M_y) under two standard deviations from the estimated motion frequency are defined as

$$\begin{cases} M_x = [\hat{f}_x - 2\sigma_{ground-truth}, \hat{f}_x + 2\sigma_{ground-truth}] \\ M_y = [\hat{f}_y - 2\sigma_{ground-truth}, \hat{f}_y + 2\sigma_{ground-truth}], \end{cases} \quad (4)$$

where $\sigma_{ground-truth}$, which is determined from the experiment illustrated in Section III C (1), is the standard deviation of ground-truth HR sequences. Subscripts x and y indicate that the signal is acquired from face position and $\hat{f}$ is the frequency bin of highest energy. With two motion sets and *SS*, a peak selection method using reference frequency (f_{ref}) is illustrated in Fig. 5.

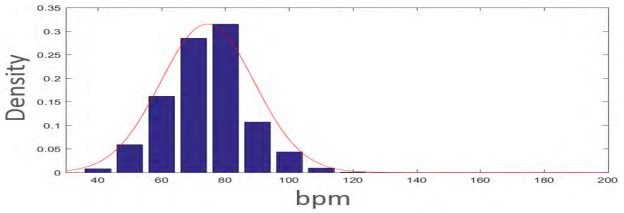


FIGURE 6. The distribution of total ground-truth HR. This result shows that ground-truth HR is distributed Gaussian. Consequently, we can apply Kalman Filter to track the estimated HR.

1) CASE 1

We developed a threshold T_h to identify multiple peaks condition. If $SS < T_h$, it represents that the spectrum is too chaotic to estimate real spectral peak. As a result, the reference frequency is previous estimated HR.

2) CASE 2

If *SS* is larger than threshold but highest spectral peak belongs to either one of the motion sets, it usually results from dominant motion signal. As a consequence, previous HR is considered as the reference frequency.

3) CASE 3

If *SS* is high and $\hat{f}$ does not belong to motion sets, the highest spectral peak is directly regarded as the reference frequency.

With reference frequency, we can derive a signal range under two standard deviations from the estimated HR, $\Delta_{HR} = [f_{ref} - 2\sigma_{ground-truth}, f_{ref} + 2\sigma_{ground-truth}]$. Next, the preliminary HR, $HR_p(t)$, can be calculated as

$$HR_p(t) = \frac{\sum_{f\in\Delta_{HR}} f \times S^2(f, t)}{\sum_{f\in\Delta_{HR}} S^2(f, t)}. \quad (5)$$

D. STAGE IV–KALMAN FILTER AND VERIFICATION

According to Fig. 6, the results show that recorded ground-truth HR is distributed Gaussian. Consequently, the state vector $X(t)$ at time t can be regarded as $HR_k(t)$. Moreover, considering the system input to be zero and both state transfer and observation model to be identity matrix, we can model the measured HR as

$$\begin{cases} X(t) = X(t-1) + W(\mathrm{t}) \\ Z(t) = X(t) + V(t), \end{cases} \quad (6)$$

where $Z(t)$ is the observed parameter (i.e. $HR_p(t)$), $W(t)$ is the processing noise with $N \sim (0, \sigma^2_{ground-truth})$ and $V(t)$ is observation noise with $N \sim (0, \sigma_p^2)$, where σ_p, which is derived from the experiment illustrated in Section III C (1), is the standard deviation of preliminary HR sequences. Using the Kalman Filter model [38], we can derive the optimal estimated pulse rate, $HR_k(t)$.

Occasionally, the peak selection algorithm may track the wrong peak owing to motion artifacts and spectrum perturbation. A verification step is developed to prevent the incorrect

tracking. Normally speaking, there is no huge HR variation (>10bpm) between relatively short time duration (0.5 sec); therefore, a regularization is proposed as follow:

$$HR(t) = \begin{cases} HR(t-1) + \Delta r(t), \\ \qquad if\ HR_k(t) - HR(t-1) > \Delta r(t) \\ HR(t-1) - \Delta r(t), \\ \qquad if\ HR_k(t) - HR(t-1) < -\Delta r(t) \\ HR_k(t), \\ \qquad otherwise, \end{cases} \tag{7}$$

where $\Delta r(t)$ is the maximum tolerated difference derived from recorded HR statistics and the variation of estimated HR(t) within the interval from t minus one minute to t. Precise definition is as follows:

$$\Delta r(t) = \sigma(t) \times \rho_{ground-truth}, \tag{8}$$

where $\sigma(t)$ is the standard deviation calculated from HR(t) sequence within the interval from t minus one minute to t and the tolerance ratio, $\rho_{ground-truth}$, represents the ratio of maximum instantaneous ground-truth HR difference to $\sigma_{ground-truth}$.

$$\rho_{ground-truth} = \frac{\max_t(|HR_{ground-truth}(t) - HR_{ground-truth}(t-1)|)}{\sigma_{ground-truth}}, \tag{9}$$

where the subscript $ground-truth$ represents that the parameters come from ground-truth heart rate.

III. ASSESSMENT DETAILS

To assess the proposed algorithm, this section presents the experimental setup, benchmark dataset, methods compared and evaluation metrics. First, we expressed general environmental setup throughout this work. Next, the benchmark dataset with different scenarios is introduced. Third, the detailed parameter of several state-of-the-art benchmark algorithms (MC-SMD, CHROM, GRD, and POS) will be presented as well. Finally, we adopted four evaluation metrics to assess the performance.

A. EXPERIMENTAL SETUP

All of the facial videos are captured under the environment listed in Table 1 with default setting of Logitech C920 webcam [39], [40]. In this study, we conducted eight realistic scenarios. The environment setup of each scenario are illustrated in Fig. 7, and the corresponding snapshots are shown in Fig. 8. Detailed scenario description is reported in Section III B.

B. BENCHMARK DATASET

A benchmark dataset containing more than one million frames has been built under different scenarios. We follow to a standard procedure, approved by Taipei Medical University. Each subject has given his/her informed consent before participating in the study. These subjects are recruited from National Chiao Tung University and none of them has history of cardiovascular issue. Total twenty volunteers with ages ranging from 21 to 32 years old participated in this study. Unless otherwise stated, the distance between the subject and the webcam ranges from 60 cm to 100 cm. The number in brackets indicates the amount of video frames recorded in the corresponding scenario.

Scenario 1 Stationary Case (Total 244,423 Frames and 13 Subjects): Subjects sitting in front of a webcam, about 0.5 meter distance, are asked to suppress any conscious movement. The duration for each trial is around 10 minutes. The illuminance source is normal fluorescent lamp and the luminance is about 200 lux.

TABLE 1. Experimental setup.

Item	Description
Webcam	Logitech C920
Quality	3 megapixels
Lens	Carl Zeiss lens
Correction	Automatic low-light correction
Resolution	VGA
Frame Rate	30 fps
Color Depth	8-bit
Image Format	bitmap
System Platform	ASUS X550JX
Ground-truth	
For Stationary Case	Alvital, AT101
For Motion Case	SCOSCHE, Rhythm+
Light Meter	TENMARS TM-205

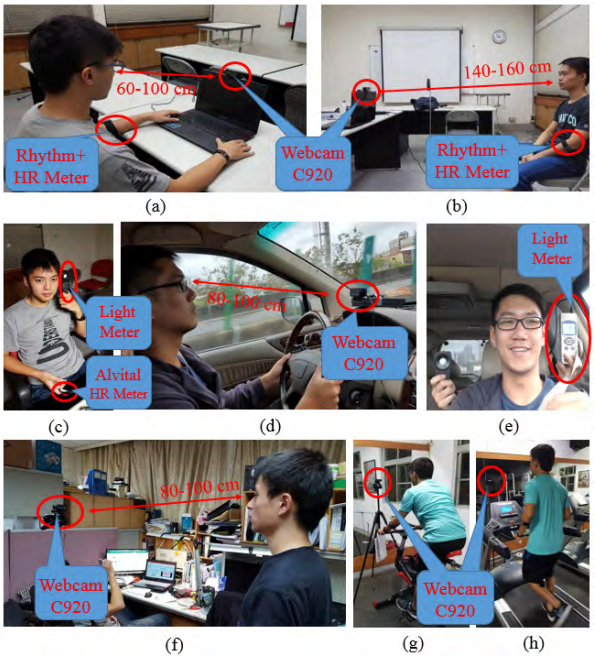


FIGURE 7. Environment Setups for each scenario. Stationary Case (Scenario 1) and Intentional Motion (Scenario 2) are set as (a). Long Distance (Scenario 3) and Illuminance Variation (Scenario 4) are depicted in (b) and (c) respectively. The setting of both Driving on Highway (Scenario 5) and Driving on Campus (Scenario 6) are illustrated in (d) and (e). The setting of Running in Place (Scenario 7) is depicted in (f). Fig. (g) and (h) represent the setting of Fitness (Scenario 8).

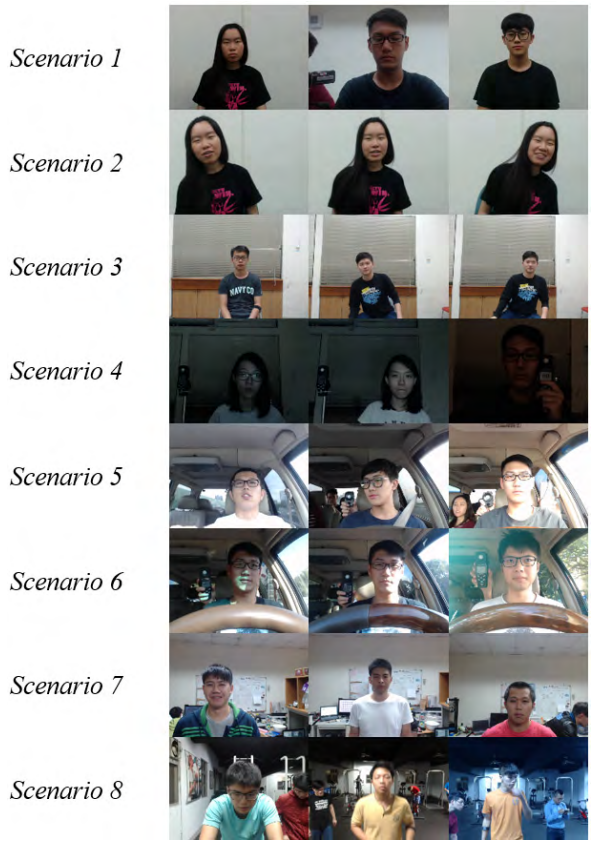


FIGURE 8. Snapshots of captured frames in the benchmark dataset. Detailed setup of each category is described in Section III B.

Scenario 2 Intentional Motion (Total 272,940 Frames and 13 Subjects): To explore the strong motion artifacts, the subjects are asked to move arbitrarily, including widely swinging their body, moving rapidly or slowly, etc. Motions like swaying is especially encouraged because ROI tracking lacks adaptability in this case [28]. All of the subjects and the other settings remain the same as described in Scenario 1.

Scenario 3 Long Distance(Total 35,392 Frames and 4 Subjects): To investigate the motion artifacts with smaller ROI, motions like swaying and swinging are also included. In comparison to Stationary Case (Scenario 1), the distance between subjects and webcam is adjusted to about 1.5 meters. Under the circumstance, ROIs are much smaller than those in Scenario 1. All of the other settings remain the same as in Scenario 1.

Scenario 4 Illuminance Variation(Total 95,493 Frames and 10 Subjects): To investigate the motion artifacts for iPPG signal under different illuminance conditions, motions like swaying and swinging are also included. In this scenario, videos are recorded with the same settings of the same camera (Logitech C920) under different illuminance conditions, ranging from 20 lux to 300 lux. Other settings and subject remain the same as in Stationary Case (Scenario 1).

Scenario 5 Driving on Highway(Total 169,279 Frames and 12 Subjects): Realistic driving situation is taken into consideration in the scenario, where both motion artifacts and luminance variations may occur simultaneously. In this scenario, the road is relatively flatter and straighter in comparison to Driving on Campus (Scenario 6). Mild motion of ROI are induced by such road conditions. At the same time, head movements (e.g., checking the side mirror) can also occur occasionally.

Scenario 6 Driving on Campus(Total 85,847 Frames and 9 Subjects): Driving around campus involves frequent direction changes which result in incident lighting from different angles. In this condition, signals may be polluted by irregular motion noise, changing luminance and varying illuminance spectrum, indicating a more challenging scenario compared with Driving on Highway Scenario (5). Motions are induced by several speed bumps throughout our driving route, as well as occasional head movements like checking the side mirror.

Scenario 7 Running in Place(Total 123,796 Frames and 12 Subjects): Subjects are required to run in place to simulate irregular motion artifacts and high HR condition. In the scenario, motion artifacts are especially strong because motions like running often leads to highly dynamic frequency with higher energy compared with those in previously mentioned scenarios. Owing to the limitation of webcam, the effect of motion blur also leads to more challenging scenario in comparison with Intentional Motion (Scenario 2). Other settings are the same as in Stationary Case (Scenario 1).

Scenario 8 Fitness(Total 112,796 Frames and 6 Subjects): Subjects are running on a treadmill or ridding on a flywheel in a gym. Compared with Running in Place (Scenario 7), in which ambient light is homogeneous, both motion artifacts and inhomogeneous illuminance are induced under such unsteady situation. As for motion artifacts, these actions involve more strenuous exercise in comparison to Running in Place (Scenario 7), which exhibits high-frequency motions at a larger scale. Other settings are the same as in the Stationary Case (Scenario 1).

C. METHODS COMPARED

The MRSPT is intended to be a post processing algorithm which addresses strong motion artifacts. Several benchmark details of methods compared issues are described as follows.

1) PARAMETER SELECTION

The proposed MRSPT method has three main parameters; the standard deviation of recorded ground truth HR ($\sigma_{ground-truth}$) and preliminary HR (σ_P), the threshold of SS (T_h). We derived all of these parameters using HR data in both Stationary Case (Scenario 1) and Intentional Motion (Scenario 2). For the remaining six scenarios, the preliminary HR is not accurate enough for its intended purpose. Then, the standard deviation are derived to be 4.35 ($\sigma_{ground-truth}$) and 5.3 (σ_P) respectively. Next, we found that if the power of maximum bin is lower than one third of the power in other bands, it is extremely likely that the multiple peaks condition occurs; therefore, the threshold of SS (T_h) is set as -4.77. Additionally, the maximum ground-truth HR difference is 2 bpm per second (1 bpm per output period). Referring to (9),

the tolerance ratio $\rho_{recorded}$, which is the ratio of maximum HR difference to $\sigma_{ground-truth}$, is set as 0.2174.

2) MAXIMUM PEAK SELECTION (MPS)

To verify the effectiveness of the proposed framework under different iPPG recovery algorithms, we compare the performance of MRSPT with maximum peak selection (MPS), which regards the highest spectral peak as HR. We test the difference between MRSPT and MPS with three state-of-the-art iPPG recovery algorithms (i.e., CHROM, GRD, and POS), but not with entire iPPG system, which means, the steps other than iPPG recovery like face detection/tracking strictly follows the method mentioned in Section 2.

All these algorithms are implemented with C++ code using Visual Studio 2015 and executing on a desktop computer with 4.2 GHz CPU (Intel-Core-i7) and 32-GB RAM. Both FFT order and window length are 512 points. The order of band pass filter is 128 and we output estimated HR every half a second. The default setting of frame rate is 30 fps, but frame rate may occasionally drop while encountering sudden illuminance variation (Scenario 4 and 6) due to limitations of the webcam. To fairly compare the methods, all the parameters remain the same while testing in different scenarios.

3) IMAGE BASED MC-SMD

Originally, the MC-SMD framework [32], which consists of two main parts, was developed to estimate contact PPG more accurately. The first part employs spectral matrix decomposition method to extract HR signal from the hybrid of two PPG channels and acceleration signals. Next, a well-designed spectral peak tracking algorithm is developed to enhance the extracted signal.

The problem of spectral matrix decomposition can be formulated as:

$$\begin{aligned} &\min_{P,Q} \ \|Y-\Phi(P+Q)\|_F^2+\lambda_1\|P\|_{1,2}+\lambda_2\|Q\|_{1,1}; \\ &s.t.: \ Y=\Phi(P+Q), \end{aligned} \tag{10}$$

where Y is the measurement matrix, P is the motion artifact matrix and Q is the desired reconstructed spectrum. To maintain the spirit of original MC-SMD method, we replace two PPG signals with bandpass POS and CHROM signals. Likewise, we substitute the acceleration signals with x, y-coordinates of facial ROI. All of the signals mentioned above are processed with 8 seconds window and 25% overlap. These measurements form a matrix of $Y \in \mathbb{R}^{M\times H}$, where M is the window size and $H=4$, consisting of 2 iPPG signals and 2 motion signals.

For spectral matrix $X=P+Q\in\mathbb{C}^{N\times H}$, the accelerated proximal gradient (APG) is adopted with parameters specified in the following for a better performance: DFT matrix $\Phi=\frac{1}{N}e^{j\frac{2\pi}{N}}$ with $N=1024$, $\lambda_1=\frac{0.2}{N}$, and $\lambda_1=\frac{4.1}{N}$. In addition, to derive the optimal solution considering general cases, we dropped the assumption of real-valued P, Q and then redefined the corresponding complex norm as $\|P\|_{1,2}=\sum_{i=1}^{N}\left(\sum_{j=1}^{R}\left|P_{i,j}\right|^2\right)^{1/2}$. Moreover, the corresponding new optimal solution for function $S_\epsilon(R)$ is used for complex variables, where

$$S_\epsilon(R)_{i,j}=\begin{cases}(1-\frac{\epsilon}{\left|R_{i,j}\right|})R_{i,j} & if \ \left|R_{i,j}\right|>\epsilon \\ 0 & otherwise.\end{cases} \tag{11}$$

Due to the limited quality of remote PPG and reconstructed sparse signals mentioned above, the thresholds of number of peaks and searching ranges are also modified as follows: $R_1=4$, $R_2=4$ and $R_3=4$, $threshold\ for\ pc1=1, threshold\ for\ pc2=1$. These parameter values are obtained and optimized by brute force searching. Although there are many essentially different characteristics between contact PPG and iPPG, we have tried our best to adapt the MC-SMD.

D. EVALUATION METRICS

The following list outlines two different types of evaluation metrics:

1) MAE, RMSE, OF PULSE RATE

The mean absolute error (MAE), root mean square error (RMSE) are applied to assess the performance of different estimators.

2) PRECISION AND SUCCESS RATE

In line with [22], the accuracy of algorithms can be assessed in term of "precision," the probability in which the absolute error between reference HR and estimated HR is below an error tolerance. In this paper, we evaluate the performance with a set of error tolerance $T\in[0,10]$ (bpm). In the precision curve, we are interested in the probability where the absolute error is under 5 or 10; the probabilities are hereafter defined as success rate-5 and success rate-10 respectively.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section consists of six parts to fairly evaluate the effectiveness of proposed MRSPT. We first reported the consistency of the best algorithm in each scenario between different metrics. Secondly, some observations with different algorithms (GRD, CHROM, POS, and MC-SMD) are discussed respectively. Next, the effectiveness of MRSPT is presented. Last comes to the overall comparison.

Several tables and figures are referenced throughout this section. Table 2-5 summarize the experimental results of MC-SMD, MPS and MRSPT with CHROM, GRD, and POS. The bold entries in each row denote the optimal algorithm in corresponding scenario. Fig. 9 illustrates the visualized version of table 2-5 for easier comparison and discussion. Also, a real-time experimental demo video is available at https://youtu.be/MyWYDkeHYx4. Table 6 denotes the time duration in the sample video for each scenario. The user interface in the sample video is also illustrated in Fig. 10.

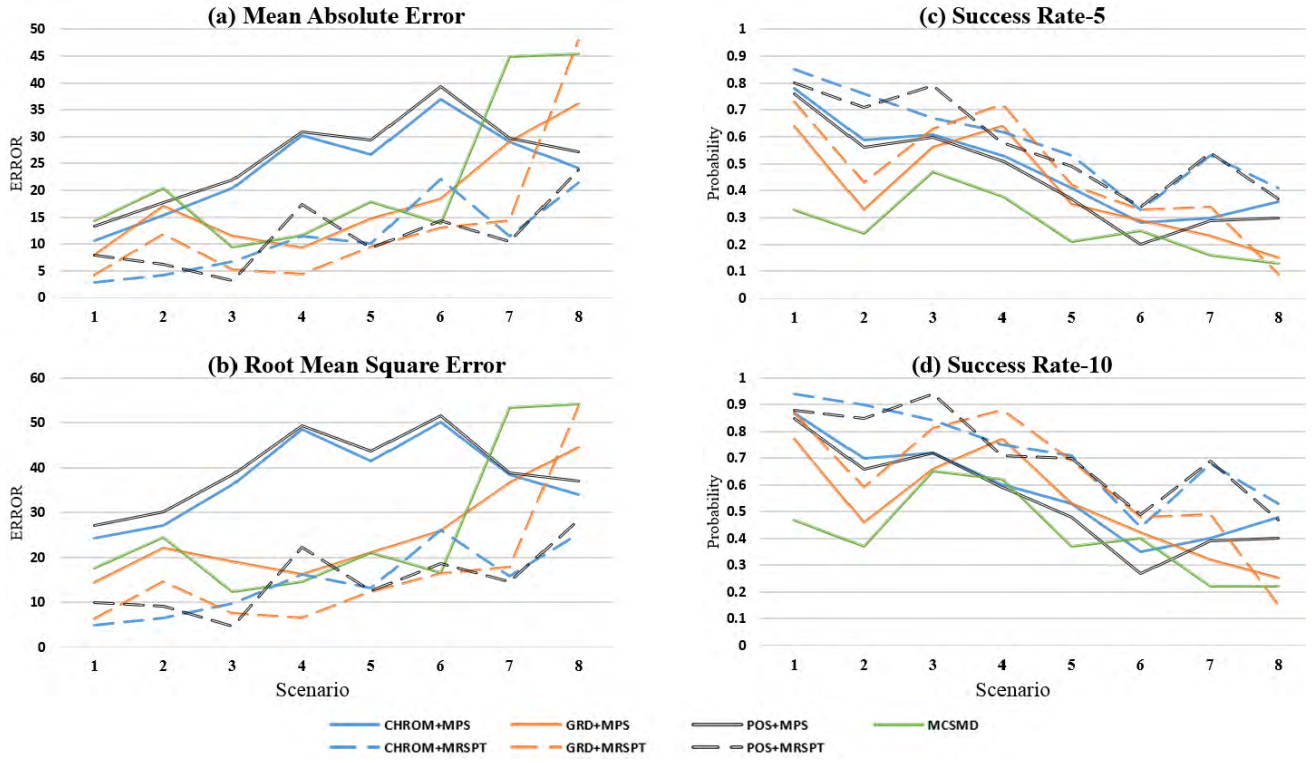


FIGURE 9. Table II-V are visualized as (a)-(d) for easier comparison and discussion. The number in the x-axis represents the order of Scenario. The dotted lines are the results of MRSPT with different iPPG recovery algorithms. In view of every metric, the best algorithms in each scenario are achieved integrating MRSPT.

TABLE 2. Mean absolute error.

Category	CHROM +MPS	GRD +MPS	POS +MPS	MCSMD	CHROM +MRSPT	GRD +MRSPT	POS +MRSPT
Stationary Case	10.63	7.92	13.26	14.3	**2.92**	4.30	7.90
Intentional Motion	15.30	17.11	17.67	20.39	**4.18**	11.8	6.25
Long Distance	20.42	11.48	21.91	9.45	6.73	5.29	**3.24**
Illuminance Variation	30.22	9.33	30.94	11.58	11.56	**4.46**	17.39
Driving on Highway	26.72	14.74	29.32	17.85	10.15	9.52	**9.22**
Driving on Campus	36.90	18.35	39.37	13.89	22.19	**13.07**	14.35
Running in Place	28.99	29.02	29.67	44.90	11.49	14.3	**10.42**
Fitness	24.22	36.19	27.26	45.48	**21.48**	48.03	23.8

unit: bpm

TABLE 3. Success rate 5.

Category	CHROM +MPS	GRD +MPS	POS +MPS	MCSMD	CHROM +MRSPT	GRD +MRSPT	POS +MRSPT
Stationary Case	0.78	0.64	0.76	0.33	**0.85**	0.73	0.80
Intentional Motion	0.59	0.33	0.56	0.24	**0.76**	0.43	0.71
Long Distance	0.61	0.56	0.6	0.47	0.67	0.63	**0.79**
Illuminance Variation	0.53	0.64	0.51	0.38	0.62	**0.72**	0.58
Driving on Highway	0.41	0.35	0.37	0.21	**0.53**	0.42	0.49
Driving on Campus	0.28	0.29	0.20	0.25	0.33	0.33	**0.34**
Running in Place	0.3	0.23	0.29	0.16	0.53	0.34	**0.54**
Fitness	0.36	0.15	0.3	0.13	**0.41**	0.09	0.37

TABLE 4. Root mean square error.

Category	CHROM +MPS	GRD +MPS	POS +MPS	MCSMD	CHROM +MRSPT	GRD +MRSPT	POS +MRSPT
Stationary Case	24.35	14.45	27.06	17.67	**4.86**	6.31	9.97
Intentional Motion	27.15	22.10	30.09	24.49	**6.44**	14.63	9.03
Long Distance	36.17	19.10	38.50	12.26	9.71	7.56	**4.67**
Illuminance Variation	48.53	16.24	49.28	14.51	16.15	**6.49**	22.26
Driving on Highway	41.41	21.05	43.66	20.98	13.16	**12.36**	12.46
Driving on Campus	50.12	25.89	51.62	16.70	26.21	**16.33**	18.69
Running in Place	38.39	36.71	38.79	53.34	**15.75**	17.83	14.61
Fitness	34.06	44.45	36.96	54.18	**25.31**	53.58	28.43

unit: bpm

TABLE 5. Success rate 10.

Category	CHROM +MPS	GRD +MPS	POS +MPS	MCSMD	CHROM +MRSPT	GRD +MRSPT	POS +MRSPT
Stationary Case	0.87	0.77	0.85	0.47	**0.94**	0.87	0.88
Intentional Motion	0.70	0.46	0.66	0.37	**0.90**	0.59	0.85
Long Distance	0.72	0.66	0.72	0.65	0.84	0.81	**0.94**
Illuminance Variation	0.60	0.77	0.59	0.62	0.75	**0.88**	0.71
Driving on Highway	0.53	0.53	0.48	0.37	**0.71**	0.69	0.70
Driving on Campus	0.35	0.42	0.27	0.40	0.44	0.48	**0.49**
Running in Place	0.40	0.32	0.39	0.22	0.68	0.49	**0.69**
Fitness	0.48	0.25	0.40	0.22	**0.53**	0.15	0.47

A. CONSISTENCY

On the whole, our benchmark metrics possess both consistency and discrimination among algorithms. The best algorithm in each scenario consistently wins the first place in all of the four benchmark metrics except for the driving scenarios (Scenario 6 and 7). For instance, MRSPT with CHROM has the lowest MAE/RMSE, as well as highest success rate-5 and -10 in Stationary Case (Scenario 1). On the other hand, the rather incongruent outcomes of MRSPT in driving scenarios (Scenario 6 and 7) result from the similar

TABLE 6. Time duration for each scenario in the sample video.

Category	from	to
Stationary Case	00:53	01:00
Intentional Motion	01:01	01:16
Long Distance	01:17	01:38
Illuminance Variation	01:39	01:52
Driving on Highway	01:53	02:11
Driving on Campus	02:11	02:30
Running in Place	02:31	02:39
Fitness	02:40	02:50

Format: mm:ss

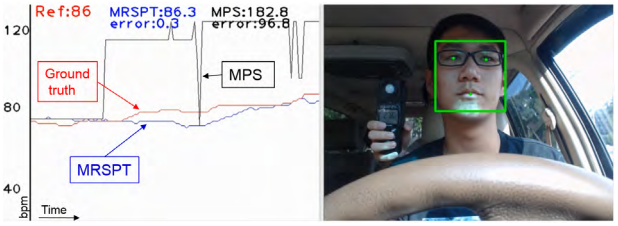


FIGURE 10. The texts within a rectangle and an arrow are only legend. The rest of this figure is the user Interface of sample video. The red, blue, and dark line indicated the ground-truth, MPS and MRSPT framework respectively. On the right hand side is the corresponding captured image and detected face.

performance using different iPPG recovery algorithms: (1) For the Driving on Campus (Scenario 6), the performance of MRSPT with CHROM, GRD, and POS are 0.33, 0.33, 0.34 in success rate-5, and 0.44, 0.48, 0.49 in success rate-10 respectively. (2) As for the Driving on Highway (Scenario 5), MRSPT with CHROM, GRD, and POS carry out 10.15, 9.52, 9.22 in MAE, and 0.71, 0.69, 0.70 in success rate 10 respectively. The major cause of this result is the fact that not only motion but also varying lighting conditions lead to complexities in motion, specular and intensity noises. Under such challenging circumstances, it is difficult to select one algorithm that outdoes the others in every benchmark metric. To sum up, the metrics provide a set of consistent and discriminative baseline to assess the performance among different algorithms.

B. GRD

GRD yields outcomes better than the others only in the Illuminance Variation (Scenario 4) among iPPG recovery algorithms supported by MRSPT. The prime reason for such exclusive results lies in its employment of the adaptive color difference method, which effectively improves the HR estimation under continuously varying illuminance. Both POS and CHROM assume consistent illuminance spectrum and in consequence perform less satisfactorily in varying illuminance condition.

Generally speaking, algorithms with MRSPT yield results better than all of those without MRSPT, but GRD in Illuminance Variation (Scenario 4) is one of the few exceptions. Interestingly, in the Illuminance Variation (Scenario 4), GRD with maximum peak selection is ranked as the first runner up in MAE, success rate-5 and -10. This result once again illustrates the importance, as well as effectiveness, of frame-dependent parameters of GRD in such a dynamic illuminance environment.

C. CHROM/POS

Because POS resembles CHROM but reducing different main expected distortions in order, we are interested in the similarity and difference of their performance. In view of [27], CHROM reveals weakness in continual specular distortion. On the other hand, POS is sensitive in nonhomogeneous luminance over ROI. The phenomenon mainly results from their assumptions of homogeneity in luminance and projection of light source in order to alleviate the artifacts. Such characteristics are manifested in Running in Place (Scenario 7) and Fitness (Scenario 8). For Running in Place (Scenario 7), regular indoor light serves as the light source and thus leads to better outcomes in POS; nonetheless, nonhomogeneous luminance makes POS inferior to CHROM in Fitness (Scenario 8). Likewise, for Long Distance (Scenario 3), the relatively smaller ROI makes iPPG signals more vulnerable to intensity distortion because the image captures less light reflected from the subject's skin but more ambient light. POS performs better because it eliminates intensity distortion first. In contrast, CHROM eliminates specular distortion first and leaves the intensity distortion as residual noise. The experimental results confirm the characteristics of CHROM and POS.

D. MC-SMD

In all of the four benchmarks, MC-SMD is not as competitive as expected. One reason for this could be the fundamental assumption in sparsity of heartbeat signals. We found that, in some challenging scenarios, multiple peaks condition contradicts the sparsity assumption in spectrum of MC-SMD. Furthermore, in the application of contact PPG, two signals are derived from nearby positions of body, suffering from similar motion artifacts. Nonetheless, among all of the iPPG signal recovery algorithms, resistance to similar motion artifacts varies from algorithm to algorithm, which also contributes to inferiority. The above-mentioned two reasons provide preliminary explanation to the unsatisfactory results.

On the other hand, it is noteworthy that the iterative nature of SMD goes against real-time applications. In contrast, MRSPT, based on characteristics of iPPG, yields enhanced results without iteration. As a consequence, it is easier to transfer original PC based MRSPT to embedded systems for broader applications.

E. EFFECTIVENESS OF MRSPT

The effectiveness of MRSPT can be verified with the following three reasons: (a) all of the optimal algorithms in each scenario are achieved by integrating with MRSPT (see Fig. 9), (b) generally speaking, MRSPT suppress MPS and MC-SMD, and (c) the worst chosen iPPG recovery

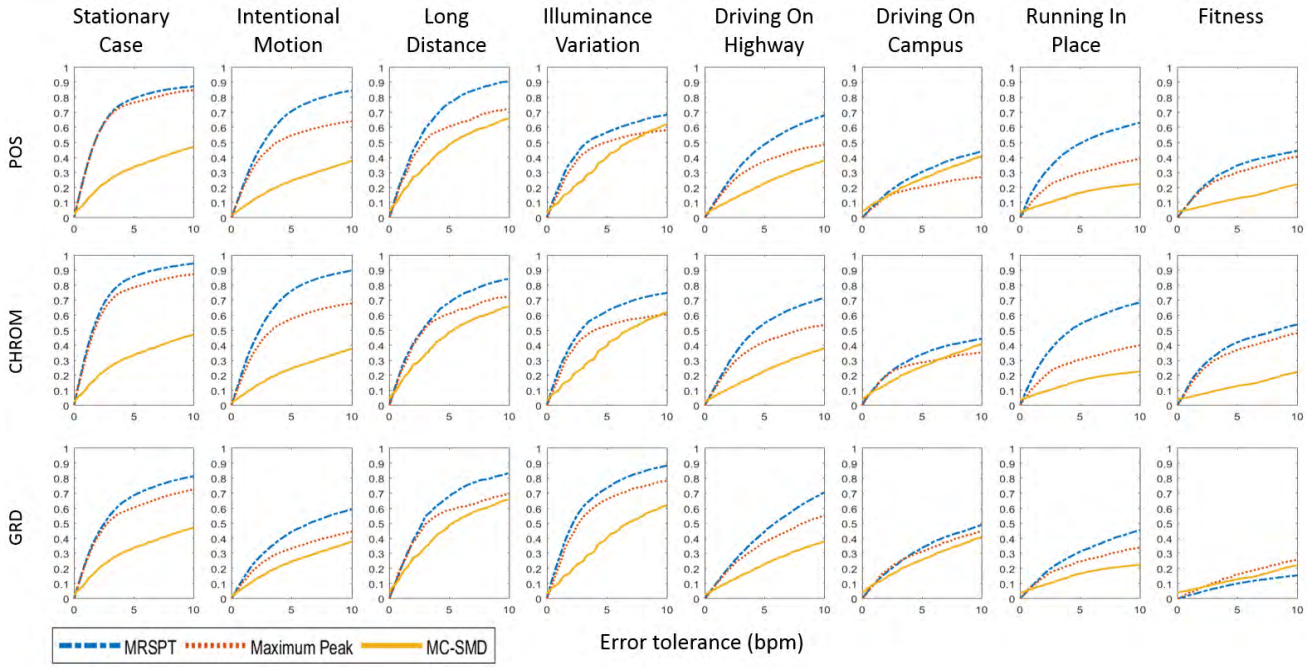


FIGURE 11. Precision curves (cumulative distribution function) of CHROM, GRD, and POS in each challenging scenarios. The x-coordinate represents absolute error between reference HR and estimated HR, and the y-coordinate indicates probabilities in which the absolute error is below the corresponding tolerance.

methods coupled with MRSPT can frequently outperform all the others using MPS.

Although there are a few cases that MRSPT seems to be falling behind MPS with different iPPG recovery algorithms, these are reasonable and can be explained. One of the few exceptions is afore-mentioned GRD in Illuminance Variation (Scenario 4) benchmarked with success rate-5 and -10. The other exceptions are CHROM algorithm in Stationary Case (Scenario 1), Intentional Motion (Scenario 2) and Fitness (Scenario 8). In these cases, all of the metrics possess relatively high value, representing high success rate-5 and -10, with high MAE and RMSE simultaneously. As shown in Fig. 12 estimated HR fluctuates owing to selecting the maximum peak, leading to severe fluctuation between real HR and incorrectly selected peaks. In consequence, the RMSE and MAE are high due to occasionally severe incorrectly selected peaks although the success rate-5 and -10 are relative high as well. In these cases, the slightly higher success rate of MPS do not imply better performance compared with the much lesser fluctuation while using MRSPT. In summary, MRSPT still outperforms MPS in these cases.

F. OVERALL COMPARISON

The precision curves of all scenarios and algorithms are drawn in Fig. 11. With the curves, first we compare the difference between Driving on Highway (Scenario 5) and Driving on Campus (Scenario 6). Next, we describe the reason why POS/CHROM suppress GRD with MRSPT but fall behind them while using MPS. Third, we briefly discuss the influence of lighting variation in different motion scenarios. Lastly, we conclude the performance with MRSPT.

We found that, whether MRSPT is applied or not, CHROM, GRD and POS in Driving on Campus (Scenario 6) consistently result in lower success rate in comparison with Driving on Highway (Scenario 5). It is because the road condition of highway is much straighter, smoother and flatter than country roads and hilly roads on campus. Additionally, incident light while Driving on Highway (Scenario 5) is more consistent than that while Driving on Campus (Scenario 6).

On the other hand, because GRD dynamically changes the parameters of linear combination, GRD with MPS is less influenced by above-mentioned fluctuating HR phenomenon and thus yields the lowest MAE/RMSE in most MPS cases. Nonetheless, GRD with MRSPT only wins in the Illuminance Variation (Scenario 4). It is because MRSPT can eliminate the fluctuating HR phenomenon of original CHROM, POS signals, leading to significant improvement.

Interestingly, in Fig. 11, we observe that MRSPT enhances success rate much more obviously in relatively consistent lighting scenarios, including Intentional Motion (Scenario 2), Long Distance (Scenario 3), Driving on Highway (Scenario 5), and Running in Place (Scenario 7); nonetheless, relatively less improvement occurs in Illuminance Variation (Scenario 4), Driving on Campus (Scenario 6) and Fitness (Scenario 8). Although it illustrates the inferior improvement under combination of motion artifacts and illuminance variation, the promising progress in the former scenarios once again guarantees the effectiveness of MRSPT in wide variety types of motion artifacts.

To conclude the effectiveness of MRSPT during motion scenarios, we take the average of success rate-5 (left) and success rate-10 (right) with respect to different algorithms in

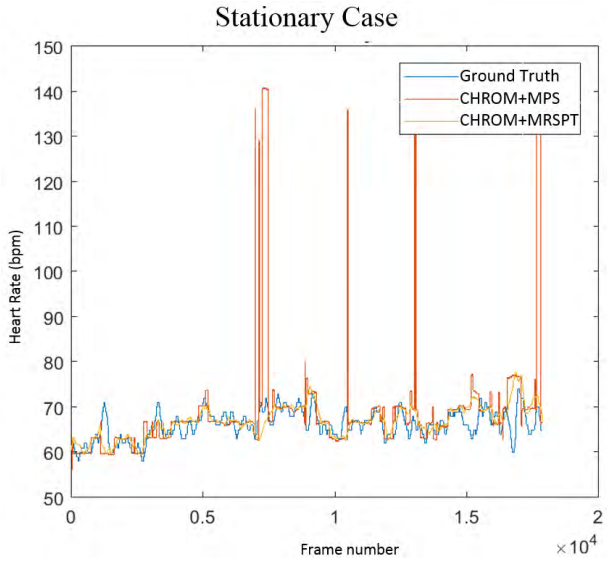


FIGURE 12. A sample of estimated HR using different algorithms. The x-axis represents the time in frame number and y-axis denotes HR (bpm). The blue, red and yellow lines represent the Ground Truth, estimated HR using MPS and estimated HR using MRSPT respectively.

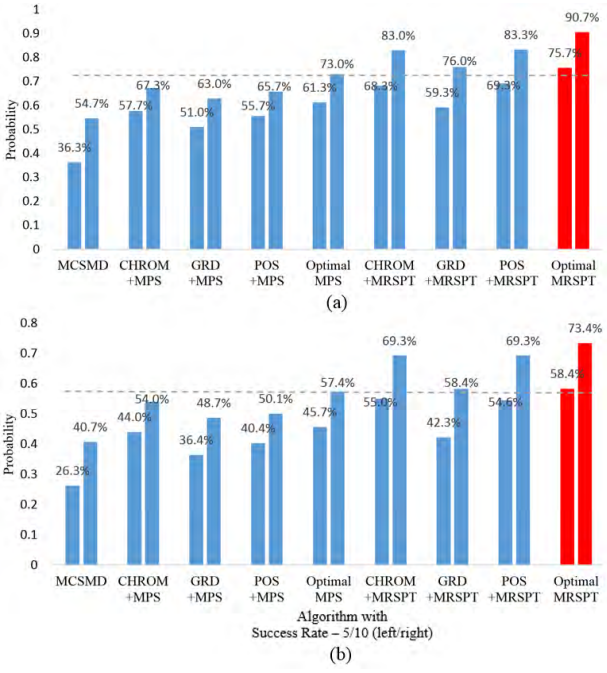


FIGURE 13. The bar charts illustrate the success rate-5 (left) and success rate-10 (right) with respect to different algorithms in experimental motion conditions (a) and total motion conditions (b). The 'Optimal' indicates that the results are derived from the optimal iPPG recovery algorithm for the framework (MPS or MRSPT) in each scenario. Please note that MRSPT coupled with any iPPG recovery algorithm outperforms the Optimal MPS. Moreover, the success rate-5 of Optimal MRSPT surpass success rate-10 of Optimal MPS.

three experimental motion conditions, including Intentional Motion (Scenario 2), Long Distance (Scenario 3), and Illuminance Variation (Scenario 4) as shown in Fig. 13. Likewise, we also take the average of all seven motion conditions including Driving on Highway (Scenario 5) and Running in Place (Scenario 7). The 'Optimal' indicates that the results are carried out with the optimal iPPG recovery algorithm in each scenario for the framework (MPS or MRSPT). Compared with MC-SMD and Optimal MPS, MRSPT enhances the success rate-10 (success rate-5) from 54.7% (36.3%) with MC-SMD and 73.0% (61.3%) with Optimal MPS to 90.7% (75.7%) with MRSPT in experimental motion scenarios where subject moves arbitrarily with different distance or lighting conditions. As for all motion conditions, MRSPT raises the success rate-10 (success rate-5) from 40.7% (26.3%) with MC-SMD and 57.4% (45.7%) with Optimal MPS to 73.4% (58.4%) with MRSPT. Moreover, it is noteworthy that the MRSPT coupled with arbitrary iPPG recovery algorithm defeats the Optimal MPS in both experimental motion conditions and realistic motion cases. Interestingly, the success rate-5 of Optimal MRSPT surpass success rate-10 of Optimal MPS. A real-time experimental demo video is available at https://youtu.be/MyWYDkeHYx4.

V. CONCLUSION

We have proposed a motion resistant image based HR monitoring system, MRSPT. Our analysis preliminarily identify and classify the spectrum distorted by motion artifacts. Next, inspired by current studies of contact PPG, we are proposing an entire framework which uplifts the accuracy of HR estimation. To the best of our knowledge, this is the first successful subject independent spectral peak tracking method applied in the image based PPG field. Our algorithm has been tested and verified to be effective among eight different challenging scenarios and compared to the three different state-of-the-art iPPG recovery algorithms. It is noteworthy that MRSPT coupled with any iPPG signal outperforms the MPS coupled with optimal chosen iPPG signal in both experimental motion conditions and realistic motion cases. The results show that MRSPT represents a promising framework to estimate HR from a regular webcam in a wide variety of motion scenarios.

REFERENCES

[1] World Health Organization. (2017). *Cardiovascular diseases (CVDs) Fact Sheet*. Accessed: Feb. 28, 2018. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/
[2] A. Høst-Madsen, N. Petrochilos, O. Boric-Lubecke, V. M. Lubecke, B.-K. Park, and Q. Zhou, "Signal processing methods for Doppler radar heart rate monitoring," in *Signal Processing Techniques for Knowledge Extraction and Information Fusion*. Springer, 2008, pp. 121–140.
[3] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1418–1426, Aug. 2007.
[4] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.
[5] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, p. R1-39, Feb. 2007.
[6] A. A. Kamshilin *et al.*, "A new look at the essence of the imaging photoplethysmography," *Sci. Rep.*, vol. 5, May 2015, Art. no. 10494.
[7] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, "Estimation of respiratory rate from photoplethysmographic imaging videos compared to pulse oximetry," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1331–1338, Jul. 2015.
[8] I. C. Jeong and J. Finkelstein, "Introducing contactless blood pressure assessment using a high speed video camera," *J. Med. Syst.*, vol. 40, no. 4, p. 77, 2016.

[9] D. Shao *et al.*, "Noncontact monitoring of blood oxygen saturation using camera and dual-wavelength imaging system," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 6, pp. 1091–1098, Jun. 2016.

[10] W. Verkruysse, M. Bartula, E. Bresch, M. Rocque, M. Meftah, and I. Kirenko, "Calibration of contactless pulse oximetry," *Anesthsia Analgesia*, vol. 124, pp. 136–145, Jan. 2017.

[11] K. Alghoul, S. Alharthi, H. Al Osman, and A. El Saddik, "Heart rate variability extraction from videos signals: ICA vs. EVM comparison," *IEEE Access*, vol. 5, pp. 4711–4719, 2017.

[12] L. K. Mestha, S. Kyal, B. Xu, L. E. Lewis, and V. Kumar, "Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 3817–3820.

[13] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 2957–2960.

[14] B. Kaur, S. Moses, M. Luthra, and V. N. Ikonomidou, "Remote stress detection using a visible spectrum camera," *Independ. Compon. Analyses, Compress. Sampling, Large Data Analyses, Neural Netw., Biosyst., Nanoeng. XIII*, vol. 9496, p. 949602, 2015.

[15] W. Wang, S. Stuijk, and G. de Haan, "Unsupervised subject detection via remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2629–2637, Nov. 2015.

[16] W. Wang, S. Stuijk, and G. de Haan, "Living-skin classification via remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2781–2792, Dec. 2017.

[17] J.-P. Couderc *et al.*, "Detection of atrial fibrillation using contactless facial video monitoring," *Heart Rhythm*, vol. 12, no. 1, pp. 195–201, 2015.

[18] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.

[19] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[20] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2011, pp. 405–410.

[21] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1422–1433, Sep. 2017.

[22] L. Xu, J. Cheng, and X. Chen, "Illumination variation interference suppression in remote PPG using PLS and MEMD," *Electron. Lett.*, vol. 53, no. 4, pp. 216–218, 2017.

[23] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomed. Signal Process. Control*, vol. 31, pp. 309–320, Jan. 2017.

[24] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[25] W. Wang, S. Stuijk, and G. de Haan, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015.

[26] W. Wang, S. Stuijk, and G. de Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1974–1984, Sep. 2016.

[27] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.

[28] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

[29] Y. Sun, S. Hu, V. Azorin-Peris, R. Kalawsky, and S. E. Greenwald, "Noncontact imaging photoplethysmography to effectively access pulse rate variability," *J. Biomed. Opt.*, vol. 18, no. 6, p. 061205, 2013.

[30] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, pp. 1023–1044, May 2017.

[31] B.-F. Wu, P.-W. Huang, T.-Y. Tsou, T.-M. Lin, and M.-L. Chung, "Camera-based heart rate measurement using continuous wavelet transform," presented at the Int. Conf. Syst. Sci. Eng. (ICSSE), Ho Chi Minh City, Vietnam, Jul. 2017, pp. 7–11.

[32] J. Xiong, L. Cai, D. Jiang, H. Song, and X. He, "Spectral matrix decomposition-based motion artifacts removal in multi-channel PPG sensor signals," *IEEE Access*, vol. 4, pp. 3076–3086, 2016.

[33] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.

[34] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.

[35] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. 27th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1867–1874.

[36] G. Lempe, S. Zaunseder, T. Wirthgen, S. Zipser, and H. Malberg, "ROI selection for remote photoplethysmography," in *Bildverarbeitung für die Medizin*. Springer, 2013, pp. 99–103.

[37] *Cardiac Monitors, Heart Rate Meters, and Alarms*, Standard ANS1/AAMI (EC13-1983), Arlington, TX, USA, 1984.

[38] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

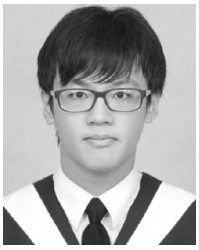
[39] Logitech International S.A. (2018). *HD-Pro Webcam C920 Specification*. Accessed: Feb. 28, 2018. [Online]. Available: http://support.logitech.com/en_us/product/hd-pro-webcam-c920/specs

[40] Logitech International S.A. (2018). *HD-Pro Webcam C920 Data Sheet*. Accessed: Feb. 28, 2018. [Online]. Available: https://www.arp.nl/webmedias/datasheet/575183859ce969320b1a2beb.pdf

BING-FEI WU (M'92–SM'02–F'12) received the B.S. and M.S. degrees in control engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1992. Since 1992, he has been with the Department of Electrical and Computer Engineering, where he was promoted to be a Professor in 1998 and a Distinguished Professor in 2010, respectively. He serves as the Director of the Institute of Electrical and Control Engineering, NCTU, in 2011. His research interests include image recognition, vehicle driving safety and control, intelligent robotic systems, intelligent transportation systems, and multimedia signal analysis. He founded and served as the Chair of Taipei Chapter of the IEEE Systems, Man and Cybernetics Society (SMCS) in 2003. He has been the Chair of the Technical Committee on Intelligent Transportation Systems of the IEEE SMCS since 2011. He is an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS and an Editor-in-Chief of the *International Journal of Computer Science and Artificial Intelligence*.

Dr. Wu is an IET fellow and CACS fellow. He was a recipient of many research honors, including the Outstanding Research Award of the Ministry of Science and Technology, Taiwan, in 2015; the Technology Invention Award of Y. Z. Hsu Scientific Award from Y. Z. Hsu Foundation in 2014; the National Invention and Creation Award of the Ministry of Economic Affairs, Taiwan, in 2012 and 2013, respectively; the Outstanding Research Award of Pan Wen Yuan Foundation in 2012; the Best Paper Award in The 12th International Conference on ITS Telecommunications in 2012; the Best Technology Transfer Contribution Award from National Science Council, Taiwan, in 2012; and the Outstanding Automatic Control Engineering Award from the Chinese Automatic Control Society in 2007.

PO-WEI HUANG (S'17) received the B.S. degree in electrical and computer engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 2017, where he is currently pursuing the Ph.D. degree with the Institute of Electrical and Control Engineering. His research is focusing on image based vital signs measurement. He was a recipient of the IEEE SMC Outstanding B.Sc. Theses Grant in 2017.



CHUN-HSIEN LIN (S'15) was born in Taipei, Taiwan. He received the B.S. degree in electrical and computer engineering and the M.S. degree in electrical control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include computer vision and machine learning.

MENG-LIANG CHUNG received the B.S. degree in naval engineering from National Taiwan Ocean University, Keelung, Taiwan, in 2005, the M.S. degree in electrical engineering from the Chung Cheng Institute of Technology, National Defense University, Taoyuan, Taiwan, in 2007, and the Ph.D. degree in electrical control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2016. He is interested in computer vision and intelligent traffic systems. His research interests include artificial intelligence, image processing, image compression, document segmentation, computer vision, and applications to intelligent transportation systems.

TSONG-YANG TSOU received the B.S. degree in electrical and computer engineering and the M.S. degree in control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2014 and 2017, respectively. His research is focused on image based photoplethysmography.

YU-LIANG WU is currently pursuing the B.S. degree of electrical engineering with National Chiao Tung University, Hsinchu, Taiwan.

His research interests include biomedical signal processing and machine learning.