Modulation Model of the Photoplethysmography Signal for Vital Sign Extraction

Mingliang Chen, *Student Member, IEEE*, Qiang Zhu, *Student Member, IEEE*, Min Wu, *Fellow, IEEE*, and Quanzeng Wang

***Abstract*—This paper introduces an amplitude and frequency modulation (AM-FM) model to characterize the photoplethysmography (PPG) signal. The model indicates that the PPG signal spectrum contains one dominant frequency component – the heart rate (HR), which is guarded by two weaker frequency components on both sides; the distance from the dominant component to the guard components represents the respiratory rate (RR). Based on this model, an efficient algorithm is proposed to estimate both HR and RR by searching for the dominant frequency component and two guard components. The proposed method is performed in the frequency domain to estimate RR, which is more robust to additive noise than the prior art based on temporal features. Experiments were conducted on two types of PPG signals collected with a contact sensor (an oximeter) and a contactless visible imaging sensor (a color camera), respectively. The PPG signal from the contactless sensor is much noisier than the signal from the contact sensor. The experimental results demonstrate the effectiveness of the proposed algorithm, including under relatively noisy scenarios.**

***Index Terms*—Modulation, photoplethysmography, vital sign.**

I. Introduction

MONITORING vital signs, such as heart rate (HR) and respiratory rate (RR), is essential in understanding a patient's physiological condition and monitoring and diagnosing diseases related to cardiovascular and lung functions. Whereas electrocardiography (ECG) is the gold standard technology used to study a patient's cardiovascular conditions, the portability, complexity, and cost of ECG equipment limit the breadth and prevalence of its use in health care, especially home care. Finding a feasible approach to track multiple vital signs from a simple, accessible, and easy-to-use sensor is desirable in daily health monitoring, especially in emerging mobile health (mHealth) care.

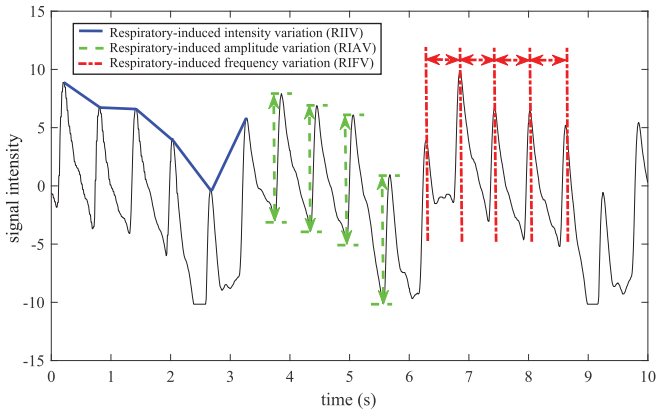

Fig. 1. Three kinds of respiratory-induced variations in the PPG signal.

Photoplethysmography (PPG) is a popular health-care technique used in clinical settings to capture vital signs by detecting blood content and volume changes in the microvascular bed of tissue. Pioneering work has proven the feasibility of extracting such vital signs as HR and RR from PPG signals [1], [2] (The words "signal" and "spectrum" indicate time domain and frequency domain information, respectively in this paper). Unfortunately, PPG signals have weak traces of respiratory information, limiting the accuracy of RR estimation in prior art.

Respiration has three different effects on PPG signals [3]–[5]: 1) The change in intrathoracic pressure leads to blood exchange between the pulmonary and systemic circulations, resulting in a change in the perfusion baseline during breathing cycles, which is referred to as respiratory-induced intensity variation (RIIV). 2) The change in ventricular filling leads to a corresponding change in cardiac output [6], which represents the change in peripheral pulse strength during breathing. This is called respiratory-induced amplitude variation (RIAV). RIAV suggests that the PPG signal is subject to amplitude modulation (AM). 3) As an autonomic response to respiration, the instantaneous HR varies to synchronize with the respiratory cycle – the HR increases during inspiration and decreases during expiration. This phenomenon is called respiratory-induced frequency variation (RIFV), indicating that the PPG signal is subject to frequency modulation (FM). Fig. 1 illustrates these respiratory-induced effects. Beyond RIIV, RIAV, and RIFV, other respiratory-induced variations in the PPG signal have been studied to estimate

Manuscript received January 15, 2020; revised June 23, 2020; accepted July 25, 2020. Date of publication August 3, 2020; date of current version April 5, 2021. *(Corresponding author: Quanzeng Wang.)*

Mingliang Chen and Min Wu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: mchen126@terpmail.umd.edu; minwu@umd.edu).

Qiang Zhu was with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA. He is now with the Facebook Inc. (e-mail: zhuqiang@umd.edu).

Quanzeng Wang is with the Center for Devices and Radiological Health, U.S. Food and Drug Administration (FDA), Silver Spring, MD 20993 USA (e-mail: Quanzeng.Wang@fda.hhs.gov).

Digital Object Identifier 10.1109/JBHI.2020.3013811

RR, such as the pulse amplitude variability and pulse width variability [7]–[9].

Considering the combined effect of RIIV, RIAV, and RIFV, time-domain approaches have been proposed to improve the accuracy of RR estimation [4], [10]. In these studies, the three respiratory-induced variations are first extracted from the PPG signal in the time domain and used individually to produce three seperate RR estimates. Averaging fusion [4] or auto-regression (AR) fusion [10] is then employed to combine the three RR estimates and produce the final RR estimate. In these studies, the fusion strategies demonstrated an improvement of more than 1 breath/min in the root mean square error (RMSE), compared with the estimates obtained from using RIIV, RIAV, and RIFV individually. Since these approaches need to examine the valley and peak points in PPG signals to extract the three types of respiratory-induced information, they are sensitive to additive noise. To obtain accurate features in the time domain, these prior approaches generally require a high data sampling rate. In addition, auto-regressive modelling [11] and Gaussian processes [12] have been used to explore the respiratory-induced variations in PPG signals.

PPG signals have also been investigated in the frequency domain [13]–[18]. In these works, only the RIIV in the PPG signal is considered with the assumption that the breath-related signal is superimposed onto the pulse signal. That is, the RR can be estimated by observing the lower frequency range of the spectrum (< 0.7 Hz). A variety of frequency analysis techniques, including periodogram [13], wavelet decomposition [14], empirical mode decomposition (EMD) [15], [17], empirical wavelet transforms (EWT) [16] and correntropy spectral density (CSD) [18] have been applied to obtain accurate RR estimates in the normal RR range. However, the superimposition model does not consider the other two important effects – RIAV and RIFV. Besides, the RIIV factor does not always appear detectable in PPG signals, especially after signal preprocessing operations such as filtering and detrending have been applied to it. In these studies, the frequency-domain methods have a low median RMSE of 0.9 breaths/min, but large variances in the estimation error, indicating that these algorithms are not robust and stable. The work in [19] detects RIAV and RIFV in PPG signals via the variable frequency complex demodulation method (VFCDM), which is a general frequency analysis framework used to estimate variable frequency. Compared with the frequency-domain methods using RIIV [13]–[18], the VFCDM has a smaller variance in estimation error but larger median error (about 2.5 breaths/min).

Traditionally, PPG signals have been collected by a pulse oximeter attached to the skin, which we refer to as the "contact PPG (cPPG) signal" to facilitate the discussions. The aforementioned studies have only been applied to cPPG signals. In the emerging mHealth care, a newly formed modality of PPG, "remote PPG (rPPG)" has garnered growing interests [20]–[23]. The rPPG signal is typically captured from a color video of the subject's face.

The principle of rPPG is that the blood volume changes under the skin influence the intensity and color of the reflected light from the skin. This pattern is consistent with heartbeat cycles. Although such subtle momentary changes in the reflected light from the facial skin are not detectable by the human eye, they can be captured by a color camera [20], [24]. Using a low-cost, accessible color camera to collect PPG signals, the rPPG technique is convenient and user-friendly, freeing users of contact sensors. One major challenge, though, in using rPPG is the low signal-to-noise ratio (SNR). Since a color camera is normally one-half to one meter away from a subject's skin, non-ideal illumination conditions and the subject's voluntary movements can influence the signal quality. Although effective denoising operations [25]–[28] have been employed to deal with realistic conditions and improve HR estimation accuracy from the rPPG signal, the research in RR estimation from such signal is still limited, and previous works [21], [29], [30] mainly estimate RR via heart rate variability (HRV) extraction [31] in the time domain.

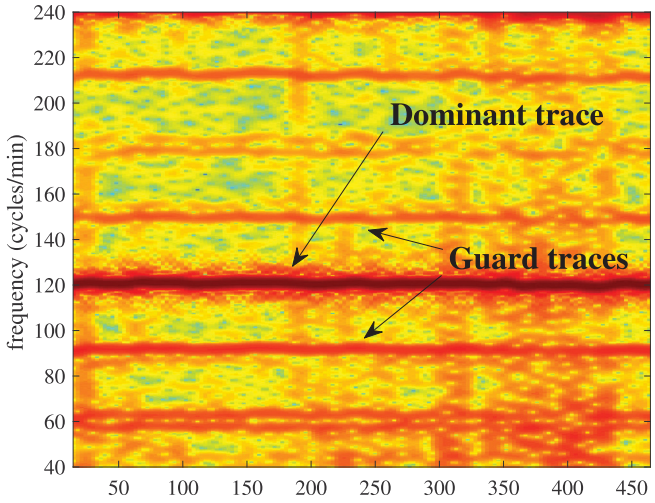

Fig. 2. The spectrogram of a subject's PPG signal in CapnoBase [4], showing one dominant frequency trace and two guard traces on both sides.

To improve the PPG signal analysis technology, we have developed a frequency-domain method based on a modulation model to extract RIAV and RIFV features from the PPG spectrum to estimate HR and RR. The motivation for this derives from the respiratory-induced effects on PPG signals and the observation of three noticeable signal traces in PPG spectrograms (Fig. 2). We present the spectrograms (i.e., time-frequency representations) in the paper to show the variation of the vital signs with time. We model PPG signals using amplitude and frequency modulation (AM-FM) and apply it to HR and RR estimation. The proposed AM-FM method utilizes relatively robust respiration-induced variation features; avoids the drawbacks of the peak/valley detection algorithms (i.e., detecting peak and valley points in PPG signals) under noisy or low-sampling rate scenarios; and improves the accuracy and robustness of RR estimation from PPG signals. The main contributions of this paper are as follows:

- We propose a modulation model, AM-FM, for PPG signals. This model can explain the dominant and guard traces in a PPG spectrogram as shown in Fig. 2.

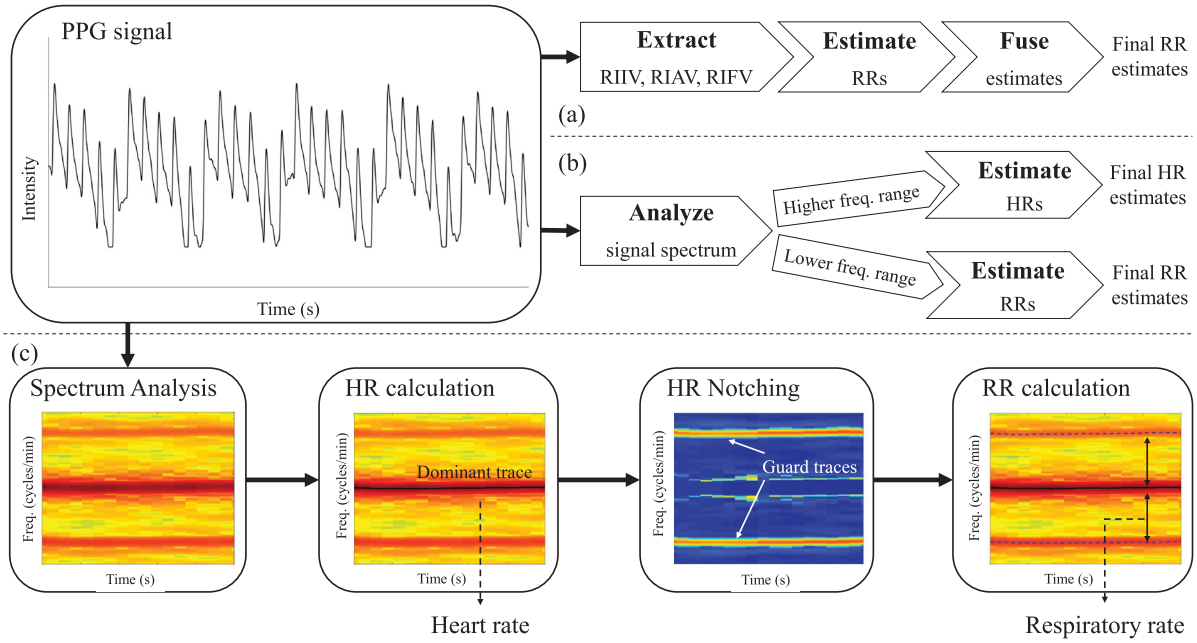

Fig. 3. Typical components of RR estimation in prior art via (a) time-domain and (b) frequency-domain approaches. The pipeline of the time-domain methods includes extracting respiratory-induced variation features in the time domain, estimating an individual RR from each variation feature, and fusing the estimated RRs to produce the final result. The pipeline of the frequency-domain methods consists of estimating the spectrum of the PPG signal and extracting HR and RR from different frequency ranges of interest. (c) The pipeline of the proposed AM-FM method. Spectrum analysis is applied to the PPG signal, where three main frequency traces can be easily observed. The dominant frequency component – HR – is extracted from the spectrum. Then, the dominant trace is adaptively notched and the two guard traces are extracted from the residual spectrum. The RR is derived from the frequency distance between the dominant and guard traces. In the spectrograms, the horizontal and the vertical axes denote time and frequency, respectively. We cropped the frequency range from 80 to 160 cycles/min for better visualization.

- We develop a robust frequency-domain algorithm based on the AM-FM model for HR and RR extraction from PPG spectra, taking advantage of the symmetry of the dominant and guard traces.
- We validate the proposed AM-FM method on both the cPPG and emerging and relatively noisier rPPG signals in terms of accuracy and robustness.

The remainder of this paper is organized as follows. Section II introduces the AM-FM model on PPG signals, and develops a robust frequency-domain algorithm based on the model to estimate HR and RR from PPG spectra. Section III shows the experimental results on the cPPG and rPPG signals and compares the proposed AM-FM method with the-state-of-art RR estimation methods. Section IV concludes the paper.

II. Proposed Algorithm

Fig. 2 presents the spectrogram of a person's PPG signal, where one dominant frequency trace and two guard traces on both sides can be easily observed. This phenomenon can be explained by our proposed AM-FM model as discussed in Section II-A. Based on this observation, we developed a frequency-domain algorithm to extract the three traces from the spectrogram to infer the HR and RR. The pipeline of the proposed method is illustrated in Fig. 3. Spectrum analysis is applied to the PPG signal after proper preprocessing (see Section II-B1 for details), where the dominant and guard traces can be easily observed. The dominant frequency component is extracted from the spectrum as the HR and then automatically notched to extract the two guard components. The RR is derived from the frequency distance between the dominant and the guard components.

A. AM-FM Model

As discussed in Section I, respiration influences the PPG signal in three aspects: RIIV, RIAV, and RIFV. RIIV is not a robust variation feature for RR estimation, since the baseline shift of the signal can be easily contaminated. In cPPG cases, high-pass filtering is usually performed in the pulse oximeter to remove the slow baseline shift in the cPPG signal. Such built-in filtering may weaken the RIIV effect. In rPPG cases, illumination variation from the external environment can influence the baseline shift. Therefore, we only consider RIAV and RIFV in modeling PPG signals. For simplicity and clarity of the derivation of our model, we assume that the HR and RR signals are purely sinusoidal functions with zero phase in a short time period.

A PPG signal $s(t)$ can be expressed as an AM-FM signal:

$$s(t) = \underbrace{(1 + k_a \sin(2\pi f_{rr} t))}_{\text{AM: } s_a(t)} \underbrace{\cos(2\pi f_{hr} t + k_f \sin(2\pi f_{rr} t))}_{\text{FM: } s_f(t)}, \tag{1}$$

where $s(t)$ is assumed to have unit amplitude; f_{hr} and f_{rr} denote HR and RR, respectively; and k_a and k_f characterize the variation strength in RIAV and RIFV, respectively. After

applying the angle sum identity to the FM part in (1), we obtain

$$s_f(t) = \cos(2\pi f_{hr}t)\cos(k_f \sin(2\pi f_{rr}t)) - \sin(2\pi f_{hr}t)\sin(k_f \sin(2\pi f_{rr}t)). \quad (2)$$

In practice, k_f is a small positive value in modeling a person's PPG signal. It was estimated to be between 0.11 to 0.32 from the CapnoBase dataset [4]. Equation (2) can be approximated by

$$s_f(t) \approx \cos(2\pi f_{hr}t) - \sin(2\pi f_{hr}t)k_f \sin(2\pi f_{rr}t). \quad (3)$$

By applying the product-to-sum identity, (3) becomes

$$s_f(t) \approx \cos(2\pi f_{hr}t) + \frac{k_f}{2}\cos(2\pi(f_{hr}+f_{rr})t) - \frac{k_f}{2}\cos(2\pi(f_{hr}-f_{rr})t). \quad (4)$$

Note further that k_a is a small positive value in practice, which is estimated to be between 0.05 to 0.23 from the CapnoBase dataset [4]. Hence, we ignore the term $k_a k_f$ after expanding the two terms $s_a(t)$ and $s_f(t)$. Thus, (1) can be simplified to

$$s(t) \approx \cos(2\pi f_{hr}t) + \frac{k_f}{2}\cos(2\pi(f_{hr}+f_{rr})t) + \frac{k_a}{2}\sin(2\pi(f_{hr}+f_{rr})t) - \frac{k_f}{2}\cos(2\pi(f_{hr}-f_{rr})t) - \frac{k_a}{2}\sin(2\pi(f_{hr}-f_{rr})t). \quad (5)$$

Equation (5) indicates that there are three main frequency components in a person's PPG signal: $f_{hr} - f_{rr}$, f_{hr}, and $f_{hr} + f_{rr}$, as shown in Fig. 2. The AM-FM model of the PPG signal reveals that the PPG spectrum contains noticeable HR and RR information that can be extracted jointly from the spectrum.

Estimation of k_a and k_f: In the signal modulation, the modulation index k_a of an AM signal is computed as (6). For the FM signal $s_f(t)$, the instantaneous frequency (equivalent to RIFV in the respiration contexts) is computed by $f = \frac{1}{2\pi}\cdot\frac{\partial\phi(t)}{\partial t} = f_{hr} + k_f f_{rr}\cos(2\pi f_{rr}t)$, where $\phi(t)$ denotes the instantaneous angle of the sinusoidal signal, which gives rise to (7) for k_f:

$$k_a = \frac{u_{a,\max} - u_{a,\min}}{u_{a,\max} + u_{a,\min}}, \quad (6)$$

$$k_f = \frac{u_{f,\max} - u_{f,\min}}{2f_{rr}}, \quad (7)$$

where $u_{a,\max}$ and $u_{a,\min}$, $u_{f,\max}$ and $u_{f,\min}$ denote the maximum and minimum of the RIAV and RIFV signal, respectively. We apply (6) and (7) to estimate the ranges of k_a and k_f in PPG signals from the CapnoBase dataset.

In the following subsections, we focus on estimating the dominant component and its two guard components from the spectrum of a collected PPG signal. In long-term monitoring, a moving window is usually employed to track the variation of the vital signs. We present the core elements of the proposed algorithm in terms of one window frame. The method can be easily extended to continuously track the frequency components in long-term monitoring, for example, by applying the AMTC algorithm [32], [33] which applies an efficient tracking algorithm of frequency traces on spectrograms using dynamic programming.

B. Heart Rate Estimation

1) Preprocessing: We preprocess the raw PPG signal using bandpass filtering and normalization. The bandwidth of the filter is 0.25 Hz to 5 Hz, which contains the normal HR frequency range. Normalization is applied using a moving window with a length of one second. We normalize one sample x in the signal by taking out the mean μ_x and scaling by the standard deviation σ_x in the x-centered one-second moving window. The normalized sample $\tilde{x}$ is

$$\tilde{x} = \frac{x - \mu_x}{\sigma_x}. \quad (8)$$

After applying filtering and normalization, the PPG signal has approximately invariant signal energy per unit time.

2) Dominant Component: For a PPG signal $s(t)$ sampled with sampling interval Δ_t, the power spectral density (PSD) is estimated through the periodogram:

$$S(f) = \frac{\Delta_t}{N}\left|\sum_{n=0}^{N-1} x_n e^{-j2\pi f \Delta_t n}\right|^2, \quad (9)$$

where x_n denotes the n-th sample of the signal; and N is the total number of samples. We estimate the heart rate f_{hr} by finding the location of the highest spectral energy in the spectrum.

C. Respiratory Rate Estimation

Two guard components, as shown in Fig. 2, appear symmetrically on both sides of the dominant component. They are estimated in the following steps.

1) Symmetrical Averaging: To utilize the symmetric property of the two guard components, we flip the spectrum around the dominant component f_{hr} to obtain the symmetric spectrum $\tilde{S}(f)$.

$$\tilde{S}(f) = \sqrt{S(f)S(2f_{hr} - f)}. \quad (10)$$

By imposing the symmetric property around the dominant peak, this operation emphasizes the symmetric structures that appear on both sides of the dominant peak and attenuates the isolated false peaks caused by noise and distortion. Fig. 4 illustrates how symmetrical averaging successfully attenuates the false alarm peaks and amplifies the symmetric guard peaks.

2) Energy Notching: To estimate the guard peaks, which are lower in intensity than the dominant peak from the symmetric spectrum, we first reduce the interference of the dominant peak on other smaller peaks by notching its energy. The energy notching step removes the dominant peak's influence on its nearby frequency range, and helps mitigate the false alarm peaks lying closely on the slope of the dominant peak. As illustrated in Fig. 5, the symmetric spectrum is decomposed into the peak spectrum $S_p(f)$ which contains the dominant component, and the residual spectrum $S_r(f)$ which contains the guard components and false alarm peaks.

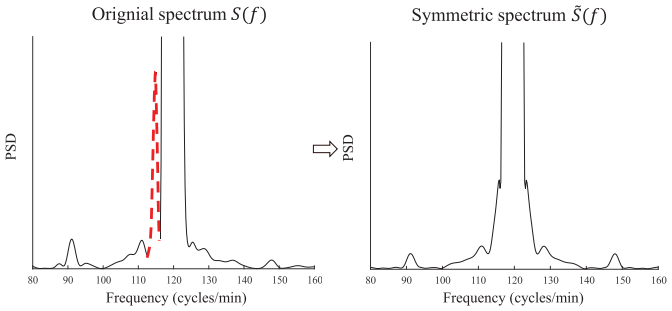

Fig. 4. Symmetrical averaging. We present the lower part of the spectra for better visualization of the guard components. The false alarm peaks (red dash) are attenuated by imposing the symmetric property along the dominant peak.

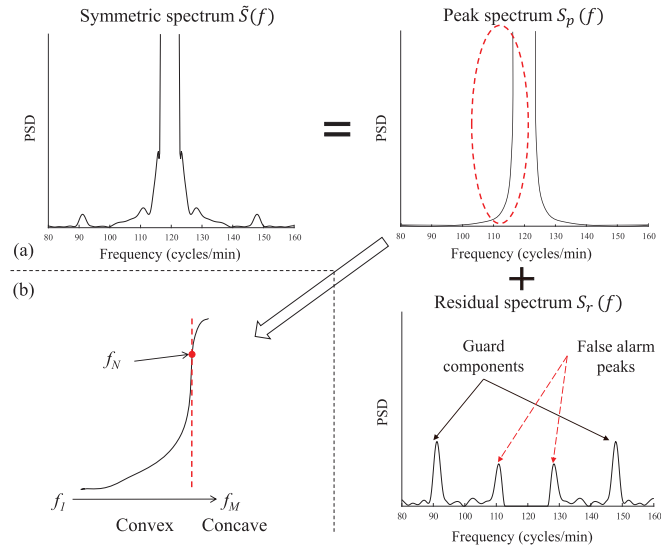

Fig. 5. Energy Notching. (a) The symmetric spectrum is the superposition of the peak and the residual spectra. We present the lower part of the symmetric and the peak spectra for better visualization of the guard components. (b) Left side of the bell-shaped curve of the peak spectrum, where f_N is the inflection point of the curve slope.

We assume the peak spectrum is a bell-like shape. Due to the symmetric property, we only focus on the left side of the peak spectrum, which has the following characteristics:

- It contains as much energy as possible, but is upper bounded by the spectrum $\tilde{S}(f)$.
- It is a non-decreasing function.
- Its slope first increases until the inflection point f_N and then decreases to zero.

Let $\{f_i\}_{i=1,2,\ldots,N,\ldots,M}$ denote the equally sampled frequency values, and f_N denotes the inflection point on the left side of the peak spectrum shown in Fig. 5(b). The left side of the peak spectrum $S_p(f_i)$ is convex before f_N and concave after f_N. To find f_N, we check the concavity of each point f_i, starting from the peak point to the left on the symmetric spectrum and stopping once reaching the first point that no longer satisfies (11):

$$2\tilde{S}(f_i) \geq \tilde{S}(f_{i-1}) + \tilde{S}(f_{i+1}),\ i = M-1, M-2, \ldots \quad (11)$$

where $\tilde{S}(f_i)$ denotes the symmetric spectrum. For the concave part where $f_i > f_N$, we have

$$S_p(f_i) = \tilde{S}(f_i),\ i = N+1, N+2, \ldots, M, \quad (12)$$

For the convex part where $f_i \leq f_N$, we can solve the peak spectrum via linear programming expressed in (13).

$$\max_{S_p(\cdot)} \sum_{i=1}^{N} S_p(f_i)$$

$$\begin{aligned} \text{s.t.}\quad & S_p(f_{i-1}) \leq S_p(f_i), && i = 2, 3, \ldots, N, \\ & 2S_p(f_i) \leq S_p(f_{i-1}) + S_p(f_{i+1}), && i = 2, 3, \ldots, N-1, \\ & 0 \leq S_p(f_i) \leq \tilde{S}(f_i), && i = 1, 2, \ldots, N. \end{aligned} \quad (13)$$

The first inequality enforces the non-decreasing property. The second inequality enforces the property of convexity. The third inequality sets the upper and lower bounds.

We can obtain the residual spectrum $S_r(f)$ by notching the peak spectrum $S_p(f)$ from the symmetric spectrum $\tilde{S}(f)$.

$$S_r(f) = \tilde{S}(f) - S_p(f). \quad (14)$$

3) Guard Components: After obtaining the residual spectrum, the locations of highest spectral energy on both sides of the dominant peak in the residual spectrum are selected as the guard components. Finally, the RR is derived from the length of interval between the dominant and guard components.

III. Experimental Results and Discussions

We conducted experiments to demonstrate the effectiveness of the proposed AM-FM method on two PPG signal datasets – a cPPG dataset collected with a contact oximeter and an rPPG dataset captured with a color camera. It is worth noting that the rPPG signal is much noisier than the cPPG signal, due to the distance between human skin and the sensor and the subject's voluntary movements. Since many studies have been conducted to estimate HR from the PPG signal captured with either contact [34], [35] or contactless sensors [20], [21], [25], [28], the evaluation of HR estimation is not the focus of our paper and we only evaluated RR estimation.

We used a moving window to extract the RR from the PPG signal throughout its duration. The performance of RR estimation algorithms was assessed using the RMSE metric:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(x_i^{true} - x_i^{est})^2}, \quad (15)$$

where n is the total number of RR estimates calculated in one PPG signal and x_i^{true} and x_i^{est} are the ground truth and estimated RR in the i-th window, respectively. We also computed the Pearson correlation coefficient (PCC) to measure the linear correlation between the ground truth and estimated RR. The PCC ranges from -1 to 1. The coefficient values -1, 0, and 1 indicate perfect negative linear correlation, no linear correlation, and perfect positive linear correlation between two variables, respectively. In our evaluation, a PCC value closer to 1 means better performance of an algorithm in RR estimations.

We compared our proposed method with several prior representative works. In the comparisons, we used the results from the

SmartFusion [4] and CSD [18] methods provided by the authors, and implemented the ARFusion [10], EMD [15], EWT [16], VFCDM [19], and HRV-based [29] methods by ourselves.

The SmartFusion and ARFusion methods first extract the three respiratory-induced variation signals in the time domain. SmartFusion estimates three RR values from the three signals, and then obtains the final result by averaging the three RR estimates. ARFusion models the three signals via AR modeling, and derives the final estimate by fusing the AR spectra of the three respiratory-induced variations with several model orders. In the rPPG case, the HRV-based method only extracts the RIFV feature for RR estimation since RIIV and RIAV are usually attenuated when face videos are captured from a distance. In the frequency-domain methods, frequency analysis techniques – EMD, EWT, and CSD – are employed, respectively, to analyze the frequency components of the PPG signal, and extract the RR from the expected RR frequency range. VFCDM analyzes the PPG signal with spectral estimation and identifies AM and FM dynamics in the signal, which contain RR information.

A. cPPG Dataset Collected with a Contact Oximeter

We used the CapnoBase dataset [4], which contains eight-minute cPPG signals of 29 pediatric and 13 adult subjects, during elective surgery and routine anesthesia. The dataset does not disclose the subjects' health status. The PPG signals were collected with an oximeter and the ground truth RR was obtained from capnography with a sampling rate of 300 Hz.

In the experiments, the length of the moving window was set to 32 seconds in all the methods, except 120 seconds in the CSD method, as [18] suggests. The window proceeded for 3 seconds between adjacent estimates. In our proposed method, the collected signal was downsampled from 300 Hz to 30 Hz to reduce the computational complexity. Downsampling also reveals that our method can be used for PPG signals collected with a low sampling rate. We compared the proposed method with the time-domain methods, SmartFusion [4] and ARFusion [10], and frequency-domain methods, EMD [15], EWT [16], CSD [18], and VFCDM [19].

For all methods, we zeroed out the measurement artifacts (provided by the CapnoBase) in the input cPPG signal and the RR was only estimated on the non-artifact duration of the signal. Note that SmartFusion may discard additional RR estimates due to its signal quality assessment strategy. For a fair comparison, we recovered the discarded estimates from SmartFusion by their nearest trustworthy neighbors.

Fig. 6 shows boxplots of the different methods' RMSE values on CapnoBase dataset. Table I summarizes the quantitative evaluation of their performance, where the best indices are shown in bold. As shown in Fig. 6, SmartFusion and ARFusion have similar performances in RMSE. EMD has the largest median RMSE and variance. EWT and CSD have smaller median RMSEs, but larger variances in RMSE than SmartFusion and ARFusion, indicating that these algorithms are less robust. VFCDM has a similar RMSE variance but a larger mean and median RMSE,

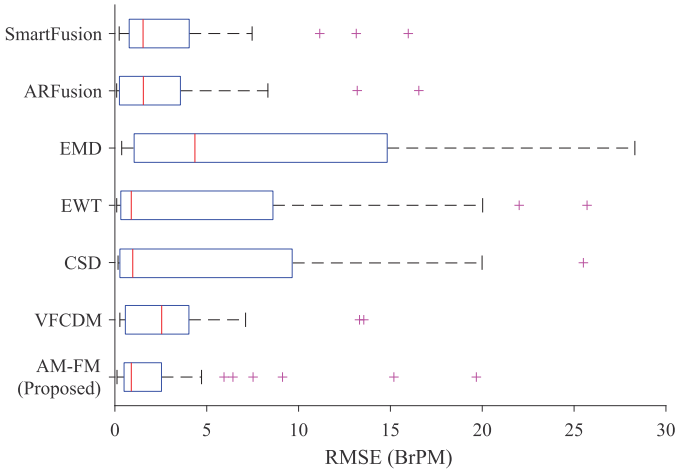

Fig. 6. Comparison of RMSE for RR estimation methods on the CapnoBase dataset. The boxplot shows distributions of RMSEs, with the first (Q1), second (median), and third quartile (Q3) values displayed as left, middle, and right vertical lines of the boxes. Whiskers represent the most extreme values within 1.5 times of the interquartile range (i.e., the range between Q1 and Q3). The outliers beyond the whiskers are displayed as magenta crosses.

TABLE I
OVERALL PERFORMANCE OF RR ESTIMATION METHODS ON CPPG DATASET (RMSE UNIT: BRPM)

Method	RMSE statistics		p-value	PCC
	Median (Q1, Q3)	Mean		
SmartFusion [4]	1.53 (0.78, 4.04)	2.92	ref.	0.64
ARFusion [10]	1.61 (0.31, 3.63)	2.67	0.48	0.65
EMD [15]	4.36 (1.11, 14.6)	5.68	5.7e-4	0.45
EWT [16]	0.90 (0.32, 7.59)	3.93	0.85	0.53
CSD [18]	0.95 (**0.27**, 6.20)	3.86	n/a	n/a
VFCDM [19]	2.55 (0.57, 3.88)	2.99	0.19	0.60
AM-FM (Proposed)	**0.89** (0.50, **2.54**)	**2.53**	3.2e-3	**0.66**

compared to SmartFusion and ARFusion. The proposed AM-FM method has the smallest RMSE median and variance values, indicating that it is the most accurate and robust among these algorithms. Table I also shows that the proposed AM-FM method has lower RMSE means – 0.39, 0.14, 3.15, 1.40, 1.33, and 0.46 breaths per minute (BrPM), compared with SmartFusion, ARFusion, EMD, EWT, CSD, and VFCDM, respectively. We can see from the RMSE statistics that the proposed AM-FM method has a noticeable reduction in both the median and mean RMSE values and provides the most accurate RR estimates, compared with the other methods. The PCC index suggests that the proposed method ($\rho = 0.66$) has the best positive correlation with the ground truth RR.

The Wilcoxon sign rank test [36] was performed to show the differences between these methods, with the SmartFusion method as the reference. The test is a nonparametric test that evaluates whether the group mean ranks differ from two non-normally distributed data groups when the observations are paired. The test results indicate that the AM-FM method ($p = 3.2 \times 10^{-3}$) is significantly different from the SmartFusion method.

B. rPPG Dataset Captured by a Color Camera

In recent years, rPPG has been attracting a lot of research attention [21], [25], [28], [37], due to its promising application in mHealth care. In this section, we also explored the robustness of the proposed AM-FM method on this new modality of PPG signal.

We evaluated the HRV-based, EMD, EWT, and proposed methods on a self-collected rPPG dataset. The study was approved by the Institutional Review Board of the U.S. Food and Drug Administration (IRB approval reference number: 17-043R-CR 0001). The dataset contains 60 one-minute face video clips of 12 subjects with ages ranging from 18 to 50 years old. The skin tones of the participants are generally categorized into three skin types based on the Fitzpatrick scale [38]: western European (skin type I and II), eastern Asian (skin type III and IV), and African/southern Asian (skin type V and VI). During the recordings, the subjects were asked to sit in front of a desk, where a regular 1080p webcam (HD Pro Webcam C920, Logitech International S.A., Lausanne, Switzerland) was used to capture videos of the frontal positions of subjects' faces at a frame rate of 30 Hz. The ground truth RR was simultaneously measured via a piezo respiratory belt transducer (Model 116-0018, Great Lakes NeuroTechnologies (GLNT), Cleveland, Ohio). The transducer was connected to a BioRadio^{TM} acquisition system (Model 700-0016, GLNT) for signal collection with a sampling rate of 500 Hz. The subjects' faces were illuminated the dedicated fluorescent lights from the ceiling, with a total illuminance of approximately 200 lux. The Subjects' voluntary rigid head motions and non-rigid face motions were allowed during the recording, including talking and facial expressions. Beyond the rPPG videos and the corresponding ground truth RR, we did not collect any information related to subjects' health status.

For fair comparison of the RR estimation methods, we processed the rPPG signal in the same way for each method. A robust extraction algorithm was deployed to extract the pulse signal from the videos. We first applied an SSD-ResNet based face detector [39], [40] to localize the face region in each frame. Then, the region of interest (ROI), defined by the entire face region, was obtained and refined with the face landmark detector [41] and the ROI selection principles [42]. A spatially averaged RGB color signal was then extracted over the detected ROI in each frame. Finally, the pulse signal was computed with the POS algorithm [37] by mapping the 3-channel RGB signal to a 1-channel rPPG signal. After obtaining the rPPG signal, the proposed AM-FM method was applied to estimate the RR and compared with the time-domain HRV-based method [29] and frequency-domain methods, EMD [15] and EWT [16]. The HRV-based method only extracted the RIFV signal from the rPPG signal, since RIIV and RIAV are usually attenuated during signal preprocessing. EMD and EWT were executed in the same way as in the cPPG case. In RR estimation, the window length was set to 32 seconds and the moving window proceeded one second between adjacent estimates.

Fig. 7 presents boxplots of the RMSE values associated with the four methods for RR estimation, showing that the proposed AM-FM method has the smallest estimation error among the methods. Table II summarizes the performance statistics of the methods on the rPPG video dataset, where the best values are in bold. As expected, the RR estimation error of the rPPG signal, in general, is larger than that of the cPPG signal (Table I), since the rPPG signal can be easily contaminated by noise from a variety of sources, such as voluntary motion and illumination variation. From Table II, we can observe that EMD and EWT have the largest error, indicating that RIIV is not a robust feature in the rPPG case. The HRV-based method performs slightly better than EMD and EWT. Overall, the AM-FM method has the best performance in terms of median RMSE (2.46 BrPM) and PCC ($\rho = 0.36$) with the ground truth RR. In the Wilcoxon sign rank test, the p-value of 6.4×10^{-4} shows that the performance of the AM-FM method is significantly different from the HRV-based method. In the rPPG case, we can see that the proposed AM-FM method is the most resistant to noise compared with the other methods.

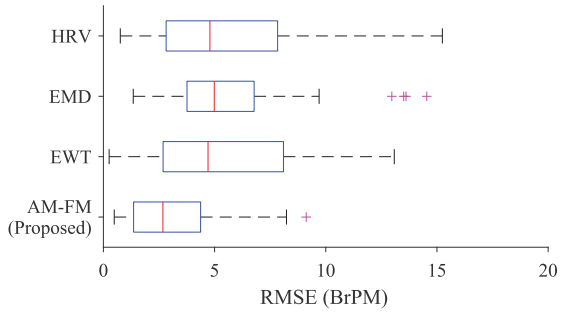

Fig. 7. Comparison of RMSE for RR estimation methods on the self-collected rPPG dataset. The boxplot settings are identical to Fig. 6.

TABLE II
OVERALL PERFORMANCE OF RR ESTIMATION METHODS ON RPPG DATASET (RMSE UNIT: BRPM)

Method	RMSE statistics		p-value	PCC
	Median (Q1, Q3)	Mean		
HRV [29]	4.10 (2.73, 7.64)	5.24	ref.	0.32
EMD [15]	4.99 (3.83, 6.62)	5.71	0.77	0.25
EWT [16]	4.71 (2.92, 8.03)	5.33	0.85	0.29
AM-FM (Proposed)	**2.46** (**1.29**, **4.08**)	**3.08**	6.4e-4	**0.36**

C. Discussions

As described in Section I, respiration can induce three effects on the PPG signal: RIIV, a baseline shift in the signal during each breathing cycles; RIAV, a change in the signal amplitude; RIFV, a variation in HR during breathing. The first effect indicates the breath-related signal is superimposed onto the PPG signal. The second and third effects reveal the AM and FM properties in the PPG signal, respectively. The previous RR estimation methods from PPG signals can be mainly categorized into two groups: time-domain and frequency-domain methods, whose typical pipelines are presented in Fig. 3(a) and (b), respectively.

Comparison with prior methodologies: To improve algorithm robustness in the cPPG signal, the time-domain methods in prior art [4], [10] utilize all three respiratory-induced

variation features in the signal, from which three individual RR estimates are extracted, but each time-domain approach fuses these estimates in a slightly different way to obtain the RR. In SmartFusion [4], averaging is employed to provide a final estimate in each window, and the algorithm will discard the RR estimate if a large discrepancy exists among the three individual estimates. However, discarding data does not truly improve the RR estimation accuracy, but rather adds an unknown final estimate. On average, 35% of the windows in each case were eliminated due to disagreement in the three estimates. In ARFusion [10], the final estimate is derived by fusing the AR spectra of the three respiratory-induced variations with several model orders. Although this fusing method is more advanced than average fusing, it still cannot avoid the peak/valley detection problem in a noisy spectrum. The p-value between SmartFusion and ARFusion ($p = 0.48$) also shows that their performances have no significant difference in terms of RMSE.

The frequency-domain methods [13]–[16], [18], unlike the time-domain ones, only consider the RIIV in the PPG signal. They model the PPG signal as the superimposition of the pulse signal and the respiration signal. These methods typically analyze the PPG spectrum and extract HR and RR from different frequency ranges of interest. Different frequency analysis techniques such as periodogram [13], wavelet [14], EMD [15], EWT [16] and CSD [18] are introduced in these algorithms to estimate the spectrum of the PPG signal. Given the assumption that the RR component is in a lower frequency range ($0.1 - 0.7$ Hz) and the HR component is in a higher frequency range ($0.7 - 4$ Hz), the HR and RR can be extracted from these ranges. Although advanced frequency analysis methods, such as EMD, EWT, and CSD, can help decompose the signal into different frequency sub-bands and improve the spectral estimation accuracy, these methods only consider RIIV in the PPG signal, which can be easily influenced by slow varying noise or removed by necessary preprocessing steps, such as detrending and filtering. The relatively large variances of RMSE in the EMD, EWT, and CSD methods (Fig. 6) demonstrate the drawback of these methods based on the superimposition assumption. Moreover, the CSD method requires a sliding window of at least 60 seconds to obtain reliable RR estimates [18], resulting in a time delay in estimation.

Compared with RIIV, we have focused on RIAV and RIFV that are more resistant to additive noise and preprocessing, analogous to the noise resistance of AM and FM in radio communication. Taking the effects of RIAV and RIFV into account, PPG signals can be modeled with amplitude and frequency modulation, rather than only considering RIIV, as the prior art of the frequency-domain methods do. VFCDM can be applied to extract AM and FM dynamics from a signal [19], but it is a general frequency analysis framework used to estimate variable frequency. The results show that VFCDM has a larger performance error than our method. Assuming the single-frequency modulation in AM and FM, we adapt and simplify the AM-FM model to meet the needs of RR estimation from PPG signals. Our resulting AM-FM model explains the three noticeable components – a dominant component and its two guard components – in the PPG spectra. By taking advantage of the symmetric property of the three components, the proposed algorithm can successfully extract them from the PPG spectrum and obtain the HR and RR, avoiding the potential problem in peak/valley detection. The performance in Table I demonstrates that the AM-FM method can estimate RR from the cPPG signal more accurately than other methods.

The AM-FM method is computationally efficient. The average processing time for 100 RR estimations is about 2 seconds with a single processor thread on a 2.4 GHz PC with 8 GB memory, performing real-time execution. The step of energy notching solves a linear programming problem for each RR estimate, which takes a substantial percentage of the overall computational load.

Performance comparison in rPPG scenarios: As presented in Section III-B, we have also tested the performance of the AM-FM method on rPPG signals and compare its performance with the HRV, EMD, and EWT methods. rPPG is an emerging modality of PPG that has two main challenges in processing rPPG data. First, an rPPG signal has much lower SNR than a cPPG signal. The HRV-based methods [21], [29], [30] extracted the RIFV signal from the rPPG signal to estimate the RR, since this kind of variation is a relatively robust feature in noisy scenarios. Due to a large amount of noise in the rPPG signal, it is difficult for the time-domain methods to provide reliable detection of peaks and valleys, leading to large estimation errors. Second, the baseline shift of the rPPG signal coming from illumination variation from the external environment can weaken the RIIV effect on the signal. We tested on the rPPG signal EMD [15] and EWT [16] that are based on the RIIV effect, and Table II shows these two methods suffer from large error and variance. The above discussion suggests that the previous methods may not be applicable to the rPPG signal. In contrast, the proposed AM-FM method considers relatively robust RIAV and RIFV features, and extracts them robustly from the frequency domain, thus achieving better performance than the time-domain methods based on peak/valley detection algorithms and the frequency-domain methods that only consider RIIV features. Overall, the performance of the AM-FM method on the rPPG signal (Table II) indicates that it improves the accuracy of RR estimation and is robust under noisy conditions.

IV. Conclusion

In this paper, we introduce the AM-FM to model PPG signals, which exploits two kinds of relatively robust respiration-induced features of PPG signals: RIAV and RIFV. The AM-FM model is consistent with the observation of two guard components lying symmetrically around the dominant component in the PPG spectrum. Based on the model, we have developed a robust and efficient frequency-domain algorithm to directly extract HR and RR from PPG spectra. The method was evaluated both on a contact-based PPG dataset collected by a pulse oximeter and on a remote PPG dataset consisting of a set of face videos collected by a color camera. The extensive experimental results demonstrate that the proposed AM-FM method is effective and robust even in relatively noisy scenarios of remote PPG data.

ACKNOWLEDGMENT

This research was partially supported by the U.S. Food and Drug Administrations Critical Path funding (CPOCD64). The authors would like to acknowledge Zachary Lazri for editorial contributions to this manuscript.

REFERENCES

[1] E. Olsson, H. Ugnell, P. Oberg, and G. Sedin, "Photoplethysmography for simultaneous recording of heart and respiratory rates in newborn infants." *Acta Paediatrica* , vol. 89, no. 7, pp. 853–861, 2000.

[2] K. H. Shelley, "Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate," *Anesth. & Analg.*, vol. 105, no. 6, pp. S31–S36, 2007.

[3] D. Meredith, D. Clifton, P. Charlton, J. Brooks, C. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: A review of relevant physiology," *J. Med. Eng. Technol.*, vol. 36, no. 1, pp. 1–7, 2012.

[4] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, 2013.

[5] D. A. Birrenkott, M. A. Pimentel, P. J. Watkinson, and D. A. Clifton, "Robust estimation of respiratory rate via ECG-and PPG-derived respiratory quality indices," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 676–679.

[6] A. J. Buda, M. R. Pinsky, N. B. Ingels Jr, G. T. Daughters, E. B. Stinson, and E. L. Alderman, "Effect of intrathoracic pressure on left ventricular performance," *N. Engl. J. Med.*, vol. 301, no. 9, pp. 453–459, 1979.

[7] P. Leonard, N. R. Grubb, P. S. Addison, D. Clifton, and J. N. Watson, "An algorithm for the detection of individual breaths from the pulse oximeter waveform," *J. Clin. Monit. Comput.*, vol. 18, no. 5-6, pp. 309–312, 2004.

[8] J. Lázaro, E. Gil, R. Bailón, A. Mincholé, and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Med. Biol. Eng. Comput.*, vol. 51, no. 1-2, pp. 233–242, 2013.

[9] R. A. Cernat, C. Ungureanu, R. Aarts, and J. Arends, "Real-time extraction of the respiratory rate from photoplethysmographic signals using wearable devices," in *Proc. Eur. Conf. Ambient Intell.*, 2014, pp. 1–17.

[10] M. A. Pimentel *et al.*, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, 2016.

[11] S. G. Fleming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," *Int. J. Biol. Med. Sci.*, vol. 2, no. 4, pp. 232–236, 2007.

[12] M. A. Pimentel, P. H. Charlton, and D. A. Clifton, "Probabilistic estimation of respiratory rate from wearable sensors," in *Proc. Wearable Electron. Sensors*. Springer, 2015, pp. 241–262.

[13] K. H. Shelley, A. A. Awad, R. G. Stout, and D. G. Silverman, "The use of joint time frequency analysis to quantify the effect of ventilation on the pulse oximeter waveform," *J. Clin. Monit. Comput.*, vol. 20, no. 2, pp. 81–87, 2006.

[14] P. Leonard, N. R. Grubb, P. S. Addison, D. Clifton, and J. N. Watson, "An algorithm for the detection of individual breaths from the pulse oximeter waveform," *J. Clin. Monit. Comput.*, vol. 18, no. 5-6, pp. 309–312, 2004.

[15] A. Garde, W. Karlen, P. Dehkordi, J. M. Ansermino, and G. A. Dumont, "Empirical mode decomposition for respiratory and heart rate estimation from the photoplethysmogram," in *Proc. Comput. Cardiol.*, 2013, pp. 799–802.

[16] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, 2013.

[17] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "Ensemble empirical mode decomposition with principal component analysis: A novel approach for extracting respiratory rate and heart rate from photoplethysmographic signal," *IEEE J. Biomed. Health Inf.*, vol. 22, no. 3, pp. 766–774, 2017.

[18] A. Garde, W. Karlen, J. M. Ansermino, and G. A. Dumont, "Estimating respiratory and heart rates from the correntropy spectral density of the photoplethysmogram," *PLoS One*, vol. 9, no. 1, pp. e86427, 2014.

[19] K. H. Chon, S. Dash, and K. Ju, "Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2054–2063, Aug. 2009.

[20] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[21] 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, 2011, pp. 7–11.

[22] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE J. Biomed. Health Inf.*, vol. 18, no. 4, pp. 1153–1160, 2013.

[23] N. Sugita, M. Akay, Y. Akay, and M. Yoshizawa, "Noise reduction technique for single-color video plethysmography using singular spectrum analysis," *IEEE J. Biomed. Health Inf.*, 2019, pp. 1788–1795.

[24] J. Zheng, S. Hu, V. Chouliaras, and R. Summers, "Feasibility of imaging photoplethysmography," in *Proc. Int. Conf. Biomed. Eng. Inf.*, vol. 2, 2008, pp. 72–75.

[25] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4264–4271.

[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, 2016.

[27] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2396–2404.

[28] Q. Zhu, C.-W. Wong, C.-H. Fu, and M. Wu, "Fitness heart rate measurement using face videos," in *Proc. IEEE Intl. Conf. Image Process.*, 2017, pp. 2000–2004.

[29] X. Li *et al.*, "The OBF database: A large face video database for remote physiological signal measurement and atrial fibrillation detection," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 242–249.

[30] M. Chen, Q. Zhu, H. Zhang, M. Wu, and Q. Wang, "Respiratory rate estimation from face videos," in *Proc. IEEE Int. Conf. Biomed. Health Inf.*, 2019, pp. 1–4.

[31] R. Favilla, V. C. Zuccala, and G. Coppini, "Heart rate and heart rate variability from single-channel video and ICA integration of multiple signals," *IEEE J. Biomed. Health Inf.*, vol. 23, no. 6, pp. 2398–2408, Nov. 2019.

[32] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving based on dynamic programming," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, 2018, pp. 1716–1720.

[33] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving for robust frequency tracking in forensic applications," 2020, *arXiv:2005.06686*.

[34] H. Fukushima, H. Kawanaka, M. S. Bhuiyan, and K. Oguri, "Estimating heart rate using wrist-type photoplethysmography and acceleration sensor while running," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 2901–2904.

[35] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1902–1910, 2015.

[36] R. Woolson, "Wilcoxon signed-rank test," *Wiley Encyclopedia Clin. Trials*, pp. 1–3, 2007.

[37] 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.

[38] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Arch. Dermatol.*, vol. 124, no. 6, pp. 869–871, Jun. 1988.

[39] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

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

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