Video-Based Heart Rate Measurement: Recent Advances and Future Prospects

Xun Chen, *Member, IEEE*, Juan Cheng, *Member, IEEE*, Rencheng Song, Yu Liu, *Member, IEEE*, Rabab Ward, *Fellow, IEEE*, and Z. Jane Wang, *Fellow, IEEE*

***Abstract*—Heart rate (HR) estimation and monitoring is of great importance to determine a person's physiological and mental status. Recently, it has been demonstrated that HR can be remotely retrieved from facial video-based photoplethysmographic signals captured using professional or consumer-level cameras. Many efforts have been made to improve the detection accuracy of this noncontact technique. This paper presents a timely, systematic survey on such video-based remote HR measurement approaches, with a focus on recent advancements that overcome dominating technical challenges arising from illumination variations and motion artifacts. Representative methods up to date are comparatively summarized with respect to their principles, pros, and cons under different conditions. Future prospects of this promising technique are discussed and potential research directions are described. We believe that such a remote HR measurement technique, taking advantages of unobtrusiveness while providing comfort and convenience, will be beneficial for many healthcare applications.**

***Index Terms*—Facial video, heart rate (HR), noncontact, region of interest (ROI), remote photoplethysmography (rPPG).**

I. Introduction

Monitoring physiological parameters, such as heart rate (HR), respiratory rate (RR), HR variability (HRV), blood pressure, and oxygen saturation is of great importance to access individuals' health status [1]–[6]. Since the heart is one of the most important organs of the body, the estimation and monitoring of HR are essential for the surveillance of cardiovascular catastrophes and the treatment therapies of chronic diseases [1], [7]. Various methods have been developed to estimate HR using contact or noncontact sensors, and a relevant review is in [8]. The aim of all the noncontact methods is to provide a more comfortable and unobtrusive way to monitor HR and avoid discomfort or skin allergy caused by the conventional contact methods [9]–[11]. Therefore, the monitoring of cardiorespiratory activity by means of noncontact sensing methods has recently spurred a remarkable number of studies that have used different techniques, such as laser-based technique [12], radar-based technique [13], capacitively coupled sensors-based technique [14], and imaging photoplethysmography (iPPG) [11], [15]–[22] technique. IPPG is also referred to as remote PPG (rPPG), due to the fact that it can measure pulse-induced subtle color variations from a distance of up to several meters using cameras with ambient illuminations [23]–[25]. The rPPG measurement is based on the similar principle to that of the traditional PPG, which the pulsatile blood propagating in the cardiovascular system changes the blood volume in the microvascular tissue bed beneath the skin within each heartbeat and thereby a fluctuation is periodically produced. The rPPG has been proven to be superior not only because subjects have no need to wear sensors, which may be suitable for cases where a continuous measure of HR is important (e.g., neonatal intensive care unit (ICU) monitoring, long-term epilepsy monitoring, burn or trauma patient monitoring, driver status assessment, and affective state assessment) [26]–[30], but also because the adopted cameras are low-cost, convenient, widespread and have the ability to access multiple physiological parameters simultaneously [18], [19], [31]–[33].

Consumer-level-camera-based rPPG was first proposed by Verkruysse *et al.* [18]. They demonstrated that HR could be measured from video recordings of the subject's face under ambient light using an ordinary consumer-level digital camera. Later, Poh *et al.* [19] proposed a linear combination of RGB channels to estimate the HR by employing blind source separation (BSS) methods. As an alternative, Sun *et al.* [34] proposed a framework of remote HR measurement during ambient light situations by employing joint time-frequency analysis. Since then, an increasing number of studies, based on realistic optical models and advanced signal processing techniques, have been conducted to remotely measure the PPG signals from facial videos [11], [20], [35]–[37]. The progress has been summarized in several relevant review articles from various aspects. Sun and Thakor [8] described the PPG measurement techniques from contact to noncontact and from point to imaging. Al-Naji and Chahl [38] provided a broad range of literature survey for remote cardiorespiratory monitoring,

Manuscript received July 30, 2018; revised October 14, 2018; accepted October 26, 2018. Date of publication November 29, 2018; date of current version September 13, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002802, in part by the National Natural Science Foundation of China under Grant 81571760, Grant 61501164, and Grant 61701160, and in part by the Fundamental Research Funds for the Central Universities under Grant JZ2016HGPA0731, Grant JZ2017HGTB0193, and Grant JZ2018HGTB0228. The Associate Editor coordinating the review process was Domenico Grimaldi. *(Corresponding author: Juan Cheng.)*

X. Chen is with the Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230026, China (e-mail: xunchen@ustc.edu.cn)

J. Cheng, R. Song, and Y. Liu are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: chengjuan@hfut.edu.cn; rcsong@hfut.edu.cn; yuliu@hfut.edu.cn).

R. Ward and Z. J. Wang are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: rababw@ece.ubc.ca; zjanew@ece.ubc.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2018.2879706

including Doppler effect, thermal imaging, and video camera imaging. Sikdar *et al.* [39] did a methodological review for contactless vision-guided pulse rate estimation updated to the year 2014, at which time most studies were still performed in a relatively stable environment. Hassan *et al.* [40] investigated both rPPG and ballistocardiography (BCG) estimation based on digital cameras, while the most recent study of rPPG in the review was reported in the year 2015. Most recently, the research focus has been shifted from demonstrating the feasibility of HR measurement under well-controlled and lablike conditions to more complex realistic conditions (e.g., including dynamical illumination variations and motion artifacts) [23], [41]. A large number of studies have been performed to reduce or eliminate the impact of noise artifacts resulting from the motions of the subject, facial expressions, skin tone, and illumination variations [31], [32], [42]–[44]. However, to the best of our knowledge, there has not been yet a thorough review of the very recent rPPG development that tackles the realistic issues of dynamical illumination variations and motion artifacts.

To fill this gap, this paper provides a timely, systematical review of the recent advances of rPPG. The main contributions of this paper are threefold. First, we provide a comprehensive review of all rPPG studies since it first appeared in 2008. Second, we summarize, compare, and discuss the methodological advancements of rPPG in detail, with a focus on solutions for illumination and motion-induced artifacts, from the signal processing perspective. Third, we present several specific prospects for future studies related to rPPG and its promising potential applications, hoping to share some new thoughts with the interested researchers.

The rest of this paper is organized as follows. In Section II, we describe the background of rPPG, including the optical model, the basic framework, and some recent research interests. The detailed progress on rPPG, with respect to the cases of varying illuminations and motions, is summarized, compared, and discussed in Sections III and IV. Future prospects of rPPG are introduced in Section V. Finally, in Section VI, conclusions are drawn.

II. Background of rPPG

A. Reflection Model of rPPG

When a light source illuminates an area of physical skin, quasi-periodical pulse-induced subtle color variations can be measured using a contact-free camera from a distance of up to several meters [23], [45], [46]. Without illumination variations and motion artifacts, color variations mainly refer to the blood volume changes in the microvascular tissue bed beneath the skin when the pulsatile blood propagates in the cardiovascular system within each heart beat circle. However, illumination variations could change both the intensity and the spectral compositions, while motion artifacts can cause the changes of the distance (angle) from (between) the light source to the skin tissue and to the camera, also leading to the changes of illumination intensity and spectral compositions. Consequently, the skin area measured by the camera has a varying color due to illumination-induced and motion-induced intensity/specular variations and pulse-induced subtle color changes. Assuming that the spectral compositions of the illuminance are fixed, illumination variations and motion artifacts can be reflected in the rPPG model in an optical and physiological sense, as illustrated in Fig. 1. As shown in [37], assuming that the processing duration of the recorded RGB image sequence is defined as T (s), and the number of the pixels included in the interested skin area is K, the reflection of the kth skin pixel in a recorded RGB image sequence can be defined as a time-varying function in the RGB channels

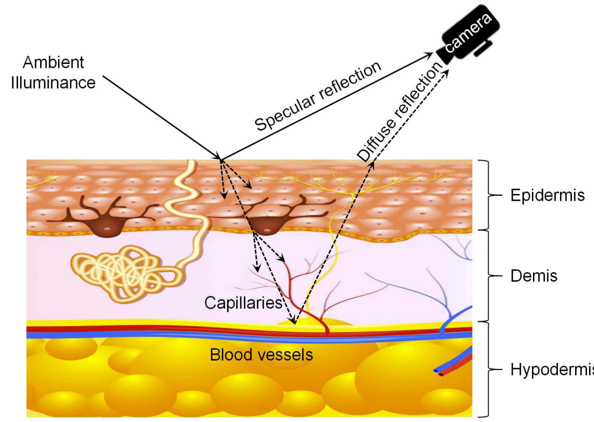

Fig. 1. Reflection model of rPPG.

$$C_k(t) = I(t) \cdot (\mathbf{v_s}(t) + \mathbf{v_d}(\mathrm{t})) + \mathbf{v_n}(\mathrm{t}), \quad 1 \le k \le K \tag{1}$$

where t represents the tth time and $1 \le t \le T$. $C_k(t)$ denotes the RGB channels (in column) of the kth skin pixel; $I(t)$ denotes the illumination intensity level; $\mathbf{v_s}(t)$ denotes the specular reflection and $\mathbf{v_d}(t)$ denotes the diffuse reflection. $I(t)$ is modulated by both $\mathbf{v_s}(t)$ and $\mathbf{v_d}(t)$. $\mathbf{v_n}(\mathrm{t})$ denotes the measurement noise of the camera sensor.

$\mathbf{v_s}(t)$ is a mirrorlike light reflection from the skin surface without pulsatile information and is time dependent since motion changes the distance (angle) from (between) the light source to the skin surface and the camera. Thereby, $\mathbf{v_s}(t)$ can be expressed as

$$\mathbf{v_s}(t) = \mathbf{u_s} \cdot (s_0 + s(t)) \tag{2}$$

where $\mathbf{u_s}$ denotes the unit color vector of the light spectrum; s_0 and $s(t)$ denote the stationary part and the varying part of the specular reflection, respectively. The varying part is mainly caused by motions.

$\mathbf{v_d}(t)$ is associated with the absorption and scattering of the light in skin tissues. In addition, $\mathbf{v_d}(t)$ is varied by the blood volume changes and can be written as

$$\mathbf{v_d}(t) = \mathbf{u_d} \cdot d_0 + \mathbf{u_p} \cdot p(t) \tag{3}$$

where $\mathbf{u_d}$ denotes the unit color vector of the skin-tissue; d_0 denotes the stationary diffuse reflection strength; $\mathbf{u_p}$ denotes the relative pulsatile strength while $p(t)$ denotes the pulse signals.

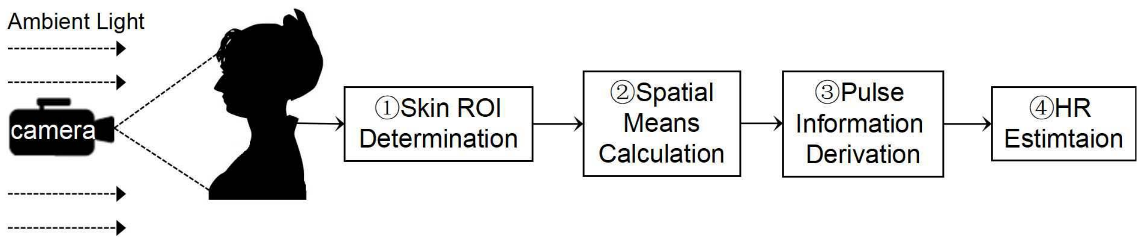

Fig. 2. Basic framework of rPPG measurement.

TABLE I
REPRESENTATIVE rPPG STUDIES UNDER WELL-CONTROLLED CONDITIONS

Ref.	ROI Defnition	ROI Detection	Parameter	Frame (fps)	Distance (m)	Methods	Main Characteristics
[18]	Forehead	Manually	HR, RR	15 or 30	1-2	Band-pass filter+FFT	It first shows the feasibility of rPPG by consumer-level camera; The green channel features the strongest PPG signals.
[19]	Face	VJ with 60% width	HR, RR, HRV	15, interpolating to 256	0.5	JADE-based ICA	It demonstrates the feasibility of multiple physiological parameters measurement using a simple webcam.
[57]	Palm	Fixed whole frame	HR, RR, HRV	200	0.4	Time-frequency	The estimations are statistically comparable to corresponding gold standards; The negative influence of low sample rate can be compensated via interpolation.
[47]	Face	Fixed spatial region	HR,RR	45	/	EVM	The HR can be well estimated by applying EVM to magnify either head motions or skin colors.
[45]	Face, Forehead	Manually	HR	20	1	PCA	Forehead ROI is recommended; The accuracy of using PCA is similar to ICA, but the time consumption is less; The order of the principal component representing PPG signal is random.
[48]	Face, Forehead	Manually	HR	15-25	0.4-0.8	PCA/ICA	HR derived by using PCA in YUV space proves feasible, similar to that in RGB space; Best results are obtained by using V signals recorded from forehead ROI.
[10]	Forehed	Manually	HR	15	0.5	Fixed signal weights and FFT	The signal weighting analysis demonstrates comparatively feasible for HR estimation.

B. Basic Framework of rPPG

Based on relevant rPPG studies in the literature, the corresponding basic framework can be summarized and described in Fig. 2. Such a framework is suitable for most of the rPPG estimation methods proposed for well-controlled conditions, except for the method proposed by Wu *et al* [47]. First, a camera is employed to capture the interested skin area of the body with a light source or just ambient illuminance. The skin region of interest (ROI) can be manually or automatically detected and tracked. Second, spatial single or multiple color channel mean(s) are calculated from the ROI [48], [49]. Third, signal processing methods (e.g., low-pass filtering and BSS methods) are applied to spatial mean(s) to derive the component including pulse information. Finally, fast Fourier transform (FFT) (or a peak detection algorithm) is usually applied to the component to estimate the corresponding frequency F_s [or the number of the peaks N_s during the processing duration T (s)]. The HR [in the form of beat per minute (bpm)] will be calculated as $60 \times F_s$ (or $N_s/T \times 60$).

Table I lists typical studies of rPPG-based HR measurement using consumer-level cameras under well-controlled conditions, meaning that the subjects are asked to keep stationary and the ambient illuminance is stable. Specifically, Verkruysse *et al.* [18] proposed to manually select the forehead ROI. Then, raw signals, calculated as the average of all pixels in the forehead ROI, were bandpass filtered using a fourth-order Butterworth filter. HR was then extracted from the frequency content using FFT for each 10-s window. The authors have found that different channels of the RGB camera feature different relative strengths of PPG signals and the green channel contains the strongest pulsatile signal. This observation is consistent with the fact that hemoglobin light absorption is most sensitive to oxygenation changes for green light.

Later, Poh *et al.* [19] presented a simple and low-cost method to measure physiological parameters, e.g., HR, RR, and HRV, by using a basic webcam. The pulse signal was extracted by applying an independent component analysis (ICA)-based BSS method to three RGB color channels of facial video recordings to derive three independent components. In their work, the facial ROI was defined as a rectangle bounding box, which was automatically identified by Viola–Jones (VJ) face detector [50]. FFT was then applied to the strongest pulsatile component and the largest spectral peak in the frequency band (0.75 to 4 Hz) was selected, corresponding to the HR in the normal range of 42 to 240 bpm. High correlations were achieved between the estimated measurements and the reference (the ground truth data) for the above-mentioned physiological parameters under well-controlled conditions.

Kwon *et al.* [51] reproduced Poh's approach and developed the FaceBEAT application on a smartphone. As an alternative, Lewandowska *et al.* [45] proposed using principal component analysis (PCA) to define three independent linear combinations of the color channels and demonstrated that PCA could be as effective as ICA. Later, Yu *et al.* [52] demonstrated the feasibility of PCA in dynamical HR estimation. The pros and cons of applying ICA and PCA, as well as other methods (i.e., direct frequency analysis, autocorrelation, and cross correlation) for the analysis of rPPG-based HR were compared in [53]. In addition, Rumiński [48], from the same group as Lewandowska, demonstrated the possibility of estimating HR from rPPG signals in the YCrCb (YUV) space using both ICA-based and PCA-based methods. The experimental results showed that the best HR estimation performance can be achieved by applying PCA to the V channel obtained from a forehead ROI.

An alternative way to measure HR from videos is skin or motion magnification framework. Typically, Wu *et al.* [47] proposed a Eulerian video magnification (EVM) framework to estimate HR by visualizing the flow of the blood, which was originally difficult or impossible to be seen with the naked eye. Since EVM has the ability of revealing subtle-motion changes based on spatiotemporal processing [47], the HR could be measured without feature tracking or motion estimation. Other skin/motion color magnification methods for measuring HR have been studied [54], [55]. It was suggested that skin color magnification algorithms followed by BSS-based signal processing methods would yield a better performance [56].

Furthermore, Sun *et al.* [57] proposed to investigate the feasibility of remote assessment of HR, RR, and HRV by applying a time-frequency representation method to the video recordings of the subjects' palm regions. All videos were recorded at a rate of 200 frames per second (fps) under the resting conditions to minimize motion artifacts. The authors demonstrated that 200-fps iPPG system could provide a closely comparable measurement of HR, RR, and HRV to those acquired from contact PPG references. It was also reported that the negative influence of a low initial sample rate could be compensated by interpolation [57]. Thereby, the frame number of the digital camera from 15 to 30 fps can be enough for the noncontact HR measurement [34], [35], [43].

C. Recent Interests in rPPG

The terminology of rPPG-based HR measurement has not yet been unified. Referring to the review in [23] and keywords in most popular articles, we chose *rPPG*, *remote PPG*, *iPPG*, *imaging*, *noncontact*, *contactless*, *contact free*, *camera-based*, *video-based* and HR to search related studies using *web of science*. After excluding the conference papers, there are 111 most relevant journal articles. Fig. 3 shows the number of articles published per year. It can be seen that rPPG techniques have drawn increasing attention from researchers. Most of such papers presented methodological solutions for suppressing the artifacts induced by illumination variations and body movements. We, therefore, later mainly review the rPPG studies from these two aspects, respectively.

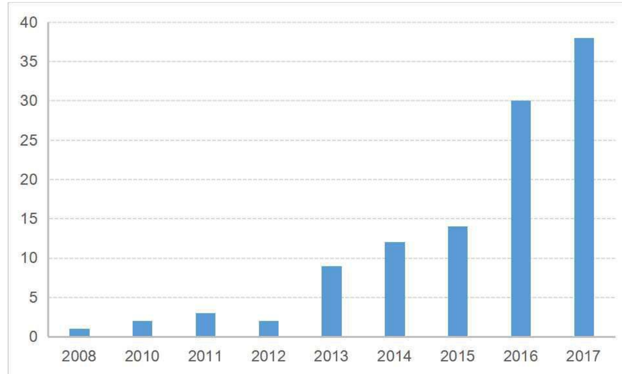

Fig. 3. Number of rPPG journal papers published per year.

III. ILLUMINATION-VARIATION-RESISTANT SOLUTIONS

In this section, rPPG studies, aiming at eliminating the impact of illumination variations, will be reviewed. Relevant representative works are listed in Table II.

A. Related Work

To suppress the influence of illumination variations, one possible way is adopting infrared cameras. For instance, Jeanne *et al.* [65] took advantages of infrared cameras to estimate HR under highly dynamic light conditions. As for RGB camera solutions, Xu *et al.* [66] proposed to extract rPPG signals by the usage of the Lambert–Beer law. They tested the feasibility of estimating HR under different illumination levels and reported a satisfactory performance.

When capturing facial RGB videos of subjects under illumination variation situations, both the periodic variation of reflectance strength corresponding to pulsatile information and the changing illumination are recorded in the raw RGB signals. Chen *et al.* [67], [68] applied an illumination-tolerant method based on ensemble empirical mode decomposition (EEMD) to the green channel for separating real cardiac pulse signals from the environmental illumination noise. The framework, using EEMD followed by a multiple-linear regression model, was later employed to evaluate HR for reducing the effects of ambient light changes [58]. Lam and Kuno [59] assumed that HR extraction from facial subregions could be treated as a linear BSS problem. With the assistance of the skin appearance model, which describes how illumination variations and cardiac activity affect the appearance of the skin over time, HR could be well estimated by randomly selecting pairs of traces in the green channel and performing majority voting.

When illumination variations occur in certain cases, both the face and the background regions contain similar variation patterns. Several HR measurement methods take the

TABLE II
REPRESENTATIVE rPPG STUDIES AGAINST ILLUMINATION VARIATIONS

Scheme	Ref.	Camera setup	Color Channel	Illumination situations	Methods	Major Characteristic
Scheme I	[58]	30 fps, distance of 0.1-0.25 m to the brow	G	indoors with ambient light changing	EEMD +MLR	EEMD has the ability of reducing the effect of ambient light changing; The proposed method outperforms ICA in terms of four performance metrics. The ambient light changing is constant.
	[59]	61 fps, public database MAHNOB-HCI	G	watching movies in front of the laptop	Select random patches	The pose-free facial landmark fitting tracker [62] is employed; Many pairs of points in ROI are randomly selected and good of them following some principles are utilized to extract HR by employing ICA. The performance is better than Poh2011 [19] and Li2014 [43].
Scheme II	[41]	61 fps, public database MAHNOB-HCI	G	watching movies in front of the laptop	NLMS	Face ROI is detected and tracked by DRMF and KLT; The background ROI is treated as a noise reference; The proposed method outperforms Poh2010 [9], Kwon2012 [53], Poh2011 [19] and Balakrishnan2013 [63], achieving an average error rate of 6.87% on all 527 samples of MAHNOB-HCI.
	[62]	59.94 fps	RGB and G	watching movie clips	MOCF	A rectangle box is manually selected as the face ROI, while the G channel of the movie is employed to calculate the brightness; The method shows better performance in the case of low-dynamic movie than the high-dynamic movie when the frequency band of the artifact is lower than HR.
	[63]	12 fps, distance of 1 m	G	strong fluorescent lights in clinical environments	AR modelling and pole cancellation	Face is registered and tracked; AR modelling to both face and background ROIs can eliminate the unwanted illumination variation components. The AR might be challenged by the periodically changing illuminance.
	[32]	30 fps, distance of 0.5 m	RGB and G	dynamic changing artificial and realistic illumination variations	JBSS + EEMD	JBSS technique helps to extract the common underlying illumination variation sources and clean facial ROI can be reconstructed. The performance is better than ICA-based, NLMS, MOCF and EEMD-based methods.
	[64]	30 fps, distance of 0.5 m	RGB	dynamic changing artificial illumination variations	PLS + MEMD	PLS-MEMD has the ability of suppressing the impact of illumination variations for remote HR measurement.

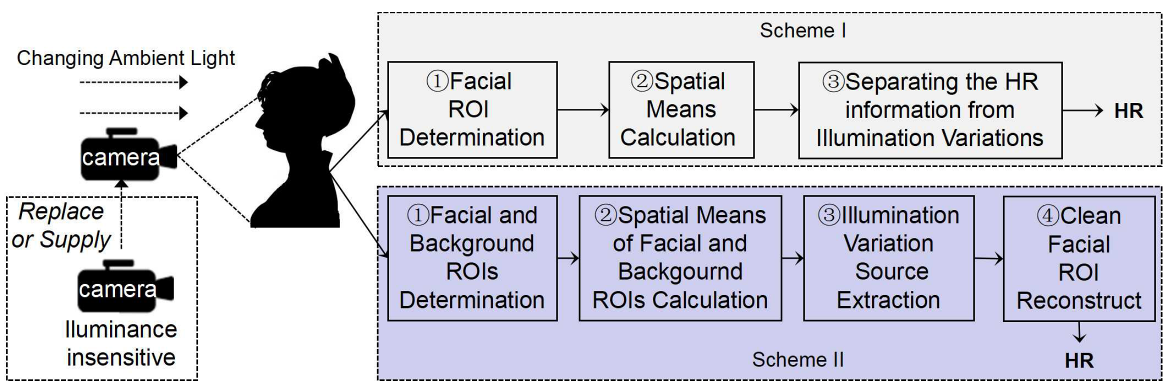

Fig. 4. Two schemes of HR estimation when tackling illumination variations using regular RGB cameras.

background region as a noise reference to rectify the interference of illumination variations. Li *et al.* [41] proposed an illumination rectification method based on the normalized least mean square (NLMS) adaptive filter, with the assumption that both the facial ROI and the background were Lambertian models and shared the same light sources. Therefore, the background can be treated as an illumination variation reference and could be filtered from the facial ROI to rectify the interference of illumination variations when subjects watched movies. Lee *et al.* [62] also assumed that the raw green trace rPPG signals from the facial video contained both pulsatile information and illumination variations when a subject watches movies in front of a laptop in a darkroom. They proposed subtracting illumination artifacts (using the extracted brightness variation signals from the movie signals) from the raw green trace rPPG signals by using the least square curve fitting method. Experimental results showed that the root mean square error (RMSE) of the estimated HR decreased. Tarassenko *et al.* [63] proposed a novel method to cancel out the aliased frequency components caused by the artificial light flicker using autoregressive (AR) modeling. The poles, corresponding to the aliased components of the artificial light flicker frequency spectrum derived by applying AR to the background ROI, are also presented in the AR model of the face ROI. Therefore, these poles could be canceled from the face ROI to find the regular HR frequency. Experimental results showed that AR modeling with pole cancellation was even suitable for strong fluorescent lights. However, due to the fact that the AR modeling is a spectral analysis method, it might be challenged by periodical illumination variations. Recently, based on the same assumption that both facial ROI and background ROI contain similar illumination variation patterns, Cheng *et al.* [32] proposed an illumination-robust framework using joint BSS (JBSS). Specifically, the authors denoised the facial rPPG signals by applying JBSS to both the ROIs to extract the underlying common illumination variation sources. Followed by EEMD, the target intrinsic mode functions (IMFs), including cardiac wave signals, were then utilized to estimate HR. The proposed method was shown effective under a number of dynamically changing illumination variation situations.[1] Furthermore, Xu *et al.* [64] proposed a novel framework based on partial least squares (PLS) and multivariate EMD (MEMD) to effectively evaluate HR from facial rPPG signals captured during illumination changing conditions. The main function of the PLS is to extract the underlying common illumination variation sources within facial ROI and background ROI, while the MEMD has the ability of extracting common modes across multiple signal channels when considering the dependent information among the RGB channels [64], [69].

B. Basic Framework and Summary

Apart from adopting illumination insensitive cameras, such as infrared cameras, two main schemes tackling illumination variations when employing RGB cameras can be concluded according to above-mentioned studies and shown in Fig. 4.

The first scheme is based on signal processing methods to separate illumination variation signals from the pulse signals. A typical illumination-tolerant solution is the EEMD algorithm, which has already been demonstrated effective during denoising situations [70], [71]. Chen *et al*. [67] applied the EEMD algorithm to the green channel for separating real cardiac pulse signals from the environmental illumination noise. The major steps are described in Scheme I in Fig. 4. First, the facial ROI is detected and tracked. Second, the spatial means of the RGB channels or only the spatial mean of the green channel is calculated from the derived facial ROI. Third, the HR information is extracted by either using EEMD to derive the target IMF representing cardiac signals or applying BSS to randomly select good local regions containing cardiac signals. Thereby, HR can be estimated from the target IMF, or from multiple local regions combined with a majority voting scheme. However, EEMD could be challenged by periodical illumination variations, especially if the frequency of which is close to the normal HR frequency range (typically from 0.75 to 4 Hz).

The second scheme is mainly based on the assumption that the raw traces of rPPG signals (e.g., facial region) contain both blood volume variations caused by the cardiac pulse and temporal illumination variations. Such illumination variations can be considered as a noise reference derived from nonskin background (or the brightness of the video) regions to denoise the rPPG signals derived from skin regions. The detailed procedures are shown in Scheme II in Fig. 4. First, both facial and background ROIs are determined, including ROI detection and tracking. Second, the spatial means of the color channels are, respectively, calculated from both facial and background ROIs. Third, the background ROI is treated as a noise reference to extract the illumination variation source, by using AR, NLMS, least square curve fitting, JBSS, or PLS. Fourth, the illumination variation source is later subtracted from the facial ROI to reconstruct the illumination-variation-free facial ROI. Finally, the HR is measured from the cleaner facial ROI. The studies demonstrated that the approach based on selecting random patches [59] is better than ICA-based method [19] and NLMS-based method [41]. The JBSS-EEMD method performs better than ICA-based [19], NLMS-based [41], multi-order curve fitting (MOCF)-based [62], and EEMD-based methods [67]. The performance of such type of rPPG methods depends on the degree of the similarity when extracting the common-underlying illumination variation source from both skin and nonskin regions. Several novel similarity measures in kernel space, proposed by Chen *et al.* [72], [73] can be used for robust filtering and regression. It should be noted that when the variation in the facial ROI is different from that in the nonskin ROI, methods in Scheme II might be ineffective. For instance, someone bursts into the room and stands behind the subject when the subject is watching a movie. To address this concern, an appropriate nonskin ROI needs to be alternatively utilized as the noise reference, such as placing a whiteboard near the subject [32], [41].

In addition, we would like to mention that almost all the above-mentioned studies aiming at eliminating the influence of illumination variations require the subject to keep still, which means that motion artifacts are out of consideration in such studies. However, in realistic applications, both illumination variations and motion artifacts are inevitable, and perhaps motion artifacts are more common. Thereby, these methods originally designed to suppress illumination variations could be challenged and need to be further improved by addressing the motion artifact concern.

IV. Motion-Robust Solutions

An increasing number of works have been done to suppress the impact of motion artifacts [79]–[83]. As shown in Fig. 1, the changes of the distance (angle) from (between) the face and the camera caused by motions can be modeled as the optical model [21], [37]. It was noted that the camera quantization noise can be reduced by spatially averaging the RGB values of all skin pixels within the facial ROI, in which case $\mathbf{v_n}(t)$ can be negligible [42]. According to (1), the averaged temporal

[1] Corresponding code can be downloaded from http://www.escience.cn/people/chengjuanhfut/admin/p/Codes

TABLE III
REPRESENTATIVE rPPG STUDIES AGAINST MOTION ARTIFACTS

Category	Ref.	Motion situations	Methods	Major Characteristic
BSS-based	[9]	Sit-move-naturally but to avoid large or rapid motions while still seated	ICA	The RMES corresponding to motion situations is reduced from 19.36 bpm to 4.63 bpm and correlation coefficient (CC) increases from 0.15 to 0.95 when using ICA compared to using raw green channel without ICA.
	[74]	The subject performed cycling exercise at 15 and 25 km/h	SCICA	The proposed method is well suited for motion-tolerant RR and HR measurements even during high-intensity physical exercise situations.
	[75]	DEAP database	JBSS	Compared to Poh2011, the proposed JBSS method achieves a better performance, especially adopting C-MCCA with an adaptive correction.
	[46]	Stationary, facial expression with head rotation; talk and stationary; different illumination levels. All scenarios including UAV camera movement	CEEMDAN + CCA	The proposed method can remove noise acquired from the illumination variations, subjects movement and camera movement. Both HR and RR can be better extracted compared to ICA and PCA methods.
Model-based	[76]	Yawing, pitching, rolling surging, swaying and heaving motions	G-R	The proposed method is suitable for HR estimation with motion artifacts in terms of four evaluation metrics (i.e., Bias, Bland-Altman plot, Pearson's CC, ANOVA).
	[21]	Subjects exercising on a stationary bike and a stepping device	CHROM	Both the percentage of the highest spectral peak corresponds to the actual HR and the SNR demonstrates that the CHROM is more robust than ICA/PCA-based BSS in motion situations.
	[77]	Subjects exercising on a range of fitness devices	PBV	PBV has a much better motion robustness than ICA/PCA-based BSS and CHROM. However, it requires accurate knowledge of the blood volume pulse vector for correct noise suppression.
	[78]	Different skin-tone, different illuminance conditions, and fitness exercise	2SR	2SR outperforms ICA-based approach, CHROM and PBV. However, multiple skin-pixels are needed in the spatial domain and the skin mask should be less noisy for 2SR.
	[37]	The same as in [80]	POS	G, G-R, PCA, ICA, CHROM, PBV, 2SR are implemented and compared. POS obtaines overall the best performance in the form of SNR.

RGB signals, marked as $C(t)$, can be written as

$$C(t)=I_0 \cdot (1+i(t)) \cdot (\mathbf{u_s} \cdot (s_0 + s(t)) + \mathbf{u_d} \cdot d_0 + \mathbf{u_p} \cdot p(t)). \tag{4}$$

Since all ac-modulation terms are much smaller than the dc term, the product modulation terms (e.g., $p(t) \cdot i(t)$ and $s(t) \cdot i(t)$) can be ignored. Therefore, $C(t)$ can be approximated as

$$C(t) = I_0 \cdot \mathbf{u_c} \cdot c_0 + I_0 \cdot \mathbf{u_c} \cdot c_0 \cdot i(t) + I_0 \cdot \mathbf{u_s} \cdot s(t) + I_0 \cdot \mathbf{u_p} \cdot p(t) \tag{5}$$

where $\mathbf{u_c} \cdot c_0 = \mathbf{u_s} \cdot s_0 + \mathbf{u_d} \cdot d_0$ and $i(t)$ is the time-varying part of the intensity strength.

It can be seen from (5) that $C(t)$ is a linear combination of three signals $i(t)$, $s(t)$, and $p(t)$. Such three signals are zero-mean signals. Depending on whether knowing the prior information of components, motion-robust methods can be mainly divided into two categories, called BSS-based and model-based methods. BSS-based methods might be ideal for demixing $C(t)$ to sources for pulse extraction without prior information, while model-based methods can use knowledge of the color vectors of different components to control the demixing. Some representative studies are listed in Table III. Besides these two categories of methods, the methods employed to determine and track ROIs can be treated as motion compensated strategies. In addition, some other motion-robust methods are also reviewed. Such four categories of motion robust solutions are shown in Fig. 5.

A. BSS-Based Methods

1) Conventional BSS: BSS refers to the recovery of unobserved signals or sources from a set of observed mixtures without prior information with respect to the mixing process. Generally, observations are the outputs of sensors and each output is a combination of sources [84]. One typical method of BSS is ICA, which has been proven feasible in many fields [85]. Based on the assumption that R, G, and B channel signals are actually a linear combination of the pulse signal and other signals, Poh *et al.* [9] proposed a joint approximate diagonalization of eigen-matrix (JADE)-based ICA algorithm to remove the correlations and the higher order dependence between RGB channels to extract the HR component during both sit-still and sit-move-naturally conditions. The RMSE corresponding to motion situations was reduced from 19.36 to 4.63 bpm, demonstrating the feasibility of ICA for HR evaluation. Sun *et al.* [74] introduced a new artifact-reduction methods consisting of planar motion compensation and BSS. Their BSS mainly referred to the single channel ICA (SCICA). The performance was evaluated through the facial video captured from a single volunteer with repeated exercises, which revealed that HR could be tracked with the proposed method. Monkaresi *et al.* [86] proposed a machine learning approach combined with the same ICA as Poh, to improve

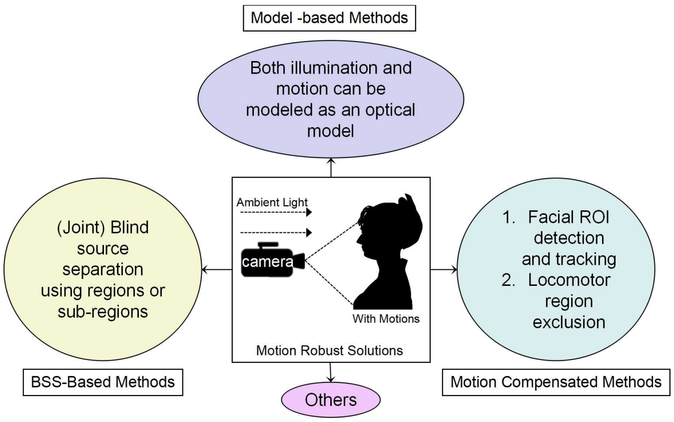

Fig. 5. Four categories of motion robust solutions for rPPG techniques.

the accuracy of HR estimation in naturalistic measurements. Wei *et al.* [22] proposed to estimate HR by applying a second-order BSS to the six-channel RGB signals that yielded from dual facial ROIs. BSS-based methods had somewhat the ability of tolerating motions but still showed limited improvement, especially in dealing with severe movements [87]. Since the orders of the extracted components via BSS are random, usually FFT is utilized to determine the most probable HR frequency. Thus, BSS-based methods cannot deal with cases in which the frequency of the periodical motion artifacts falls into the normal HR frequency range. Recently, Al-Naji *et al.* [46] proposed the combination of complete EEMD with adaptive noise (CEEMDAN) and canonical correlation analysis (CCA) to estimate HR from video sequences captured by a hovering unmanned aerial vehicle (UAV). The proposed CEEMDAN followed by CCA method achieved a better performance than that using ICA or PCA methods in the presence of noises induced from illumination variations, subject's motions, and camera's own movement.

2) Joint BSS: Conventional BSS techniques are originally designed to handle one single data set at a time, e.g., decomposing the multiple color channel signals from the single facial ROI region into constituent independent components [32]. Recently, color channel signals from multiple facial ROI subregions were employed for more accurate HR measurement [22], [31]. With the increasing availability of multisets, various joint BSS (JBSS) methods have been proposed to simultaneously accommodate multisets. Chen *et al.* [88] provided a thorough overview of representative JBSS methods. Several realistic neurophysiological applications from multiset and multimodal perspectives highlighted the benefits of the JBSS methods as effective and promising tools for neurophysiological data analysis. The goal of JBSS is to extract underlying sources within each data set and meanwhile keep a consistent ordering of the extracted sources across multiple data sets [85]. Guo *et al.* [27] first introduced the JBSS method into rPPG fields, mainly applying the independent vector analysis (IVA) to jointly analyze color signals derived from multiple facial subregions. Preliminary experimental results showed a more accurate measurement of HR compared to ICA-based BSS method. Later, Qi *et al.* [75] proposed a novel method for noncontact HR measurement by exploring correlations among facial subregion data sets via JBSS. The testing results on a large public database also demonstrated that the proposed JBSS method outperformed previous ICA-based methodologies.

The HR estimation by using JBSS methods is preliminary. In the future, other types of multisets in addition to color signals from facial subregions and even multimodal data sets can be utilized for more accurate and robust noncontact HR measurement via JBSS.

B. Model-Based Methods

Since the information of color vectors can be utilized by model-based methods to control the demixing for component derivation, the model-based methods have in common that the dependence of $C(t)$ on the averaged skin reflection color channels can be eliminated [37]. The model-based methods typically refer to methods based on the chrominance model (CHROM), methods using blood volume pulse signature (PBV) to distinguish pulse signals from motion distortions [77], and methods based on a plane orthogonal to the skin (POS) [37].

de Haan and Jeanne [21] developed a CHROM to consider diffuse reflection components and specular reflection contributions, which together made the observed color varied depending on the distance (angle) from (between) the camera to the skin and to the light sources. Therefore, the impact of such motion artifacts could be eliminated by a linear combination of the individual R, G, and B channels. Experimental results demonstrated that CHROM outperformed previous ICA-based and PCA-based methods in the presence of exercising motions. Relying on the same CHROM method, Huang *et al.* [89] applied an adaptive filter (taking the face position as the reference) followed by discrete Fourier transform (DFT) to rPPG signals. Experimental results showed the motion-robust feasibility of the proposed method even under the situation that subjects performed periodical exercises on fitness machines. Still relying on the CHROM method as a baseline, Wang *et al.* [90] proposed a novel framework to suppress the impact of motion artifacts by exploiting the spatial redundancy of image sensors to distinguish the cardiac pulse signal from the motion-induced noise.

Afterward, de Haan and van Leest [77] proposed a PBV-based method for improving the motion robustness. The PBV-based method utilized the signature of blood volume change to distinguish pulse-induced color changes from motion artifacts in temporal RGB traces. Experimental results during the conditions that subjects were exercising on five different fitness-devices showed a significant improvement of the proposed method compared to the CHROM-based method.

Recently, Wang *et al.* [37] proposed another model-based rPPG algorithm, referred to POS. The POS method defined a POS tone in the temporally normalized RGB space for pulse extraction. A privately available database, involving challenges regarding different skin tones, various illuminance, and motions, was utilized for the benchmark of evaluating HR methods including G (2007) [18], ICA (2011) [19], PCA

(2011) [45], CHROM [21], PBV [77], 2SR [78], and POS [37]. POS obtained the overall best performance among them, mainly due to the fact that the defined POS tone was physiologically reasonable. It made POS especially advantageous in fitness challenges where the skin-mask was noisy. They also proved that POS and CHROM performed well during both stationary and motion situations although both of them may have problems in distinguishing the pulsatile component from close amplitude-level distortions, whereas the PBV was particularly designed for motion situations.

C. Motion Compensated Methods

The motion mentioned here is mainly referred to global rigid and local nonrigid motions. Rigid motions usually include head translation and rotation while nonrigid motions generally refer to eye blinking, emotion expressing, and mouth talking. In general, reliable ROI (i.e., facial ROI) detection and tracking is one of the crucially important steps for rPPG-based HR estimation, which can also be treated as a global motion compensation way to guarantee the accuracy of HR estimation [32], [41], [76]. Meanwhile, methods aiming to exclude the regions that are easier to be locomotor can be regarded as local motion compensation strategies [9], [41].

1) Global Motion Compensation: All the exposed skin areas can be utilized as the ROI, such as face, forehead, cheese, palm, finger, forearm, and wrist [57], [81], [91], [92]. In this paper, the ROI mainly refers to the whole face or subregion(s) of the face.

Preliminary rPPG studies tended to manually select ROIs [18], [48], [74]. Later, some researchers employed the popular VJ face detector to determine facial ROIs [19], [43]. Without any ROI detecting and tracking algorithms, even minor movements of the observed regions were not permitted. In a sense, all the studies focusing on refined ROI determination by employing face detection or/and tracking strategies can be treated as the compensation of global motions.

In the beginning, a zoomed out version of the whole rectangle box obtained by using VJ face detector was utilized to determine the face ROI, avoiding nonfacial pixels [9], [43]. Later, benefiting from the development of image processing techniques, more and more advanced face detection (mainly referring to feature landmark localization) and tracking algorithms have been introduced to rPPG field. For instance, discriminative response map fitting (DRMF), proposed in [93], was employed in [41] to automatically detect the 66 facial landmarks on the face.[2] In addition, Kanade–Lucas–Tomasi (KLT) was then employed to track these feature landmarks frame by frame. By this means, the global motions such as shaking your head while keeping the frontal face were acceptable. Tulyakov *et al.* [94] adopted the supervised descent method to define the facial ROI by locating and tracking facial landmarks. Cheng *et al.* [95] used the approximated structured output learning approach in the constrained local model technique to efficiently detect facial landmarks[3] [32] and KLT was still the tracker. Lam and Kuno [59] stated that the pose-free facial landmark fitting tracker proposed by Yu *et al.* [60] was very effective[4] even suitable for large range of motion situations.

Although all the above-mentioned face detection and tracking algorithms had the ability of tolerating global motions, frontal faces in most cases must be guaranteed, which may not meet the practical usage of rPPG applications. Recently, an efficient facial landmark localization algorithm proposed in [96] was employed to detect facial ROI even under different nonfrontal face viewpoint.[5] In addition, since several video-based rPPG frameworks have already been implemented on mobile phones, the execution speed of the detection and tracking algorithms should be acceptable. In this case, the circulant structure of tracking-by-detection with kernels, developed by Henriques *et al.* [97] could be considered owing to the processing ability of hundreds of fps. We believe that with more advanced facial landmark detection and tracking algorithms employed for ROI detection in the wild, the performance of rPPG-based HR estimation will be further promoted. Other effective facial landmark detection and tracking algorithms can be found in [98].

2) Local Motion Compensation: After the facial ROI detection, the spatial channel means of all the pixels within each ROI are usually calculated and temporally concatenated to compose rPPG signals. Such averaging will guarantee the quality of rPPG signals unless the noise level is comparable. However, the image-by-image variations in skin pixels from a mouth region of a talking subject, or from a blinking eye region might be more stronger than that from the stationary forehead. Thus, with the consideration of eliminating the influence of local motions, the detected rectangle box is segmented to only keep the relatively stationary forehead or cheek regions. In addition, several out of all the detected facial landmarks will further be selected to exclude eye region, mouth region, and other regions that are prone to be locomotor [31], [99]. In addition, studies aiming to find the optimal facial ROI could achieve a better HR estimation [100]–[103].

Besides the above-mentioned methods, other methods can also compensate local motions. For instance, Wang *et al.* [90] exploited the spatial redundancy of image sensors to distinguish the pulse signals from motion-induced noise. The possibility of removing the motion artifacts was based on the observation that a camera could simultaneously sample multiple skin regions in parallel, and each of them could be treated as an independent sensor for HR measurement. Specifically, a pixel-track-complete (PTC) method extended the face localization with spatial redundancy by creating a pixel-based rPPG sensor and employing a spatiotemporal optimization procedure. Experimental results, derived from 36 challenging benchmark videos consisting of subjects that differed in gender, skin types, and motion types, demonstrated that the proposed PTC method led to significant motion robustness improvement and excellent computational efficiency.

[2]The code can be downloaded in https://ibug.doc.ic.ac.uk/resources/drmf-matlab-code-cvpr-2013/

[3]The detailed information and related code can refer to http://kylezheng.org/facial-feature-mobile-device/

[4]The corresponding code can be found in http://research.cs.rutgers.edu/∼xiangyu/face_align.html

[5]The code can be downloaded in https://sites.google.com/site/chehrahome/

D. Other Methods

Apart from the above-mentioned three types of motion-robust methods, wavelet transform was another effective strategy for motion-tolerant HR estimation. For instance, Bousefsaf *et al.* [80] obtained PPG signals from facial video recordings using a continuous wavelet transform and achieved high degrees of correlation between physiological measurements even in the presence of motion. The combination of BSS and machine learning technique had an excellent performance when selecting the best independent component for HR estimation during both controlled lablike tasks and naturalistic situations [86]. By considering a digital color camera as a simple spectrometer, Feng *et al.* [76] built an optical rPPG signal model to clearly describe the origins of rPPG signals and motion artifacts. The influences of motion artifacts were later eliminated by using an adaptive color difference operation between the green and red channels. Immediately, following the PBV-based method, Wang *et al.* [78] proposed a conceptually novel data-driven rPPG algorithm, namely, spatial-subspace rotation (2SR), to improve the motion robustness. Numerical experiments demonstrated that given a well-defined skin mask, the proposed 2SR method outperformed ICA-based, CHROM-based, and PBV-based methods in challenges of different skin tones and body motions. In addition, the proposed 2SR algorithm took advantages of simplicity and easy extensibility. In addition, Fallet *et al.* [16] designed a signal quality index (SQI) and demonstrated the feasibility of SQI as a tool to improve the reliability of iPPG-based HR monitoring applications.

E. Dealing With Both Illuminance and Motions

Till now, many researchers focused on simultaneously dealing with both illumination variations and motion artifacts. Li *et al.* [41] proposed a novel HR measurement method to reduce the noise artifacts in the rPPG signals caused by both illumination variations and rigid head motions. The problem of rigid head motions was first solved by using DRMF and KLT algorithms for face detection and tracking. The NLMS filter was then employed to reduce the interference of illumination variation by treating the green value of background as a reference. However, the signals corresponding to nonrigid motions were segmented and sheared out of the analysis, which might contain significant information related to physiological status. To overcome the difficulty of contactless HR detection caused by subjects motions and dark illuminance, Lin *et al.* [87] proposed to detect subjects motion status based on complexion tracking and filter the motion artifacts by motion index (MI). The near-infrared (NIR) LEDs was also employed to measure the HR in a dark environment.

The above two studies provided solutions to suppress the impact of illumination variations and motion artifacts independently. To deal with them simultaneously, Kumar *et al.* [31] recently reported that a weighted average over skin-color variation signals from different facial subregions (i.e., rejecting bad facial subregions contributing large artifacts) helped to improve the signal-to-noise ratio (SNR) of video-based HR measurement in the presence of different skin tones, different lighting conditions, and various motion scenarios. Profiting from the mathematical optical model that treated both illumination variations and motion artifacts as optical factors, all the model-based methods can deal with the impacts of both synchronously.

V. Future Prospects

Since video-based rPPG is a low-cost, comfortable, convenient, and widespread way to measure HR, it is of great potential for circumstances where a continuous measure of HR is important and physical contact with the subject is not preferred or inconvenient, i.e., neonatal ICU monitoring [26], [104], [105], long-term monitoring, burn or trauma patient monitoring, driver status assessment [106], [107], and affective state assessment [108]. In order to accurately achieve the remote HR measurement anytime and anywhere, future prospects of rPPG include the following aspects.

A. Use Prior Knowledge

With the help of recently proposed mathematical models [37], the commonalities and differences between existing rPPG methods in extracting HR can be better understood. In general, each method might be more appropriate under some assumptions for certain specific situations. The first assumption of the mathematical model is that the light source has a constant spectral composition but varying intensity, which indicates that the changing of the spectral composition of the light source will be an additional challenge. In this case, if such spectral composition changing information is prior known, a specialized method can be designed to better estimate HR. In addition, the conventional BSS-based methods help to demix the raw averaged RGB channels into independent or principal components without any prior information, while the generally utilize the information of color vectors to control the demixing. Thereby, the performance of HR measurement achieved by model-based methods is usually better than that by conventional BSS-based methods. Furthermore, the data-driven based rPPG methods can achieve an even better performance by creating a subject-dependent skin-color space and tracking the hue change over time. Hereto, including accurate knowledge or soft priors made both model-based and data-driven-based rPPG methods more robust to motion artifacts when compared to conventional BSS-based methods. It is generally accepted that when developing a robust rPPG engine for a broad range of applications, typical properties of rPPG should be considered. This suggests that certain known information can be utilized as a prior to improve the optical model. Also, considering that BSS techniques can incorporate prior information, such semi-BSS techniques might be a promising attempt to eliminate the impact of artifacts [109].

B. Establish Public Database Benchmark

In practice, another challenge in developing robust HR measurement approaches is the lack of publicly available data sets recorded under realistic situations. In other words, most papers published in recovering HR from facial videos have been assessed on privately owned databases. It is noteworthy that several public databases originally designed for emotion recognition or analysis using both physiological and video signals have been utilized as the benchmark to evaluate the performance of existing rPPG methods [113]. The most popular and challenging one is MAHNOB-HCI, which is a multimodal database recorded in response to affective stimuli with the goal of emotion recognition and implicit tagging research [114]. In the MAHNOB-HCI data set, the face videos, audio signals, eye gaze, and peripheral/central nervous system physiological signals including HR are synchronized recorded, which is obviously suitable for rPPG evaluation [40], [41], [59], [113]. Several researchers have already evaluated their own algorithms on this database. For instance, Li *et al.* [41] tested their method by using face tracking and NLMS adaptive filtering methods on the public database MAHNOB-HCI, demonstrating the feasibility of countering the impact of illumination and motion artifacts. In their testing, each of 27 subjects has 20 frontal face videos, and altogether 527 videos (excluding 13 lost-information cases) were available. They chose 30 s (frame 306 to 2135, video frame rate: 61 fps) from each video for the test. Lam *et al.* chose the same videos as Li *et al*. (excluding those without ECG ground truth) from MAHNOB-HCI to evaluate Li2014, Poh2011, and their own proposed method (BSS combined with selecting random patches), and reported that their proposed method outperformed other two methods. Lam and Kuno [59] recently published a reproducible study on remote HR measurement by comparing CHROM, Li2014, and 2SR methods on both publicly available MAHNOB-HCI and self-established COHFACE. A thorough experimental evaluation of the three selected approaches was conducted, demonstrating that only CHROM yields a stable behavior during all experiments but highly depends on the associated optimization parameters. It should be noted that the maximum Pearsons correlation coefficient was only 0.51 under all evaluation conditions, and thus, it is clear that more advanced rPPG algorithms or target-oriented rPPG algorithms are still needed [113].

Another useful database is DEAP, which is a public multimodal database for the analysis of human affective states in terms of levels in arousal, valence, like/dislike, dominance, and familiarity. It provides electroencephalography and other peripheral physiological signal recordings of 32 participants under designated multimedia emotional stimuli [115]. DEAP database recently has been utilized by Qi *et al.* [75] to evaluate the performance of their proposed JBSS-based rPPG algorithm. Their results showed that JBSS outperformed ICA-based methods.

However, both MAHNOB-HCI and DEAP involve illumination variations related to the movie itself and motions related to the reaction of the induced emotions. They might not be the best choice as the benchmark of evaluating rPPG algorithms for more complex practical applications. Consequently, a new publicly available database, directly related to rPPG-suitable practical applications, is an urgent need.

C. Multimodel Fusion

Many studies have demonstrated that HR can be recovered by using ordinary RGB cameras even during relatively dark illumination situations, but it would be useless during totally dark conditions. In order to monitor HR uninterruptedly, a thermal/infrared camera, combined with RGB cameras and possibly also other cameras insensitive to dark illuminance, will be an appropriate approach for robust and continuous noncontact HR measurement. The feasibility has been demonstrated in [43], [83], and [117].

It has been pointed out that HR can also be estimated based on motion-induced changes. These changes are caused by the cyclical movement of blood from heart to head via the carotid arteries giving rise to periodic head motion at the cardiac frequency. These cardiac-synchronous changes in the ambient light can also be remotely detected from the facial videos and it is called remote BCG (rBCG) [23], [61]. By this means, the only rBCG or the combination of rPPG and rBCG will be another prospect [117], [118].

D. Multipeople, Multiview, and Multicamera Monitoring

In realistic applications, when a camera is installed in a room, more than one person can be captured by the camera. Besides, apart from the frontal face, other views of the face (even the disappearance of the face) will appear, which bring challenges to existing rPPG methods. Poh *et al*. [9] have already demonstrated that their proposed method of remotely measuring HR can be easily scalable for simultaneous assessment of multiple people in front of the camera. Al-Naji and Chahl [119] proposed to simultaneously estimate HR from multiple people using noise artifact removal techniques. It is encouraging that remote HR measurement for multipeople is feasible and can be further promoted with the development of multiface detection and tracking techniques [120]. As for the multiview problem, most of the dominant face alignment algorithms employed in the rPPG field can only handle the frontal face within a sight deviation. Although the algorithm developed by Asthana *et al*. [93] and recently employed by Qi *et al*. [75] can provide a more unconstrained strategy for HR estimation, it is not good enough yet. More advanced face alignment algorithms aiming to provide a free face-view of the subject should be developed and introduced [121], [122]. Furthermore, in order to realize a space-seamless rPPG-based HR measurement, a single camera may no longer meet the need since the face of the subject may be turned away from the camera or be obscured by other objects, resulting in missing observations [99], [123]. A preliminary research fusing partial color-channel signals from an array of cameras has been conducted to enable physiology measurements from moving subjects [124]. In the future, other fusion mechanisms or advanced signal processing methods with respect to the optimal ROI selection or HR component extraction can be developed when using multiple cameras.

E. Multiple Parameters Evaluated in Multiple Applications

This review paper mainly concentrates on HR measurement by using rPPG technique. Apart from HR, several other physiological parameters related to health status can also be measured by rPPG. For instance, HRV and RR [125]–[129], blood oxygen saturation [5], [130], blood perfusion, pulse

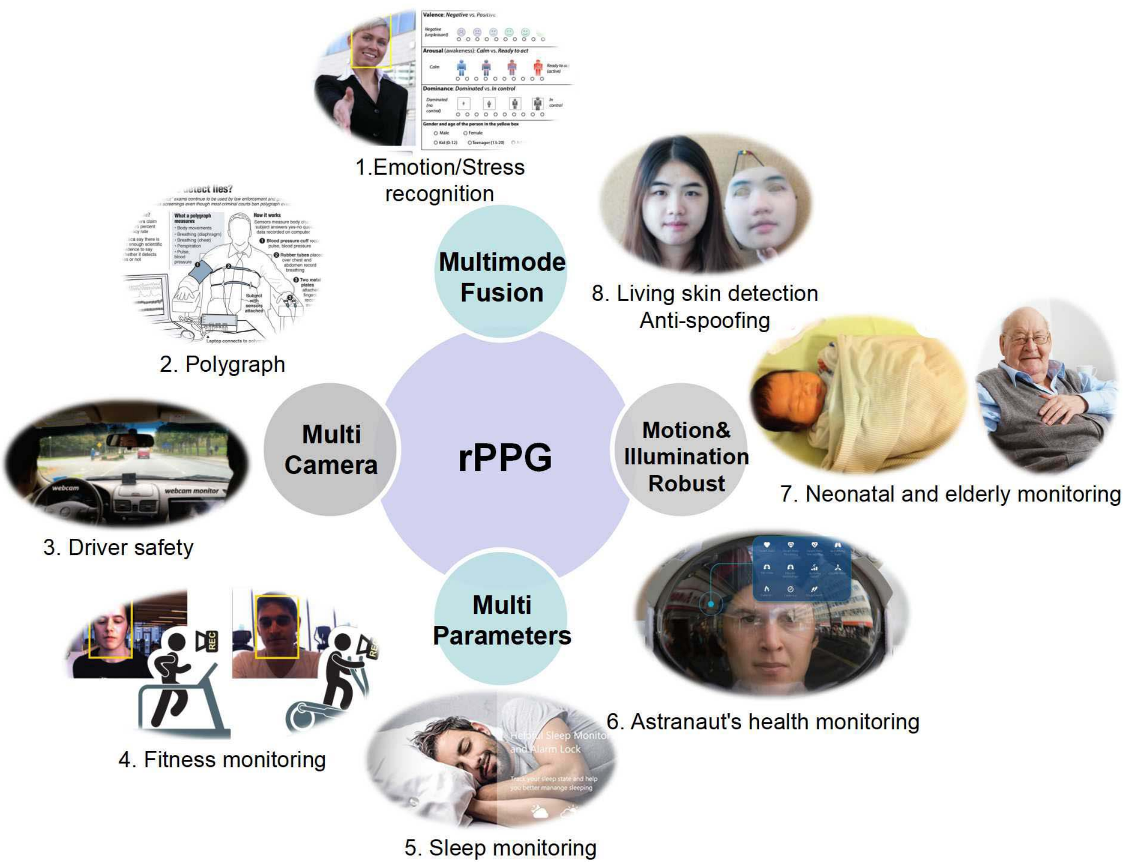

Fig. 6. Potential applications of rPPG techniques.[6]

transit time/pulse wave velocity [131], blood pressure [4], as well as systolic and diastolic peaks can also be measured by using rPPG [132]. The detailed information can be found in [8]. However, the related methods have not been evaluated during rigorous situations. Thereby, the robust measurement of these parameters under more challenging conditions is another important direction.

Since rPPG overcomes the disadvantages related to fragile skin injury or infection when using contact HR sensors, such noncontact HR monitoring technique has been demonstrated feasible and appealing to many potential applications, as illustrated in Fig. 6. For instance, it can provide a comfortable way for monitoring of infants, elderly and chronic-patients at home, in ICU or remote healthcare situations [26], [43]. Since telling a lie involves the activation of the autonomic nervous system (ANS), which leads to the changes of mental stress or physiological parameters, the rPPG can be extended as a polygraph while someone is questioned. In addition, emotional states are also important to indicate the healthy status of individuals. In particular, fatigue and negative emotions (such as irritation that may make drivers more aggressive and less attentive) of the driver are risk factors for driving safety. Consequently, monitoring the fatigue, the engagement and the emotional states of individuals by rPPG is a great potential prospect [133]–[135]. In addition, rPPG-based noncontact physiological parameter measurement will provide an efficient way, or/and combined with facial expressions, to remotely aware emotions [136]. As for fitness applications, it is important to retrieve the health status of the exerciser and an optimized training program can be customized according to the changing physiological parameters. The rPPG, instead of the conventional contact handhold or thoracic-band HR electrodes, will be more attractive. As for time-seamless HR monitoring, the combination of visible RGB cameras and infrared cameras will be promising. The rPPG based on infrared cameras will be particularly appropriate for sleep monitoring during the night.

Furthermore, a recent study, which demonstrated that HR and RR can be well derived from video sequences captured by a hovering UAV by using a combination of CEEMDAN and CCA, suggests potential applications of detecting security threats or deepening the context of human–machine interactions. More recently, researchers have proposed to classify living skin by using the rPPG technique based on the idea of transforming the time-variant rPPG-signals into signal shape descriptors (called multiresolution iterative spectrum) [137], [138]. This breakthrough in the rPPG technique can be employed as a biometric authentication tool, i.e., to prevent the adversary that may pretend to be a trusted device by generating

[6]Figures in application 1 to 8 are sequentially from [110], [World Book Science and Invention, Encyclopedia American Polygraph Association, federal polygraphers] [106], [111], https://www.amazon.cn/dp/B06XWTYMZV, http://www.softwaretestingnews.co.uk [43], and https://www.telecare24.co.uk [112].

a similar ID without physical contact and thus bypassing one of the core security conditions [139].

Apart from the above-mentioned prospects of rPPG, there exists one crucial problem that has not yet been addressed much. Currently, most of the existing rPPG methods are effective based on uncompressed video data. However, the uncompressed videos will occupy a large amount of storage space, which causes difficulty to the sharing of the data online. In addition, since the required data transfer rate of the uncompressed video data largely exceeds the transmission capability of current telecommunication technology, it is hardly possible to apply rPPG methods to the cases that need telecommunication. [140]. Previous studies have demonstrated that the performances of applying some rPPG methods to lossless compressed videos are close to those corresponding to uncompressed ones but with much less storage space (about 45% with FFV1 codec) [146]–[148]. In the future, developing more robust rPPG methods that suit for compressed video data is also a prospect.

VI. Conclusion

rPPG has been attracting increasing attention in the literature. This paper provides a comprehensive review of this promising technique, with a particular focus on recent contributions to overcome challenges in the presence of illumination variations and motion artifacts. A general scheme for measuring HR under either condition was illustrated, and dominating methods for each condition were summarized, compared, and discussed to reveal their principles, pros, and cons. Among all such methods, those employed for eliminating motion-induced artifacts were then classified into four subcategories, namely, BSS-based, model-based, motion compensated, and others. Finally, certain future prospects of rPPG were proposed, including: 1) the design of advanced methods with prior information; 2) establishing a public database benchmark; and 3) realizing a continuous, robust and space seamless HR measurement using different strategies. We believe that this paper can provide the researchers a more complete and comprehensive understanding of rPPG, facilitate further development of rPPG, and inspire numerous potential applications in healthcare.

References

[1] C. Brüser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and unobtrusive cardiorespiratory monitoring techniques," *IEEE Rev. Biomed. Eng.*, vol. 8, pp. 30–43, Aug. 2015.

[2] L. Iozzia, L. Cerina, and L. Mainardi, "Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA," *Physiol. Meas.*, vol. 37, no. 11, p. 1934, 2016.

[3] A. P. Prathosh, P. Praveena, L. K. Mestha, and S. Bharadwaj, "Estimation of respiratory pattern from video using selective ensemble aggregation," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2902–2916, Jun. 2017.

[4] 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.

[5] L. Kong *et al.*, "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light," *Opt. Express*, vol. 21, no. 15, pp. 17464–17471, 2013.

[6] U. Bal, "Non-contact estimation of heart rate and oxygen saturation using ambient light," *Biomed. Opt. Express*, vol. 6, no. 1, pp. 86–97, 2015.

[7] P. K. Jain and A. K. Tiwari, "Heart monitoring systems—A review," *Comput. Biol. Med.*, vol. 54, pp. 1–13, Nov. 2014.

[8] 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.

[9] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[10] Y. Yan, X. Ma, L. Yao, and J. Ouyang, "Noncontact measurement of heart rate using facial video illuminated under natural light and signal weighted analysis," *Biomed. Mater. Eng.*, vol. 26, no. s1, pp. 903–909, 2015.

[11] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15776–15790, 2017.

[12] A. D. Kaplan, J. A. OrSullivan, E. J. Sirevaag, P.-H. Lai, and J. W. Rohrbaugh, "Hidden state models for noncontact measurements of the carotid pulse using a laser Doppler vibrometer," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 744–753, Mar. 2012.

[13] W. Hu, Z. Zhao, Y. Wang, H. Zhang, and F. Lin, "Noncontact accurate measurement of cardiopulmonary activity using a compact quadrature Doppler radar sensor," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 725–735, Mar. 2014.

[14] A. E. Mahdi and L. Faggion, "Non-contact biopotential sensor for remote human detection," *J. Phys., Conf. Ser.*, vol. 307, no. 1, p. 012056, 2011.

[15] J. Kranjec, S. Beguš, J. Drnovšek, and G. Geršak, "Novel methods for noncontact heart rate measurement: A feasibility study," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 838–847, Apr. 2014.

[16] S. Fallet, Y. Schoenenberger, L. Martin, F. Braun, V. Moser, and J.-M. Vesin, "Imaging photoplethysmography: A real-time signal quality index," *Computing*, vol. 44, Sep. 2017, pp. 1–4.

[17] Q. Fan and K. Li, "Non-contact remote estimation of cardiovascular parameters," *Biomed. Signal Process. Control*, vol. 40, pp. 192–203, Feb. 2018.

[18] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, 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] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6398–6404.

[21] 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.

[22] B. Wei, X. He, C. Zhang, and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions," *Biomed. Eng. Online*, vol. 16, no. 1, p. 17, 2017.

[23] P. V. Rouast, M. T. P. Adam, R. Chiong, D. Cornforth, and E. Lux, "Remote heart rate measurement using low-cost RGB face video: A technical literature review," *Frontiers Comput. Sci.*, vol. 12, no. 5, pp. 858–872, 2016.

[24] R. Amelard *et al.*, "Feasibility of long-distance heart rate monitoring using transmittance photoplethysmographic imaging (PPGI)," *Sci. Rep.*, vol. 5, Oct. 2015, Art. no. 14637.

[25] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 40–48, May/Jun. 2016.

[26] L. A. M. Aarts *et al.*, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study," *Early Hum. Develop.*, vol. 89, no. 12, pp. 943–948, 2013.

[27] Z. Guo, Z. J. Wang, and Z. Shen, "Physiological parameter monitoring of drivers based on video data and independent vector analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4374–4378.

[28] S. Rasche *et al.*, "Camera-based photoplethysmography in critical care patients," *Clin. Hemorheol. Microcirculation*, vol. 64, no. 1, pp. 77–90, 2016.

[29] F. Zhao, M. Li, Z. Jiang, J. Z. Tsien, and Z. Lu, "Camera-based, non-contact, vital-signs monitoring technology may provide a way for the early prevention of SIDS in infants," *Frontiers Neurol.*, vol. 7, Dec. 2016, Art. no. 236.

[30] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, and F. Meriaudeau, "Novel health monitoring method using an RGB camera," *Biomed. Opt. Express*, vol. 8, no. 11, pp. 4838–4854, 2017.

[31] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[32] 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.

[33] P. Sahindrakar, "Improving motion robustness of contact-less monitoring of heart rate using video analysis," Ph.D. dissertation, Eindhoven Univ. Technol., Eindhoven, The Netherlands, Aug. 2011.

[34] Y. Sun, V. Azorin-Peris, R. Kalawsky, S. Hu, C. Papin, and S. E. Greenwald, "Use of ambient light in remote photoplethysmographic systems: Comparison between a high-performance camera and a low-cost webcam," *J. Biomed. Opt.*, vol. 17, no. 3, p. 037005, 2012.

[35] J. Przybyło, E. Kańtoch, M. Jabłoński, and P. Augustyniak, "Distant measurement of plethysmographic signal in various lighting conditions using configurable frame-rate camera," *Metrol. Meas. Syst.*, vol. 23, no. 4, pp. 579–592, 2016.

[36] S. A. Siddiqui, Y. Zhang, Z. Feng, and A. Kos, "A pulse rate estimation algorithm using PPG and smartphone camera," *J. Med. Syst.*, vol. 40, no. 5, p. 126, 2016.

[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] A. Al-Naji and J. Chahl, "Simultaneous tracking of cardiorespiratory signals for multiple persons using a machine vision system with noise artifact removal," *IEEE J. Transl. Eng. Health Med.*, vol. 5, 2017, Art. no. 1900510.

[39] A. Sikdar, S. K. Behera, and D. P. Dogra, "Computer-vision-guided human pulse rate estimation: A review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 91–105, 2016.

[40] M. Hassan *et al.*, "Heart rate estimation using facial video: A review," *Biomed. Signal Process. Control*, vol. 38, pp. 346–360, Sep. 2017.

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

[42] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, no. 6, p. 1023, 2017.

[43] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PLoS ONE*, vol. 8, no. 10, p. e71384, 2013.

[44] P. S. Addison, D. Jacquel, D. M. H. Foo, and U. R. Borg, "Video-based heart rate monitoring across a range of skin pigmentations during an acute hypoxic challenge," *J. Clin. Monitor. Comput.*, vol. 32, no. 5, pp. 871–880, 2017.

[45] M. Lewandowska, J. Ruminski, 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.

[46] A. Al-Naji, A. G. Perera, and J. Chahl, "Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle," *Biomed. Eng. Online*, vol. 16, Aug. 2017, Art. no. 101.

[47] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, 2012, Art. no. 65.

[48] J. Rumiński, "Reliability of pulse measurements in videoplethysmography," *Metrol. Meas. Syst.*, vol. 23, no. 3, pp. 359–371, 2016.

[49] Y. Yang, C. Liu, H. Yu, D. Shao, F. Tsow, and N. Tao, "Motion robust remote photoplethysmography in CIELab color space," *J. Biomed. Opt.*, vol. 21, no. 11, p. 117001, 2016.

[50] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. 1.

[51] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, San Diego, CA, USA, Aug. 2012, pp. 2174–2177.

[52] Y.-P. Yu, P. Raveendran, C.-L. Lim, and B.-H. Kwan, "Dynamic heart rate estimation using principal component analysis," *Biomed. Opt. Express*, vol. 6, no. 11, pp. 4610–4618, 2015.

[53] B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, "Signal recovery in imaging photoplethysmography," *Physiol. Meas.*, vol. 34, no. 11, pp. 1499–1511, 2013.

[54] K. H. Suh and E. C. Lee, "Contactless physiological signals extraction based on skin color magnification," *J. Electron. Imag.*, vol. 26, no. 6, p. 063003, 2017.

[55] A. Sarkar, Z. Doerzaph, and A. L. Abbott, "Video magnification to detect heart rate for drivers," Nat. Surf. Transp. Saf. Center Excellence and Virginia Tech Transp. Inst., Blacksburg, VA, USA, Tech. Rep. 17-UT-058, 2017.

[56] C. J. Dorn *et al.*, "Automated extraction of mode shapes using motion magnified video and blind source separation," in *Topics in Modal Analysis & Testing*, vol. 10. Cham, Switzerland: Springer, 2016, pp. 355–360.

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

[58] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Face-based heart rate signal decomposition and evaluation using multiple linear regression," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1351–1360, Mar. 2016.

[59] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3640–3648.

[60] X. Yu, J. Huang, S. Zhang, W. Yan, and D. N. Metaxas, "Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1944–1951.

[61] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2013, pp. 3430–3437.

[62] D. Lee, J. Kim, S. Kwon, and K. Park, "Heart rate estimation from facial photoplethysmography during dynamic illuminance changes," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2758–2761.

[63] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, pp. 807–831, Mar. 2014.

[64] 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, Feb. 2017.

[65] V. Jeanne, M. Asselman, B. den Brinker, and M. Bulut, "Camera-based heart rate monitoring in highly dynamic light conditions," in *Proc. Int. Conf. Connected Vehicles Expo (ICCVE)*, Dec. 2013, pp. 798–799.

[66] S. Xu, L. Sun, and G. K. Rohde, "Robust efficient estimation of heart rate pulse from video," *Biomed. Opt. Express*, vol. 5, no. 4, pp. 1124–1135, 2014.

[67] D.-Y. Chen *et al.*, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," *IEEE Sensors J.*, vol. 15, no. 1, pp. 618–627, Jan. 2015.

[68] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward, "Removing muscle artifacts from EEG data: Multichannel or single-channel techniques?" *IEEE Sensors J.*, vol. 16, no. 7, pp. 1986–1997, Apr. 2016.

[69] X. Chen, X. Xu, A. Liu, M. J. Mckeown, and Z. J. Wang, "The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 359–370, Feb. 2018.

[70] J. Jenitta and A. Rajeswari, "Denoising of ECG signal based on improved adaptive filter with EMD and EEMD," in *Proc. IEEE Conf. Inf. Commun. Technol. (ICT)*, Apr. 2013, pp. 957–962.

[71] X. Chen, Q. Chen, Y. Zhang, and Z. J. Wang, "A novel EEMD-CCA approach to removing muscle artifacts for pervasive EEG," *IEEE Sensors J.*, to be published, doi: 10.1109/JSEN.2018.2872623.

[72] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Príncipe, "Kernel risk-sensitive loss: Definition, properties and application to robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2888–2901, Jun. 2017.

[73] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[74] S. Yu, S. Hu, V. Azorin-Peris, J. A. Chambers, Y. Zhu, and S. E. Greenwald, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, p. 077010, 2011.

[75] 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.

[76] 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.

[77] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiol. Meas.*, vol. 35, no. 9, pp. 1913–1926, 2014.

[78] 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.

[79] G. Cennini, J. Arguel, K. Akşit, and A. van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Express*, vol. 18, no. 5, pp. 4867–4875, 2010.

[80] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 568–574, 2013.

[81] A. V. Moço, S. Stuijk, and G. de Haan, "Motion robust PPG-imaging through color channel mapping," *Biomed. Opt. Express*, vol. 7, no. 5, pp. 1737–1754, 2016.

[82] W. J. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Amplitude-selective filtering for remote-PPG," *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1965–1980, 2017.

[83] M. V. Gastel, S. Stuijk, and G. D. Haan, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.

[84] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[85] X. Chen, Z. J. Wang, and M. McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Process. Mag.*, vol. 33, no. 3, pp. 86–107, May 2016.

[86] 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 Informat.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014.

[87] Y. C. Lin, N. K. Chou, G. Y. Lin, M. H. Li, and Y. H. Lin, "A real-time contactless pulse rate and motion status monitoring system based on complexion tracking," *Sensors*, vol. 17, no. 7, p. 1490, 2017.

[88] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in EEG data," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1770–1779, Jul. 2017.

[89] R.-Y. Huang and L.-R. Dung, "A motion-robust contactless photoplethysmography using chrominance and adaptive filtering," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Oct. 2015, pp. 1–4.

[90] 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.

[91] A. A. Kamshilin, V. V. Zaytsev, and O. V. Mamontov, "Novel contactless approach for assessment of venous occlusion plethysmography by video recordings at the green illumination," *Sci. Rep.*, vol. 7, Mar. 2017, Art. no. 464.

[92] A. V. Moço, S. Stuijk, and G. de Haan, "Skin inhomogeneity as a source of error in remote PPG-imaging," *Biomed. Opt. Express*, vol. 7, no. 11, pp. 4718–4733, 2016.

[93] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3444–3451.

[94] 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. (CVPR)*, Jun. 2016, pp. 2396–2404.

[95] S. Zheng, P. Sturgess, and P. H. S. Torr, "Approximate structured output learning for constrained local models with application to real-time facial feature detection and tracking on low-power devices," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Apr. 2013, pp. 1–8.

[96] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1859–1866.

[97] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 702–715.

[98] D. Rathod, A. Vinay, S. S. Shylaja, and S. Natarajan, "Facial landmark localization—A literature survey," *Int. J. Current Eng. Technol.*, vol. 4, no. 3, pp. 1901–1907, 2014.

[99] O. Gupta, D. McDuff, and R. Raskar, "Real-time physiological measurement and visualization using a synchronized multi-camera system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 46–53.

[100] S. Kwon, J. Kim, D. Lee, and K. Park, "ROI analysis for remote photoplethysmography on facial video," in *Proc. 37th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 4938–4941.

[101] R.-C. Peng, W.-R. Yan, N.-L. Zhang, W.-H. Lin, X.-L. Zhou, and Y.-T. Zhang, "Investigation of five algorithms for selection of the optimal region of interest in smartphone photoplethysmography," *J. Sensors*, vol. 2016, Nov. 2016, Art. no. 6830152.

[102] F. Bousefsaf, C. Maaoui, and A. Pruski, "Automatic selection of webcam photoplethysmographic pixels based on lightness criteria," *J. Med. Biol. Eng.*, vol. 37, no. 3, pp. 374–385, 2017.

[103] D. Wedekind *et al.*, "Assessment of blind source separation techniques for video-based cardiac pulse extraction," *J. Biomed. Opt.*, vol. 22, no. 3, p. 035002, 2017.

[104] 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.

[105] M. Villarroel *et al.*, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare Technol. Lett.*, vol. 1, no. 3, pp. 87–91, Sep. 2014.

[106] H. Qi, Z. J. Wang, and C. Miao, "Non-contact driver cardiac physiological monitoring using video data," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2015, pp. 418–422.

[107] Q. Zhang, Q. Wu, Y. Zhou, X. Wu, Y. Ou, and H. Zhou, "Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers," *Measurement*, vol. 100, pp. 311–321, Mar. 2017.

[108] F. Bousefsaf, C. Maaoui, and A. Pruski, "Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam," *Comput. Biol. Med.*, vol. 53, pp. 154–163, Oct. 2014.

[109] M. S. Pedersen, U. Kjems, K. B. Rasmussen, and L. K. Hansen, "Semi-blind source separation using head-related transfer functions [speech signal separation]," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, May 2004, p. V–713.

[110] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Emotion recognition in context," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1960–1968.

[111] W. Wang, B. Balmaekers, and G. de Haan, "Quality metric for camera-based pulse rate monitoring in fitness exercise," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2430–2434.

[112] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao, "3D mask face anti-spoofing with remote photoplethysmography," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 85–100.

[113] G. Heusch, A. Anjos, and S. Marcel. (2017). "A reproducible study on remote heart rate measurement." [Online]. Available: https://arxiv.org/abs/1709.00962

[114] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[115] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Oct./Mar. 2012.

[116] M. N. H. Mohd, M. Kashima, K. Sato, and M. Watanabe, "Facial visual-infrared stereo vision fusion measurement as an alternative for physiological measurement," *J. Biomed. Image Process.*, vol. 1, no. 1, pp. 34–44, 2014.

[117] C. H. Antink, H. Gao, C. Brüser, and S. Leonhardt, "Beat-to-beat heart rate estimation fusing multimodal video and sensor data," *Biomed. Opt. Express*, vol. 6, no. 8, pp. 2895–2907, 2015.

[118] D. Shao, F. Tsow, C. Liu, Y. Yang, and N. Tao, "Simultaneous monitoring of ballistocardiogram and photoplethysmogram using a camera," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1003–1010, May 2017.

[119] A. Al-Naji and J. Chahl, "Simultaneous tracking of cardiorespiratory signals for multiple persons using a machine vision system with noise artifact removal," *IEEE J. Transl. Eng. Health Med.*, vol. 5, 2017, Art. no. 1900510.

[120] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2017.2781233.

[121] S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, 2015, pp. 643–650.

[122] Y. Wang, Y. Liu, L. Tao, and G. Xu, "Real-time multi-view face detection and pose estimation in video stream," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, Aug. 2006, pp. 354–357.

[123] J. R. Estepp, E. B. Blackford, and C. M. Meier, "Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography," in *Proc. IEEE Int. Conf. Syst., Man (SMC)*, Oct. 2014, pp. 1462–1469.

[124] D. J. McDuff, E. B. Blackford, and J. R. Estepp, "Fusing partial camera signals for noncontact pulse rate variability measurement," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1725–1739, Aug. 2017.

[125] 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.

[126] M. van Gastel, S. Stuijk, and G. de Haan, "Robust respiration detection from remote photoplethysmography," *Biomed. Opt. Express*, vol. 7, no. 12, pp. 4941–4957, 2016.

[127] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, Sep. 2014.

[128] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *Biomed. Eng. Online*, vol. 15, no. 1, p. 11, 2016.

[129] K. Y. Lin, D. Y. Chen, and W. J. Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors J.*, vol. 16, no. 9, pp. 3263–3271, May 2016.

[130] A. R. Guazzi *et al.*, "Non-contact measurement of oxygen saturation with an RGB camera," *Biomed. Opt. Express*, vol. 6, no. 9, pp. 3320–3338, Sep. 2015.

[131] S. Dangdang, Y. Yuting, L. Chenbin, T. Francis, Y. Hui, and T. Nongjian, "Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 11, pp. 2760–2767, Nov. 2014.

[132] D. McDuff, S. Gontarek, and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2948–2954, Dec. 2014.

[133] C. Maaoui, F. Bousefsaf, and A. Pruski, "Automatic human stress detection based on webcam photoplethysmographic signals," *J. Mech. Med. Biol.*, vol. 16, no. 4, p. 1650039, 2016.

[134] C. R. Madan, T. Harrison, and K. E. Mathewson, "Noncontact measurement of emotional and physiological changes in heart rate from a webcam," *Psychophysiology*, vol. 55, no. 4, p. e13005, 2018.

[135] P. V. Rouast, M. T. P. Adam, D. J. Cornforth, E. Lux, and C. Weinhardt, "Using contactless heart rate measurements for real-time assessment of affective states," in *Information Systems and Neuroscience*. Cham, Switzerland: Springer, 2017, pp. 157–163.

[136] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affective Comput.*, vol. 8, no. 1, pp. 15–28, Jan./Mar. 2017.

[137] 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.

[138] 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.

[139] R. M. Seepers, W. Wang, G. de Haan, I. Sourdis, and C. Strydis, "Attacks on heartbeat-based security using remote photoplethysmography," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 714–721, May 2018.

[140] C. Zhao, C.-L. Lin, W. Chen, and Z. Li, "A novel framework for remote photoplethysmography pulse extraction on compressed videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1299–1308.

[141] D. J. McDuff, E. B. Blackford, and J. R. Estepp, "The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Jun. 2017, pp. 63–70.

[142] L. Cerina, L. Iozzia, and L. Mainardi, "Influence of acquisition frame-rate and video compression techniques on pulse-rate variability estimation from vPPG signal," *Biomed. Eng./Biomedizinische Technik*, to be published, doi: 10.1515/bmt-2016-0234.

[143] E. B. Blackford and J. R. Estepp, "Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography," *Proc. SPIE*, vol. 9417, p. 94172D, Mar. 2015.