Combining Photoplethysmography and Ballistocardiography to Address Voluntary Head Movements in Heart Rate Monitoring

JEAN-PIERRE LOMALIZA[1], HANHOON PARK[1], AND MARK BILLINGHURST[2]
[1]Department of Electronic Engineering, Pukyong National University, Busan 48513, South Korea
[2]School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, SA 5001, Australia

Corresponding author: Hanhoon Park (hanhoon.park@pknu.ac.kr)

ABSTRACT Daily vital signs monitoring is very important for detecting diseases in early stages and for preventive treatments. Such a task can be achieved by taking advantage of the omnipresence of cameras in people's personal space. As heart-related diseases are part of the leading causes of deaths worldwide, monitoring heart-related vital signs appear to be very crucial. In this article we aim to provide a touchless approach and propose a robust method for estimating heart rate through analysis of face videos. In particular, we consider a challenging scenario, i.e., the user is on a video call and may often move his/her head. Existing touchless, vision-based methods use either photoplethysmography (PPG) or ballistocardiography (BCG). PPG methods exploit color changes in human skin during heartbeats caused by blood volume variations, but this is very sensitive to unstable lighting conditions. On the other hand, BCG methods exploit subtle head motions caused by Newtonian reaction to blood influx into the head at each heartbeat, thus being sensitive to a subject's voluntary head movements. Unlike conventional studies where either a PPG method or a BCG method is used, we propose to combine both to overcome the weakness faced by each method. We use BCG methods as the main approach due to their better accuracy on heart rate estimation, and PPG methods are used as the secondary backup to improve the accuracy in cases of large and frequent voluntary head movements. To this end, we introduce a dynamic voting system that effectively combines results of several variants of PPG and BCG methods. Experiments conducted on 20 healthy subjects with different skin tones in different lighting conditions show that our method has better accuracy compared to state-of-the-art methods, well addressing large voluntary head movements. Our method had a mean absolute error of 1.23 beats per minute (BPM) in the cases without voluntary head movements and 2.78 BPM in the cases with voluntary head movements.

INDEX TERMS Heart rate estimation, photoplethysmography, ballistocardiography, face video, voluntary head movement, hybrid method, dynamic voting, video call.

I. INTRODUCTION

Heart-related diseases are some of the most leading causes of deaths worldwide as reported by the American health association [1]. Therefore, studying heart-related vital signs has attracted the attention of many healthcare researchers working on healthcare applications through diverse fields. The most efficient way to reach as many potential users as possible is to exploit commonly and daily used devices

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Sanchez.

such as cameras. Fortunately, the visual information for the user obtained by cameras can be exploited to estimate some heart-related vital signs such as heart rate. There are two main types of approaches to estimate heart rate: touch-based approaches [2]–[6] and touchless approaches [7]–[10]. However, touchless approaches appear to be ideal for non-invasive heart rate monitoring where the user would not need any uncomfortable (or unsafe) contact with any sensor. Therefore, this article focuses on touchless approaches.

In the modern human society, online collaborative works through video calls are becoming increasingly common.

As one of the ideal scenarios is to measure heart rate while subjects are doing other activities without interfering, exploiting cameras to measure heart rate during video calls appear to be advantageous. Moreover, in most of video conference calls, the human face is the part of the body that is all the time visible on cameras. Therefore, we aim at developing touchless heart rate estimation methods that exploit visual information from human face for video conference call applications (Figure 1).
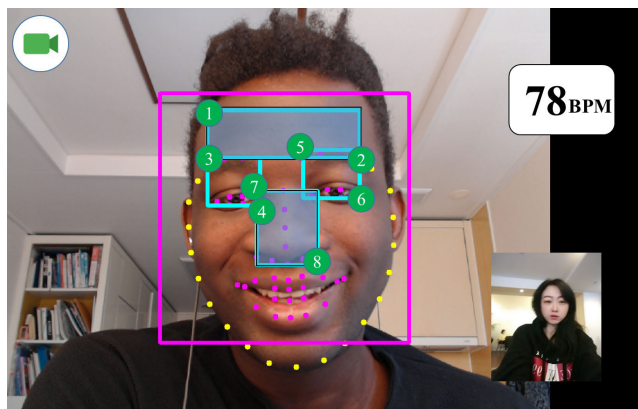

FIGURE 1. Interface of the prototype system. The heart rate is measured by combining PPG and BCG methods. Areas from eyes, forehead and nose are tracked using region-based optical flow tracking. Circled numbers represent tracked corners for the BCG signals. Blue rectangles (forehead and nose) are the regions used to extract the PPG signals.

Depending on the way physiological information is extracted, touchless approaches can be grouped into two main categories: photoplethysmography (PPG, sometimes called videoplethysmography) methods and ballistocardiography (BCG) methods. PPG methods such as [11]–[16] exploit color changes on the human skin during heartbeats. Such color changes are caused by a cyclic variation of blood volume on arteries. As the blood volume variations coincide with heartbeats rhythm, so heart rate information can be extracted accurately using such information. The main way for doing so is to track the human face throughout each frame and compute the average of pixel intensities within the selected regions of interest (ROI). The frequency of the variations of pixel intensities strongly correlates with the frequency associated with the heart rate. Hence, the heart rate can be estimated by analyzing such variations. On the other hand, BCG measures the recoil forces of the body when it reacts to cardiac ejection of blood into the vasculature. BCG methods such as [17]–[22] exploit subtle head motions caused by the recoil forces when blood cyclically travels from the heart to the head through the aorta and carotid artery during heartbeats. Those blood flows cause the head to move vertically in a cyclic way. The general way to get such head motions is to track feature points attached to the human face within some chosen ROIs throughout each frame. The time series of the vertical locations of those feature points are then considered as heartbeats related signals. Those signals strongly correlate with the heart rate since the tracked head motions are mostly dictated by heartbeats. Therefore, the heart rate can successfully be estimated by analyzing such signals.

The problem is that, although both the PPG and BCG methods work well in ideal situations such as stable lighting conditions or minimal voluntary head movements, real life scenarios mostly happen to be non-ideal. Some recent PPG or BCG methods [11], [19], [20] proposed mechanisms for reducing the noises caused by lighting changes and voluntary head movements (they will be explained in more detail in the next section). However, none of those methods showed good robustness to large head movements because they focused on handling small head movements caused by facial expression changes or breathing. In addition, since those methods are still one of the PPG and BCG methods, thus cannot be completely free from the inherent limitations. Thus, we attempt to combine both the PPG and BCG methods to achieve a better accuracy than using either method alone, proposing a dynamic voting system to combine output from each method. In addition, we clearly show that combining the PPG and BCG methods can effectively address issues with voluntary head movements in a video call scenario. In this regard, we propose a way to detect and remove such voluntary movements from signals to make our method robust and stable. Although previous attempts have been made to combine PPG and BCG signals [23]–[25], we focus on designing several variants of promising PPG and BCG methods and combining them. Therefore, our method gains additional consistency by cooperatively estimating the heart rate using the variants. Moreover, different from the existing combination methods, the main purpose of our study is to overcome challenges of real-life videos in a video call scenario. This will be further discussed in Section II-C.

In summary, our method concentrates in two key contributions:

1) **Voluntary movement detection/removal**: we propose a simple yet effective way to detect large head movements and remove portions of the signals related to those movements. This makes our method more robust to voluntary movements that may often occur in real-life scenarios like a video call.
2) **Combining several BCG and PPG variants through a voting system**: the combination allows our method to be more accurate and robust than the single use of the BCG and PPG methods.

II. RELATED WORK
As mentioned before, vision-based, touchless heart rate estimation methods can be classified into two categories: PPG methods and BCG methods.

A. PPG METHODS
The phenomenon of PPG was first introduced in the 1940s [26]. As explained in [27], blood absorbs light more than other surrounding tissues. Hence, the PPG approach

is based on simply analyzing cyclic variations of light reflectance caused by blood volume changes that occur during heartbeats.

In 2008, Verkruysse *et al.* [28] introduced the first practical remote PPG method, measuring PPG signals from subjects within one meter using a digital camera as input sensor. They generated raw PPG signals by averaging the pixel intensities of red, blue and green channels within the ROIs over time. They then applied a fourth order Butterworth bandpass filter (BPF) to refine signals and remove noises.

Poh *et al.* [29] proposed the first face-based PPG system for heart rate estimation. They used a simple webcam to extract PPG signals from subjects' facial regions. They detected the subject's face using the Viola-Jones face detector [30] and then used the whole face as the ROI. They extracted the mean of pixels intensities for each of red, green, and blue color channels. The time series of pixels intensities of each channel is used as raw signals which were then normalized and an independent component analysis (ICA) was applied for blind source separation. After the ICA decomposition, they selected the component that resembled a PPG signal the most and computed the heart rate from that.

Tasli *et al.* [31] proposed to detect and track facial landmarks using an improved version of the framework introduced in [32]. The purpose of using facial landmarks was to dynamically track designated ROIs despite facial expressions. To obtain the raw PPG signal, they proposed to average the intensity of pixels within the ROIs for the green channel in each frame. After obtaining the PPG signal, they used a detrending technique from [33] to remove noises caused by ambient lighting condition changes. After noise removal, they used both time domain information through peaks detection and frequency domain information through most dominant harmonic selection to estimate the heart rate.

Jain *et al.* [14] showed an additional video preprocessing step of cropping and stabilizing the video around human face. The cropped video is located in the facial area and has a static defined size of $M \times M$ pixels. After generating the cropped video, they only considered the red channel and constructed a matrix **A** where each column represents a video frame. Then, they used a principal component analysis (PCA) to obtain a reconstruction matrix **A'** and computed the reconstruction error matrix A_{PPG} from the difference between **A** and **A'**. Each column of A_{PPG} was reshaped back to the original frame size of $M \times M$ and the PPG signal was obtained by averaging the pixel intensity in each reconstructed frame. After applying a BPF with cut-off frequencies of 0.5 Hz and 5 Hz, they only considered portion of the signal having good quality, with significant visual features like peaks, foots and dicrotic notches. Finally, they computed the heart rate relative to the number of detected peaks within a signal duration.

Li *et al.* [11] introduced mechanisms to handle noises caused by illumination changes, rigid and non-rigid facial motions. To rectify noise caused by illumination changes, they used background regions (i.e., regions not belonging to the face) as references and proceeded to iteratively minimize the noise using a normalized least mean square so that they could subtract it from the signal. They handled rigid noise by correcting tracking facial feature points with a transformation matrix computed with respect to the feature point locations in consecutive frames. To handle the noises from non-rigid motions, they divided the signal into smaller parts of the same length, computed the standard deviation for each subpart, and removed subparts having standard deviations too different from others.

Finally, more recent studies are learning-based; they attempted to learn the mapping from facial videos to heart rate-related signals from a huge dataset using deep neural networks [34], [35].

B. BCG METHODS

Balakrishnan *et al.* [17] introduced the first BCG method accurate enough to be used for daily heart rate monitoring. To detect the head region, they used the Viola-Jones face detector that comes pre-installed with the OpenCV library [36]. The face detector returns a rough rectangular localization of the head region that does not always fit well to the real region on the image. Hence, in order to make sure that the facial region is selected, they used a ROI that is the middle 50% of the returned rectangle widthwise and 90% heightwise. In order to avoid the eye blinking artifacts, sub-rectangle spanning from 25% to 50% heightwise was removed from the ROI. After ROI selection, feature points within selected regions were tracked using the Lucas-Kanade optical flow algorithm. A time-series of the vertical locations of those feature points were considered as raw signals of head motion related to heartbeats. After signal extraction, a fifth order Butterworth filter with a passband of [0.75, 5] Hz was applied to remove noises from other motions. After signal filtering, the PCA was used to extract the most dominant Eigenvectors within those motion signals. Finally, the most periodic signal within the five most dominant components was selected as the one related to the heartbeats. To obtain the related heart rate, a fast Fourier transform (FFT) was applied to the selected signal to find the frequency having the highest magnitude in the frequency domain. The heart rate was then computed in beats per minute (BPM) by converting the frequency by multiplying it to 60. Shan and Yu [18] improved the previous BCG method by reducing ROI to only forehead regions in order to avoid signals noises caused by deformable motions from voluntary or involuntary facial expressions around cheeks or upper lip. In addition, they used ICA instead of PCA for the latent signal extraction because ICA had a more consistency and accuracy performance.

Haque *et al.* [19] used a combination of OpenCV's good features to track and facial landmarks detected with a supervised descent method (SDM) [37]. Using additional feature points detected with SDM, they could tackle the problem of feature points tracking loss occurring during facial expressions or other rigid motions. After obtaining raw signals using the feature point detection and tracking, they applied an eighth order Butterworth filter followed by a moving average

TABLE 1. PPG and BCG methods used in our method.

Method	Processing steps			
	Raw signal acquisition	Signal filtering	Heart rate-related signal extraction	Heart rate computation
M1 (BCG)	Tracking facial features	BPF [17] + MAF	Averaging raw signals	Constructing an optimization tree [5]
M2 (BCG)	Tracking facial features	BPF [17] + MAF	Averaging raw signals	Adaptive peak averaging [4]
M3 (BCG)	Tracking facial features	BPF [17] + MAF	Averaging raw signals	Finding the most dominant frequency [17], [19]
M4 (PPG)	Averaging intensities of three color channels	BPF [17] + MAF	Selecting the most periodic independent component [29]	Finding the most dominant frequency [17], [19]
M5 (PPG)	Averaging intensities of three color channels	BPF [17] + MAF	Selecting the most periodic principal component [14]	Adaptive peak averaging [4]

filter (MAF) in order to remove noises. Then, they computed the PCA followed by a discrete cosine transform to find the most periodic signal. Heart rate was then estimated similarly to [17].

Lee *et al.* [38] proposed a similar method to [17]. However, for ROI selection, they considered only forehead and nose regions. Moreover, instead of simply selecting the frequency with the most dominant magnitude as the heart rate, they proposed to train a K-means model. They did so by extracting trainable features from the most periodic PCA component. The ground truth data was obtained from an electrocardiograph.

Wang *et al.* [20] proposed a three-layer filtering method to obtain a better BCG signals than previous methods, filtering out the noises caused by facial expression, talking, and breathing.

Finally, Lomaliza and Park [39] introduced the first practical BCG system working on smartphones and addressed the hand motion artifacts caused by the non-static camera.

C. COMBINING PPG AND BCG METHODS
In recent years, several methods have attempted to combine PPG and BCG signals. However, as our method uses a single camera as input sensor to acquire both signals, we introduce only the methods using a single camera here. Antink *et al.* [23] proposed a way for fusing inputs from the video stream recorded with a camera output and an additional BCG sensor. They fused PPG and BCG signals acquired from the video stream by using a Bayesian approach. As they aimed to have a signal highly accurate in term of beat-to-beat average time estimation, their evaluation metric was the mean absolute interval estimation error. Such metric allows to measure beats displacements in second as compared to a ground truth device such as an electrocardiograph sensor. Although their full method having fusion with an additional BCG sensor had a better error rate of 24.4 ms of average displacement, their video-only variant also had a good error rate of 31.8 ms of average displacement. Shao *et al.* [24] proposed another method to simultaneously track PPG and BCG signals using a single camera. However, unlike [23], they processed those signals separately instead of fusing them. They aimed to prove the feasibility of simultaneously tracking PPG and BCG signals from a single camera. To evaluate the signal quality, they proposed to use signal-to-noise ratio by analyzing power spectrum frequencies in the frequency domain.

They later proposed to use a similar approach to track blood pressure and abnormal heart rhythm monitoring [25].

Note that the existing combination methods fused directly PPG and BCG signals or used BCG signals to obtain additional information such as stroke volume and blood pressure. They do neither consider nor fit our scenario, i.e. video calls with voluntary head movements. In contrast, we combine several variants of recent PPG and BCG methods to address the inherent weaknesses of PPG and BCG methods. Therefore, our method can support a wider range of applications in more challenging conditions.

III. OUR METHOD
As mentioned before, our method is a hybrid one, combining PPG and BCG methods. The PPG and BCG methods are variants of the conventional ones, listed in Table 1, because the designed variants are exclusive to each other and yields the best performance in our voting system used for combining PPG and BCG methods. Both types of PPG and BCG methods are usually divided into four steps: raw signal acquisition, signal filtering, heart rate-related signal extraction, heart rate computation. PPG and BCG are fundamentally different in the methods for signal acquisition, but they can use the same or similar methods in other steps. We created three BCG and two PPG variants by choosing different methods in each step. The variants are the same in most steps, which minimizes the computational burden of implementing multiple methods to estimate heart rate. Methods used in each variant and how to combine the variants to handle voluntary head motions will be explained in the following subsections.

A. RAW SIGNAL ACQUISITION
Using facial landmarks [37] that come with contribution modules of the OpenCV library, we created four ROIs as shown in Figure 1. These were used to obtain BCG raw signals while only forehead and nose ROIs are used to obtain PPG raw signals. To obtain the raw signals from ROIs, all the conventional PPG and BCG methods track feature points frame to frame using optical flow. However, such a technique introduces noise artifacts in the signal due to tracking loss of some feature points usually located on underexposed or overexposed regions. Therefore, instead of tracking feature points, we track regions using the median flow (MF) algorithm that comes with the OpenCV contribution modules. The region-based tracking approach offers better tracking
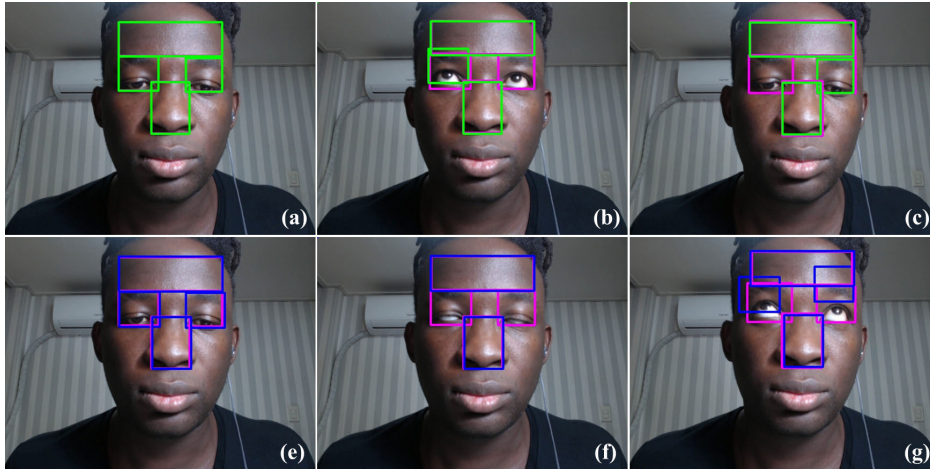
FIGURE 2. Sensitivity of different tracking methods to eye movements. In (a) to (c), pink and green rectangles represent the regions tracked by MF and KCF, respectively. In (e) to (g), pink and blue rectangles represent the regions tracked by MF and MOSSE, respectively. KCF and MOSSE often fail to track eye regions.

accuracy and robustness while reducing signal noises. As the MF algorithm tracks rectangular regions, we use top-left and bottom-right vertical locations of those rectangle to get the BCG raw signals.

Note that the OpenCV library contains several tracking algorithms, such as MIL, KCF, TLD, MOSSE, and CSRT [36]. So, we have fully tested whether each algorithm can be used for our purposes. In our preliminary experiments, MF appeared to be the best for our method. Table 2 shows the computation time in ms of each algorithm for each frame in an ordinary PC. Only three algorithms (KCF, MF, and MOSEE) had a speed fast enough to be used in real time. In addition, after conducting intensive experiments, we found out that KFC and MOSSE are very sensitive to eye movements while MF is robust (Fig. 2). As a result, MF achieves a good trade-off between accuracy, speed, and reliability.

TABLE 2. Per-frame computation time (in ms) of each tracking algorithm.

Boosting	MIL	KCF	TLD	MF	MOSSE	CSRT
126.45	168.35	8.72	263.04	16.07	3.59	122.25

For the PPG signals, we average intensities of the red, green, and blue channels of pixels within the forehead and nose regions.

B. SIGNAL FILTERING AND HEART RATE-RELATED SIGNAL EXTRACTION

Raw PPG and BCG signals still contain noises that may compromise the accuracy of heart rate estimation. Therefore, we apply signal processing techniques to remove the noise, but use different sequences of processing for each PPG and BCG method, as shown in Table 1.

For the PPG signals: At this stage, we have three PPG raw signals representing each of the red, green and blue color channels (Figure 3-(a)). We first apply a normalization with mean equal to 0 and variance equal to 1 (Figure 3-(b)). After that, we apply a fifth order Butterworth filter with a passband of [0.75, 2.5] Hz and a MAF with a window size of 5 to each signal (Figure 3-(c) and Figure 3-(d)). Then, we decompose the obtained signals using both PCA and ICA to end up with six signals in totals (Figure 3-(e)). For both decomposed principal and independent components, we select the most periodic ones that are expected to contain heartbeat-related information. Inspired by the approach in [17], we quantify the periodicity as the percentage of total spectral magnitude accounted for by the frequency with the most dominant power and its harmonic. Finally, we apply a MAF to the selected signals again.

For the BCG signals: Similarly to the PPG signals, we apply normalization and a MAF to the obtained raw signals. However, as the BCG signals describe the same head trajectory from different locations on the face, we found that simply averaging those signals is sufficient to capture the essential information. After averaging, we sequentially apply a Butterworth filter and a MAF similar to those applied to the PPG signal. The process flows of M1, M2, and M3 are represented in Figure 4.

For all the PPG and BCG methods in Table 1, we used both the Butterworth BPF [17] and MAF in the signal filtering step because recent methods, such as [19], [20], [38], saw a clear improvement by using several filters together.

C. VOLUNTARY HEAD MOVEMENT DETECTION

In realistic scenarios such as a conference video call, the user should be able to freely move his/her head. Such free and voluntary head movement may introduce artifacts to the PPG and BCG signals that would decrease the accuracy of heart rate estimation. The effect of such head movements is shown in Figure 5 where both the PPG and BCG signals are greatly distorted. For PPG signals, when the lighting conditions are

FIGURE 3. Process flows of the PPG methods (M4 and M5) in Table 1. M4 is represented by processes from (a) to (e1) and (f1). M5 is represented by processes from (a) to (e2) and (f2).
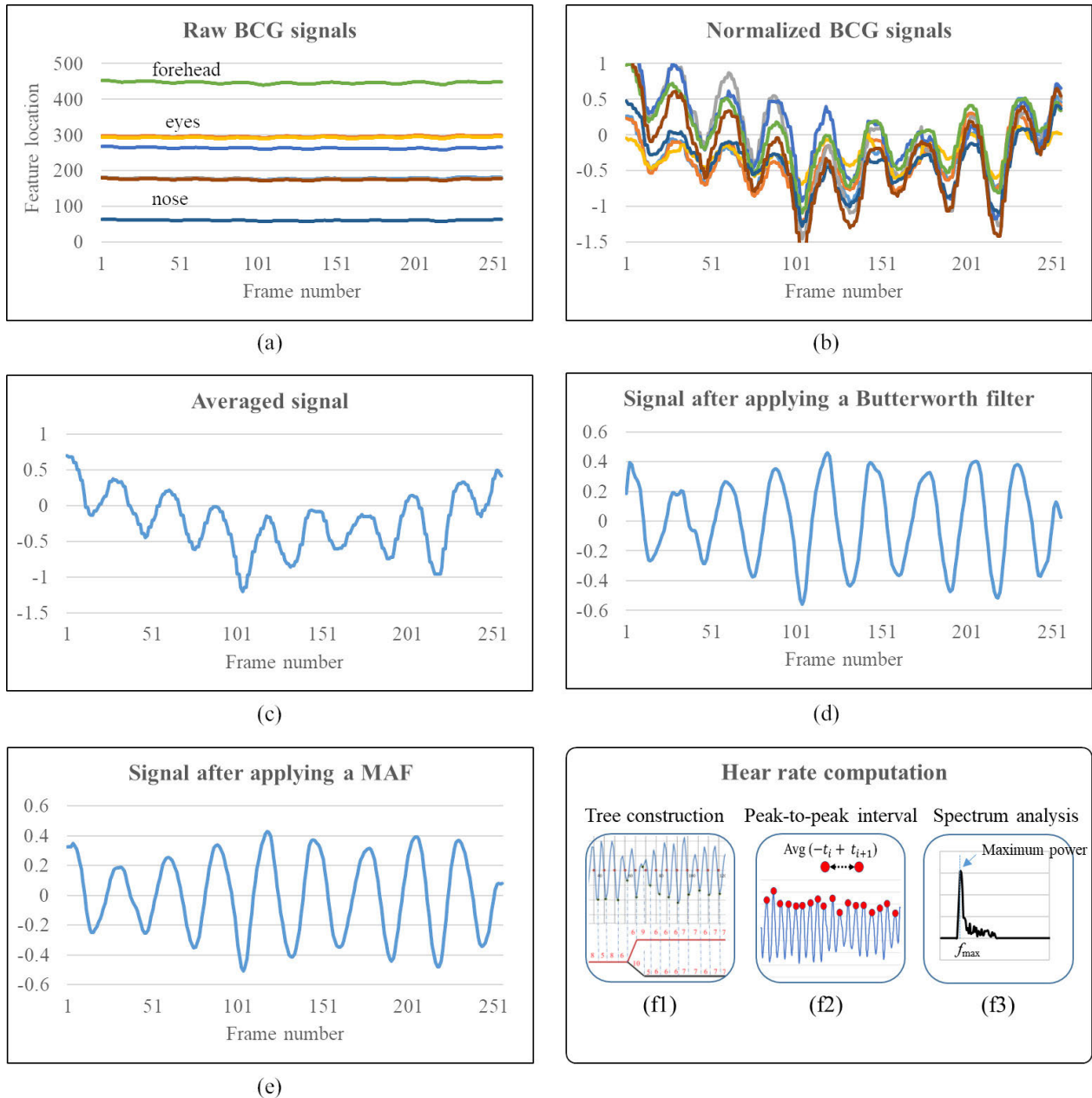
Raw BCG signals

(a)

Normalized BCG signals

(b)

Averaged signal

(c)

Signal after applying a Butterworth filter

(d)

Signal after applying a MAF

(e)

Hear rate computation

Tree construction   Peak-to-peak interval   Spectrum analysis

Avg $(-t_i + t_{i+1})$

Maximum power

f_{max}

(f1)   *(f2)*   *(f3)*

FIGURE 4. Process flows of the BCG methods (M1, M2, and M3) in Table 1. (f1) is the optimization tree [5], (f2) is the adaptive peak averaging [4], and (f3) is the method that finds the most dominant frequency in the power spectrum [17], [19]. M1 uses (a) to (f1), M2 uses (a) to (f2) and M3 uses (a) to (f3).

stable and diffused, voluntary head movements tend to have negligible impact on the extracted signals. However, in a realistic scenario where there might be several light sources with some being specular, voluntary head movements do influence the pixel intensities.

On the other hand, as BCG signals are also motion-based, voluntary head movements have a significant influence on the trajectories of tracked features. Moreover, heartbeat-related movements tend to be overshadowed by large movements within a relatively short period of time. Therefore, to handle such movements, we first introduce a simple way of detecting

them. The whole process is illustrated in Figure 6. First, we normalize and average BCG signals into one as shown in Figure 6-(b) and Figure 6-(c) respectively. Next, we apply a MAF with a large window, 25 in our case, to flatten the high frequency peaks as shown in Figure 6-(d). Then, we apply a curve approximation to convert the signal into simplistic line segments as shown in Figure 6-(e). The obtained line segments are then used to detect and remove portions corresponding to rapid head movements. Such portions are defined in this article as segments forming an angle greater than 45 degrees with the horizontal axis. Those portions are shown
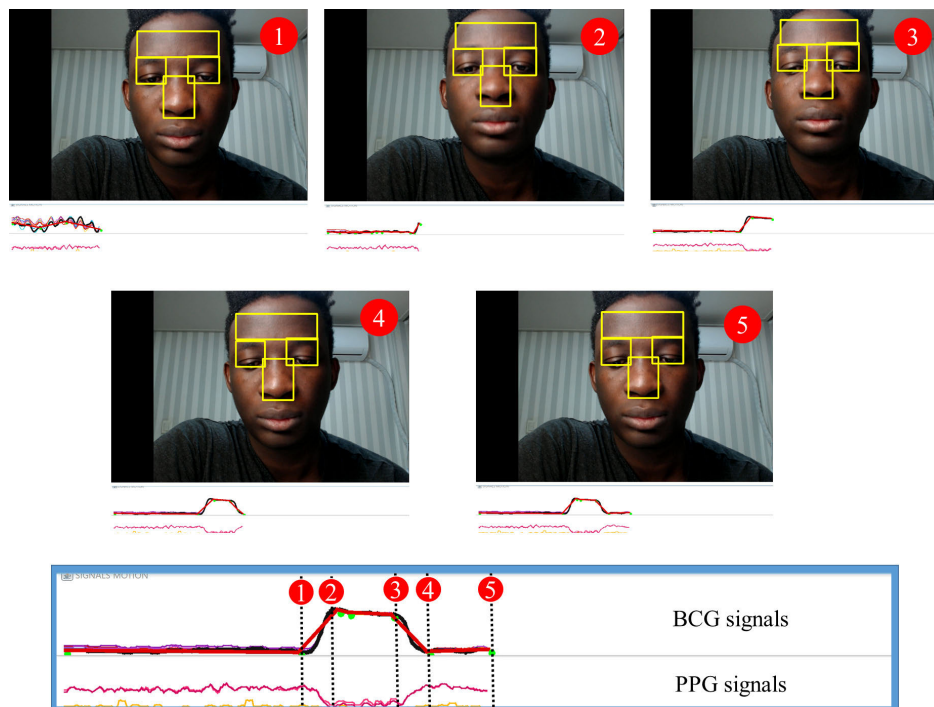
FIGURE 5. The impact of voluntary head movements on signals. Key frames with numbers correspond to timestamps on signals on the bottom. Voluntary head movements occurred in the segments between 1 and 2 and those between 3 and 4; the segments will be removed from the processed signals for both PPG and BCG methods since they form an angle greater than 45 degrees with the horizontal axis.

as red rectangles in Figure 6-(a) and red line segments in Figure 6-(e). Finally, we remove such portions for both PPG and BCG signals. The signal portions that are considered as not having rapid head movements are recombined to construct the final signals. It is also important to mention that, when such head movements happen, we conduct more tracking to reach the predefined signal length.

D. HEART RATE COMPUTATION WITH A DYNAMIC VOTING SYSTEM

Each PPM or BCG method computes heart rates from the heart-related signals using different methods, such as optimization tree construction [5], adaptive peak averaging [4], and most dominant frequency computation [17], [19], and then their resulting heart rates are combined together. Basically, the BCG methods are the main algorithms used in this article because of their better accuracy (this will be shown in Section IV-B). When there is no large voluntary head movement detected, we may only use those BCG methods (M1, M2, and M3). However, when some large movements are detected, we consider using PPG methods (M4 and M5) together. Nevertheless, the PPG methods are used only when the estimated heart rates from those methods are similar to those of the BCG methods. In our experiments, we used a threshold distance of 5 BPM for the difference between the PPG and BCG results. That means, to be used, a PPG method's estimation should be at most 5 BPM away from one of the BCG method's estimations. Finally, after selecting

the methods to be used, we use a dynamic voting system that takes advantage of all them. The idea of our system is to select a heart rate that falls within the range of all heart rates estimated by methods in Table 1. To do so, we create a table of heart rates ranging from 45 BPM to 150 BPM. At the initial step, each heart rate in the table is associated with 0 votes. Then, when a heart rate is estimated by one of the methods, we use a 1×5 mask to vote over heart rates. In our experiments, the mask [1 2 5 2 1] appeared to be the optimal to capture the best heart rate estimation. It is important to mention that the mask is centered around the estimated heart rate on the voting table. Finally, the heart rate value having the largest number of votes is determined to the final estimation. An illustration of the voting system is shown in Figure 7. In this example, all the five methods are used since head movements were detected. Masks were placed around estimations from each method and the heart rate related to 60 BPM received the largest number of votes.

IV. EXPERIMENTS AND ANALYSIS

Our method was implemented to run on desktop platform. An Apple MacBook Pro with Intel core i7 processor was used to run the experiments. The Logitech C922 webcam was used as input sensor with an image resolution of 1280×720 and the Polar chest belt H7 [40] was used as ground truth measuring device (The ground truth heart rates are obtained by averaging the measurements collected every 1 s). To test the accuracy of our method, we recorded videos
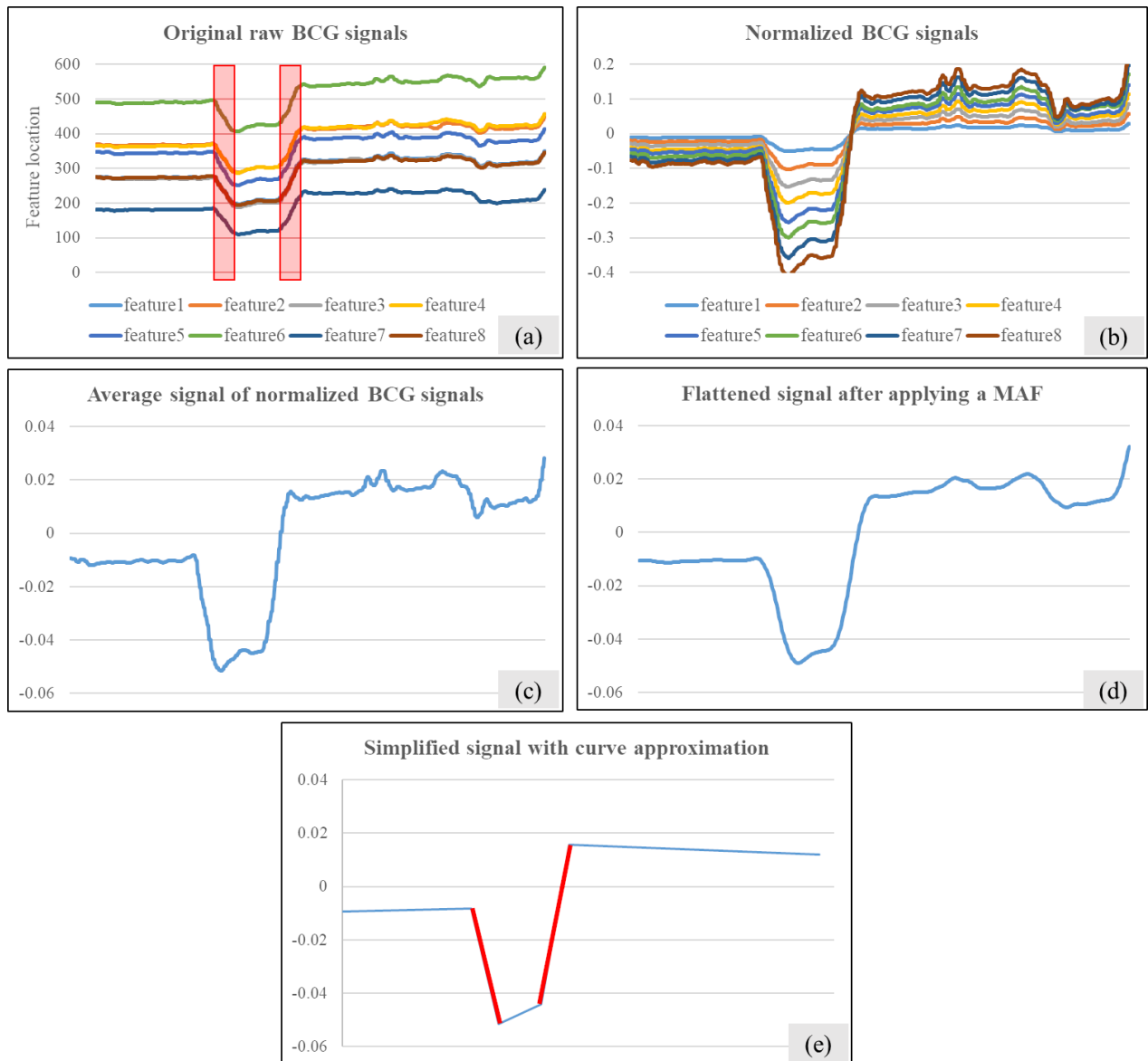
FIGURE 6. Our method for detecting voluntary head movements that are not related to heartbeats. (a) The raw BCG signals where voluntary head movements occurred in the portions marked with red rectangles, (b) BCG signals after normalization to a mean of 0 and a variance of 1, (c) averaged signal from normalized signals of (b), (d) flattened signal after applying a MAF with a large window, (e) curve approximation to divide the signal into line segments. Red segments in (e) represent areas having an angle bigger than 45 degrees with the horizontal axis and are removed from the raw signals.

of 20 subjects of different ethnicities and skin tones, and also different ages ranging from 25 to 40, with a duration of 24 s for each video and in JPEG2000 lossless video compressing format. The intended measurement length of our method is set to 20 s corresponding to 500 frames at 25 fps. However, the FFT implementation used in some steps of our method, such as most dominant frequency computation, requires the signal size to be a power of 2. Therefore, we chose to use 512 frames for each measurement. Subjects were asked to stay still without any movement while recording the first set of videos. After that, they were asked to behave like they were in a video call with natural movements while recording

the second set of videos. As some subjects were recorded a couple of times for each condition, both sets consisted of 25 videos, respectively. However, videos with head movements were recorded longer (thus, with a duration of 50 s) to make sure that, after removing the portion regarding head movements, we still have enough signal lengths to meet the required length of our method. Although all recordings were done in front of the laptop to which the camera was attached, different light conditions were considered. We recorded in daytime indoors, daytime outdoors, nighttime indoors, and nighttime outdoors. All subjects were asked to stay at most 1.2 meters away from the camera.

TABLE 3. Mean absolute errors (in BPM) of each method with or without voluntary head movements.

	[11]	[17]	[19]	[20]	[24]	[38]	Our method
Without movements	3.58	3.71	3.44	2.93	3.11	3.05	1.23
With movements	5.72	13.23	12.57	14.05	11.70	11.71	2.78
Average	4.65	8.47	8.05	8.49	7.40	7.38	2.01

TABLE 4. Pearson's correlation coefficients of each method with or without voluntary head movements.

	[11]	[17]	[19]	[20]	[24]	[38]	Our method
Without movements	0.99	0.99	0.99	0.99	0.99	0.99	0.99
With movements	0.98	0.86	0.87	0.82	0.88	0.87	0.99

		57	58	59	60	61	61	63	
M1		1	2	5	2	1	0	0	
M2		0	1	2	5	2	1	0	
M3	...	0	1	2	5	2	1	0	...
M4		0	0	1	2	5	2	1	
M5		1	2	5	2	1	0	0	
Total votes		2	6	15	21	11	4	1	

FIGURE 7. Voting system example where the ground truth value is 60 BPM. Each method's estimation is where the mask value is 5.

A. COMPARISON TO THE CONVENTIONAL METHODS

We compared our method with six conventional methods [11], [17], [19], [20], [24], [38]. As our method mainly depends on BCG methods, most of the methods chosen for the comparison were BCG methods. We added the only PPG method [11], since it also introduced an approach for handling head movements. We also added the method [24] that provides a framework for simultaneously tracking reliable PPG and BCG signals, to show that simply using PPG and BCG signals together (without providing an effective solution to how to combine them or explicitly handling voluntary head movements) is not the solution to our scenario. However, since the method does not provide a specific method for computing heart rate from both signals, we computed the most dominant frequencies (having the highest magnitude in the power spectrum) in both signals and averaged them. To measure the estimation accuracy of each method, mean absolute errors (MAE) and Pearson's correlation coefficients (PCC) are computed as follows.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| h_{GT}^i - h_{Est}^i \right|,$$

$$PCC = \frac{\sum_{i=1}^{N} \left(h_{GT}^i - \overline{h_{GT}} \right) \left(h_{Est}^i - \overline{h_{Est}} \right)}{\sqrt{\sum_{i=1}^{N} \left(h_{GT}^i - \overline{h_{GT}} \right)^2} \sqrt{\sum_{i=1}^{N} \left(h_{Est}^i - \overline{h_{Est}} \right)^2}},$$

where N is the number of measurements, h_{GT}^i is the ith ground-truth value and h_{Est}^i is the ith estimation in BPM. Table 3 and Table 4 show the MAE and PCC values of each method and Figure 8 shows the Bland-Altman plots for the estimation errors.

For the first set of videos without voluntary head movements, most of the conventional methods had good accuracy (MAE < 4 BPM and PCC > 0.99) especially when lighting conditions were good. It is also shown with blue circle samples of the Bland-Altman plot in Figure 8 that most conventional methods perform within an acceptable margin of error. However, our method showed much higher accuracy and had an estimation error of 1.23 BPM. For the second set of videos that contained voluntary head movements, all related BCG methods had a large decrease in accuracy (estimation errors increased by 9.61 BPM on average). In contrast, the PPG method [11] showed less sensitivity to voluntary head movement (PCC > 0.98) but the estimation errors were not small (5.72 BPM). Using PPG and BCG signals together [24] did not improve the results and the estimation errors were much higher than the PPG method. Clearly, our method showed the least sensitivity to voluntary head movements (PCC > 0.99) and had still a low estimation error of 2.78 BPM. Our method showed better responses regardless of the existence of head movements with the smallest errors in Table 3 and the densest samples on plots in Figure 8.

We also proceeded to analyze the accuracy of each method for heart rates of different ranges, from 55 BPM to 100 BPM. Figures 9 and 10 show the estimation errors of each method for different ranges of heart rates without and with voluntary head movements, respectively. Generally, all the BCG methods showed a decrease in accuracy as the heart rate ranges increased. This is the same even when using PPG and BCG signals together. However, the PPG method did not result in a decrease in accuracy for cases both with and without voluntary head movements. Our method also maintained high accuracy when there was voluntary head movements. Without head movements, the errors of our method increased as the heart rate increased. This will be because our method mainly depends on BCG methods. However, our method still had much lower errors than the PPG method at the high ranges of heart rates. The phenomena of decrease in accuracy at higher heart rate ranges is also shown in Figure 8 where the distribution of samples of BCG methods get sparse as the heart rates increase.

Finally, as the main concern of this article is to handle the voluntary head movements, we plotted the errors of each method against the percentage of the signal having the head
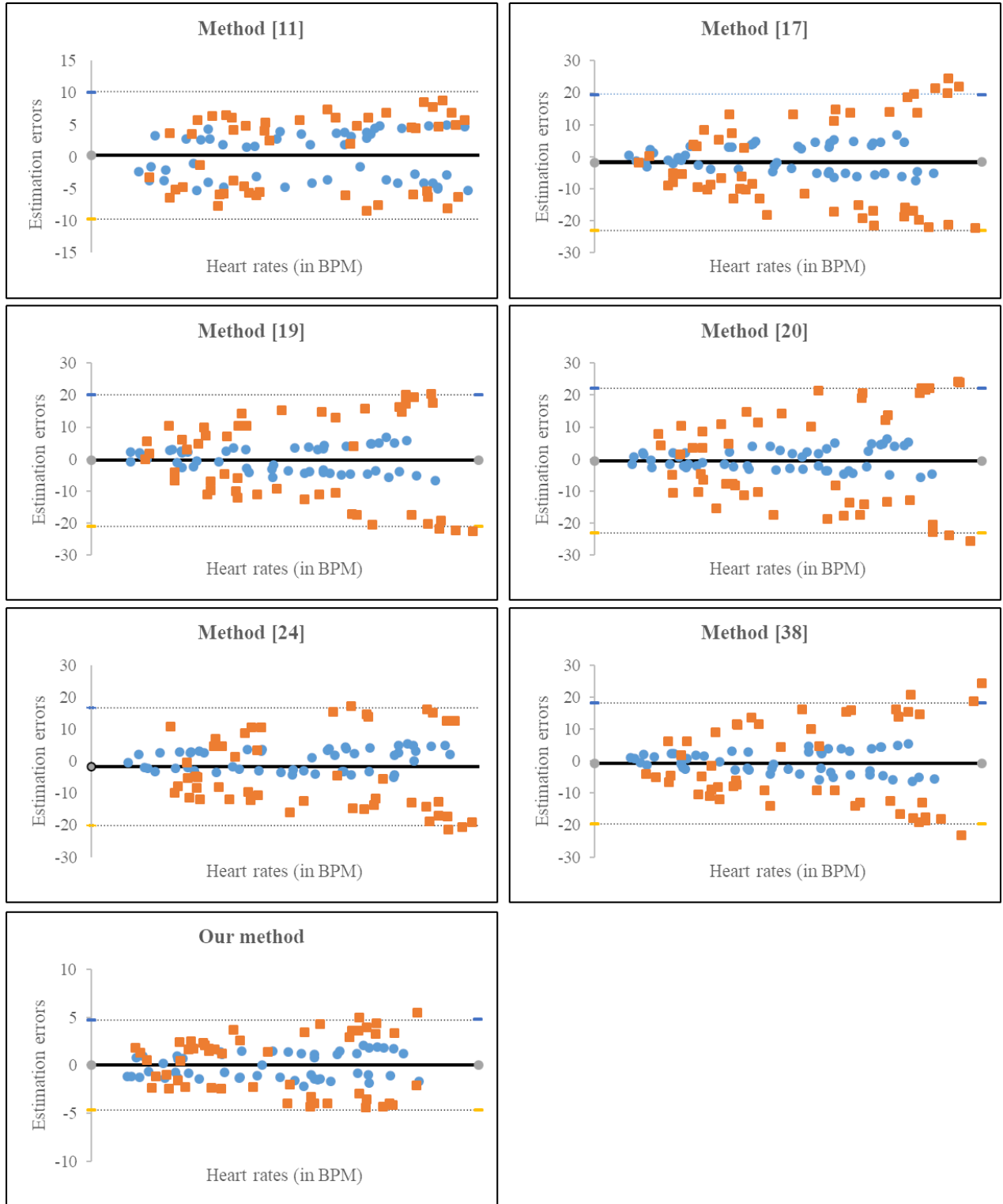
FIGURE 8. Bland-Altman plot for estimation errors of each method. For each graph, blue circle and orange rectangle points are samples for measurements without and with voluntary head movements, respectively.

movements. The percentage ranges from 5% to 50% of the signal as shown in Figure 11. As expected, the errors of the BCG methods increased significantly as the percentage increased. Even the errors of the PPG method increased in proportion to the percentage. The combination of PPG and BCG signals could not mitigate the increase in error
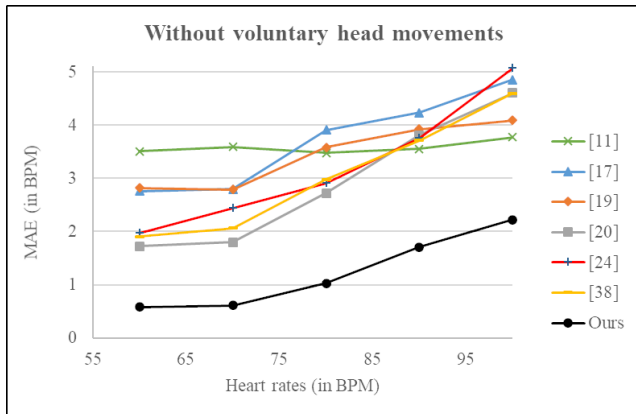
FIGURE 9. The mean absolute errors over different heart rate ranges. Cases without head movement where subjects stayed still.
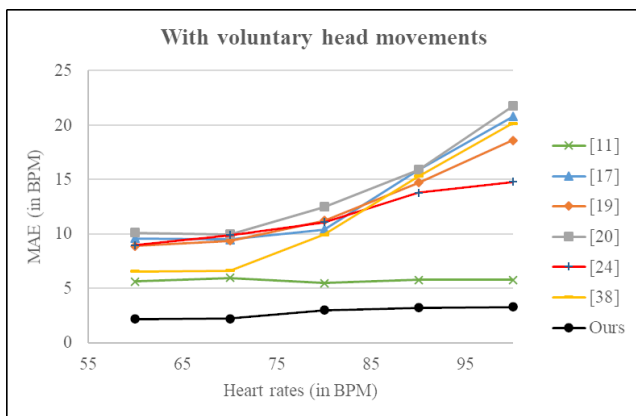

FIGURE 10. The mean absolute errors over different heart rate ranges. Cases with head movement where subjects moved naturally like they were in a video call scenario.

FIGURE 11. The mean absolute errors of different methods for different percentages of voluntary head movements contained in the signals.
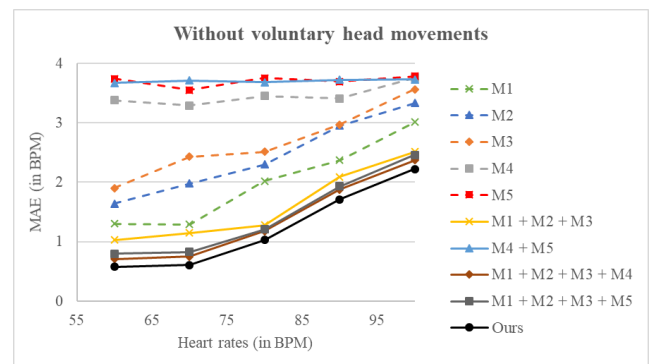
FIGURE 12. The mean absolute errors of different combinations of the variants in Table 1. Cases without head movement where subjects stayed still.

(see the results of [24]). For our method, we tested two different approaches for handling voluntary head movements. One is to simply remove the portion related to voluntary head movements and estimate the heart rate using the remaining signal, which is Ours-1 in Figure 11. The other is to remove the portion related to voluntary head movements but keep the feature tracking within the ROIs to prolong the signal, then, estimate heart rates using the prolonged signal, which is Ours-2 in Figure 11. Both Ours-1 and Ours-2 outperformed the conventional PPG and BCG methods. However, Ours-1 resulted in an increase in errors in proportion to the portion of data related to head movements. The tendency was similar to the PPG method. Nevertheless, the errors of Ours-1 were much lower throughout all the percentages. On the other hand, Ours-2 maintained good accuracy even with the high percentages and worked very stably. Given that, in the real scenarios, the percentage of head movements is rarely more than 50%, it is no doubt that Ours-2 provides a reliable solution for heart rate estimation during a video call.

B. MORE DETAILED ANALYSIS OF OUR METHOD
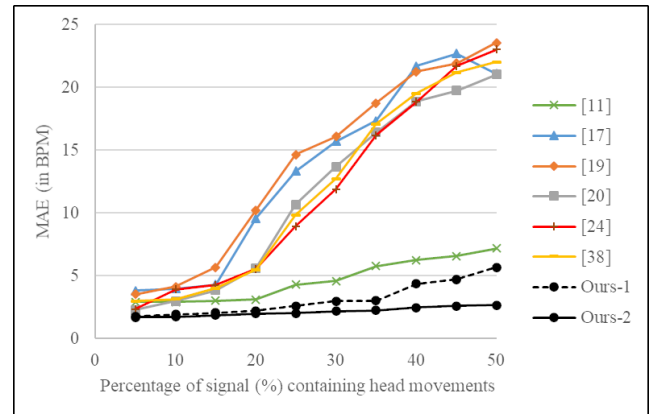To analyze in depth the accuracy and robustness of our combination method, we conducted more experiments where our

method was compared with other combinations of the variants used in the proposed voting system and also compared with the individual use of each variant. As shown in Fig. 12 and Fig. 13, our method showed the best results by combining all the variants. Using subsets of the variants caused a drop in accuracy. Specifically, the BCG variants (M1, M2, and M3) were more accurate than the PPG variants (M4 and M5) when there was no voluntary head movement. The combination of BCG variants alone provided an acceptable accuracy.
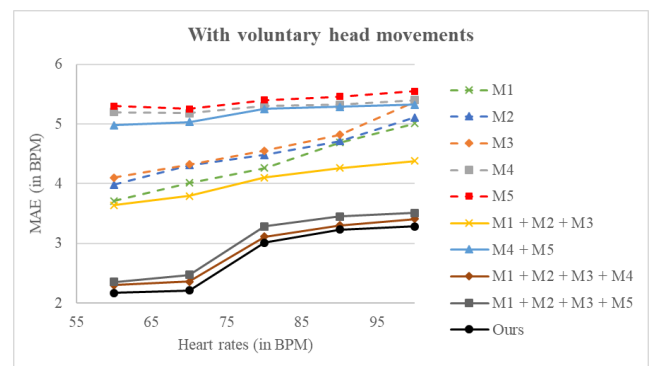

FIGURE 13. The mean absolute errors of different combinations of the variants in Table 1. Cases with head movement where subjects moved naturally like they were in a video call scenario.

FIGURE 14. Examples of subjects with different skin tones. From left, dark skin, fairly dark skin, fairly bright skin, and bright skin.
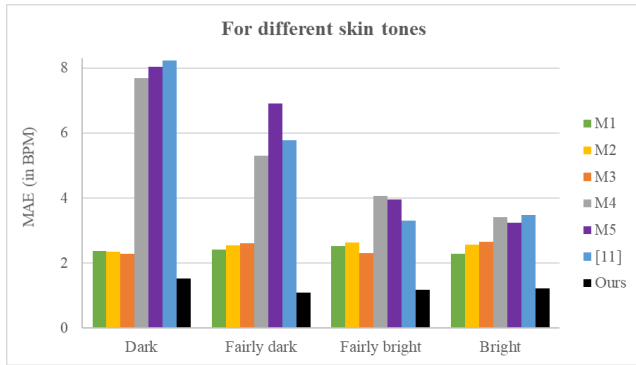

FIGURE 15. The mean absolute errors over different skin tones without voluntary head movements.

method (Section III-C), the BCG variants could be robust to voluntary head movements, thus their estimation errors were still lower than the PPG variants. However, the combination of BCG variants alone produced high errors. By using the PPG variants (M4 and M5) together, the estimation errors greatly decreased. As a result, it is evident that each variant contributes to the accuracy of our method.

Next, PPG methods tend to be sensitive to ambient illumination and skin tone of subjects. To analyze the sensitivity of our method, we divided the subjects into four groups: dark skin, fairly dark skin, fairly bright skin, and bright skin groups. Examples of each group are shown in Fig. 14. For each group, we conducted experiments without voluntary head movements for different lighting conditions and averaged the results. Figure 15 shows the estimation errors of different methods for each group. As expected, the PPG variants (M4 and M5) and the PPG method [11] showed high sensitivity to skin tone while the BCG variants (M1, M2, and M3) showed robustness. Although the estimation errors slightly increased for the dark skin tone, our method could also be robust to skin tone thanks to the robustness of BCG variants. As a result, when the subjects had no voluntary head movements, our method had an estimation error of less than 1.53 BPM regardless of the skin tone of subjects.

However, using more PPG variants together has increased accuracy. When there are large voluntary head movements, BCG methods are usually less accurate than PPG methods. However, with our voluntary head movement handling

C. COMPUTATION TIME OF OUR METHOD
Our hybrid method uses multiple PPG and BCG methods for each frame, which can increase the computational time complexity. To this end, we designed the variants that are the

Signal acquisition (each frame)	
Regions tracking [A1]	**Pixel intensity averaging** [A2]
26 ms	10 ms

Signal filtering (last frame)			
MAF [B1]	**Butterworth BPF** [B2]	**PCA** [B3]	**ICA** [B4]
1 ms	16 ms	3 ms	13 ms

Heart-related signal extraction (last frame)	
Most periodic signal [C1]	**Averaging raw signals** [C2]
3 ms	1 ms

Heart rate computation (last frame)		
Most dominant frequency [D1]	**Adaptive peaks averaging** [D2]	**Optimization tree** [D3]
2 ms	4 ms	45 ms

(a)

For each frame		For the last frame					
Variant	Signal acquisition	Head movement detection	Signal filtering	Signal extraction	Heart rate computation	Voting system	Total
M1			B1 + B2 + B1: 18 ms	C2: 3 ms	D3: 45 ms		
M2			B1 + B2 + B1: 18 ms	C2: 3 ms	D2: 4 ms		
M3	A1 + A2: 36 ms		B1 + B2 + B1: 18 ms	C2: 3 ms	D1: 2 ms		
M4			B1 + B2 + B1+B3: 21 ms	C1: 1 ms	D1: 2 ms		
M5			B1 + B2 + B1+B4: 31 ms	C1: 1 ms	D2: 4 ms		
	36 ms	4 ms	106 ms	11 ms	57 ms	0 ms	178 ms

(b)

FIGURE 16. Processing time of each step used on different PPG and BCG variants of our method. Times taken (a) by sub-processes in each step and (b) by each step.
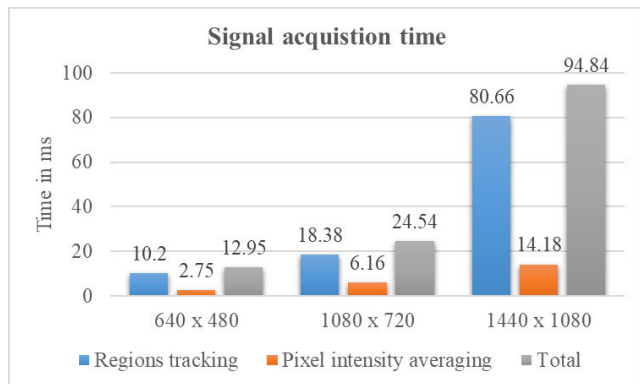
FIGURE 17. Processing time of signal acquisition step according to the input video resolution.

same in most processing steps, instead of using conventional PPG and BCG methods. This helped to minimize the computational burden of implementing multiple methods, but the computational time complexity is still high. Therefore, our implementation uses a multi-threading scheme to parallelize some operations. For example, we need to track ROIs for obtaining the PPG and BCG raw signals. Since each ROI is tracked individually, the tracking is parallelized. In our experiment with the large image resolution of 1280 × 720, the tracking time for each ROI is around 20 ms. However, the total tracking time for five ROIs was reduced to 36 ms on

average by parallelization. Therefore, even with the process of averaging the pixel intensity in the forehead and nose ROIs to obtain the PPG raw signals, the parallelization allowed the average computation time for each frame to be kept bellow 40 ms and our method can run in real time and stably at around 25 fps. Figure 16 shows details of the computation time for each step of our method. Since the signal acquisition step is done within the capture time of each frame, it is not included in the total computation time that is the sum of computation times of the other steps only done in the last frame. However, as the signal acquisition step takes 20.48 s to capture 512 frames at 25 fps, the user has to wait for 20.658 (= 20.48 + 0.178) s in total to read the resulting heart rate. Nevertheless, as our method mainly targets video calls, the whole estimation processes are implemented to run in background. Therefore, users do not have to wait since they would have their heart rates measured while having a video call.

The resolution of input video may change the computation time of our method. However, only the signal acquisition step is influenced by the video resolution. Figure 17 shows the computation time of the signal acquisition step according to the video resolution. The computation time increased rapidly as the video resolution increased. Therefore, considering the real-time capability, 1280 × 720 was the maximum resolution for our experimental setup.
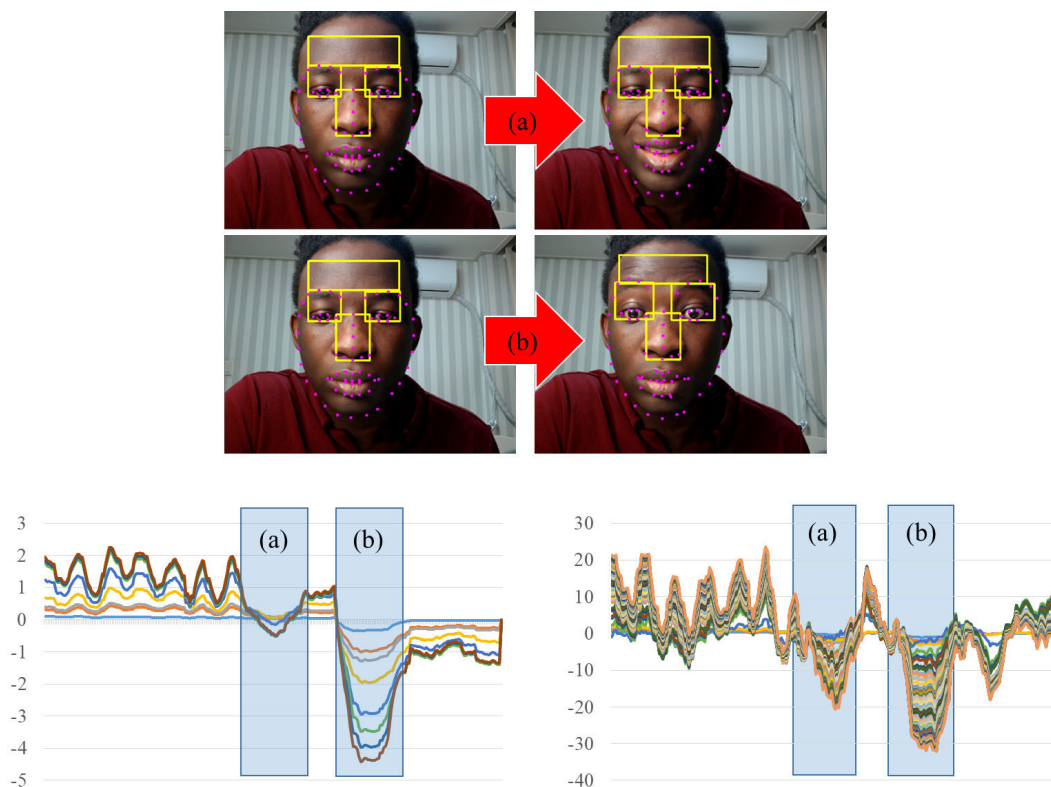

FIGURE 18. Videos with two different facial expressions (non-rigid motions). (a) Smile and (b) eyebrows up. Both facial expressions cause the head to move slightly upwards, resulting in sudden changes in the signals. The left-side signal is one obtained by the region tracking, while the right-side signal is obtained by the feature point tracking.

D. OTHER FACTORS TO BE CONSIDERED IN A VIDEO CALL SCENARIO

In this article, we focused on addressing voluntary head movements. However, in the video call scenario, large change in facial expressions is another factor that decreases the accuracy of PPG and BCG methods using face videos [11]. Therefore, we tested our method for some videos with large changes of facial expressions as shown in Figure 18. Unlike what was expected, the region tracking method (Section III-A) did not have a clear advantage over the feature tracking method in reducing the influence of facial expression change. However, non-rigid motions caused by facial expressions also resulted in rapid vertical changes in the signals representing the trajectories of facial regions or features; the head movement detection method (Section III-C) could successfully detect and remove such non-rigid motions. Therefore, it was roughly confirmed that our method can also be used to address facial expression changes.

In the video call scenario, talking actions may also create extra noise. However, talking can be considered a similar factor to the facial expression because it causes the non-rigid motions in local regions. Besides, talking-related non-rid motions are mostly observed around the mouth; to this end, we excluded the mouth region from the ROI (Figure 1). As a result, our method is rarely affected by talking actions.

E. LIMITATION OF OUR METHOD

In the video call scenario, lighting conditions can be locally and dynamically changed on the face. Such challenging lighting conditions would negatively influence the accuracy of our method. In this article, we did not consider challenging lighting conditions in depth because our main concern was to handle voluntary head movements.

In this article, we assume that voluntary head movements, facial expressions, and talking actions do not frequently and continuously occur. However, in the video call scenario, this assumption may not be always valid. In other words, our method may not be available when the assumption is not met. This is because our method simply removes the related portions from the signal, thus cannot obtain a signal long enough to estimate the heart rate.

Considering users often use their smartphones or tablets in the video call scenario, our method should also be available on the mobile devices. However, in this article, we assumed that the camera is static (not moving), thus our method may not work on mobile devices where the cameras are not static and hand motion artifacts would be a concern [39].

V. CONCLUSION

In this article, we presented a hybrid heart rate estimation method that combined different variants of PPG and BCG methods through a well-designed voting system. Our method outperformed the conventional PPG, BCG, and combination methods for different conditions of measurements. It worked accurately and consistently even in cases of voluntary head movements that may occur in real environments. Detailed analysis of the experimental results with different heart rate ranges and different percentage of head movement noise showed that our method can be a reliable solution for the challenging scenarios, such as heart rate estimation during a video call.

Future studies will focus on improving our method in more challenging lighting conditions and developing a sophisticated, dedicated method to address the facial expressions briefly discussed in Section IV-D. In addition, we plan to analyze in depth the advantage of handling such challenging conditions during video calls with regards to various metrics of cardiovascular health. We also plan to test more variants of PPG and BCG methods on the voting system to make our method even more robust.

REFERENCES

[1] E. J. Benjamin, P. Muntner, A. Alonso, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, S. R. Das, and F. N. Delling, "Heart disease and stroke statistics-2019 update a report from the American heart association," *Circulation*, vol. 139, no. 10, pp. e526–e528, 2019.

[2] A. D. Choudhury, A. Misra, A. Pal, R. Banerjee, A. Ghose, and A. Visvanathan, "Heartsense: Estimating heart rate from smartphone photoplethysmogram using adaptive filter and interpolation," in *Internet of Things. User-Centric IoT* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 150. Basel, Switzerland: Springer, 2014, pp. 203–209.

[3] L.-M. Po, X. Xu, L. Feng, Y. Li, K.-W. Cheung, and C.-H. Cheung, "Frame adaptive ROI for photoplethysmography signal extraction from fingertip video captured by smartphone," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 1634–1637.

[4] J.-P. Lomaliza and H. Park, "A highly efficient and reliable heart rate monitoring system using smartphone cameras," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21051–21071, Oct. 2017.

[5] J.-P. Lomaliza and H. Park, "Improved peak detection technique for robust PPG-based heartrate monitoring system on smartphones," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17131–17155, Jul. 2018.

[6] B. Mallick and A. K. Patro, "Heart rate monitoring system using finger tip through Arduino and processing software," *Int. J. Sci., Eng. Technol. Res.*, vol. 5, no. 1, pp. 84–89, 2016.

[7] T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2012, pp. 544–545.

[8] 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.*, Aug. 2012, pp. 2174–2177.

[9] 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.

[10] W. Zeng, Q. Zhang, Y. Zhou, G. Xu, and G. Liang, "Infrared video based non-invasive heart rate measurement," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 1041–1046.

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

[12] 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.

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

[14] M. Jain, S. Deb, and A. V. Subramanyam, "Face video based touchless blood pressure and heart rate estimation," in *Proc. IEEE 18th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2016, pp. 1–5.

[15] 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, Sep. 2014.

[16] K. H. Shelley, "Photoplethysmography: Beyond the calculation of arterial oxygen saturation and heart rate," *Anesthesia Analgesia*, vol. 105, pp. S31–S36, Dec. 2007.

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

[18] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, Dec. 2013, pp. 160–164.

[19] 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 2016.

[20] L. Wang, S. Geng, B. Liu, and Y. Jin, "Ballistocardiogram heart rate detection: Improved methodology based on a three-layer filter," *Measurement*, vol. 149, Jan. 2020, Art. no. 106956.

[21] M. A. Hassan, A. S. Malik, D. Fofi, N. M. Saad, Y. S. Ali, and F. Meriaudeau, "Video-based heartbeat rate measuring method using ballistocardiography," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4544–4557, Jul. 2017.

[22] O. T. Inan, P.-F. Migeotte, K.-S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G. K. Prisk, and M. D. Rienzo, "Ballistocardiography and seismocardiography: A review of recent advances," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1414–1427, Jul. 2015.

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

[24] 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.

[25] D. Shao, Y. Yang, F. Tsow, C. Liu, and N. Tao, "Non-contact simultaneous photoplethysmogram and ballistocardiogram video recording towards real-time blood pressure and abnormal heart rhythm monitoring," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 273–277.

[26] A. B. Hertzman and J. B. Dillon, "Applications of photoelectric plethysmography in peripheral vascular disease," *Amer. Heart J.*, vol. 20, no. 6, pp. 750–761, Dec. 1940.

[27] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali, and F. Meriaudeau, "Heart rate estimation using facial video: A review," *Biomed. Signal Process. Control*, vol. 38, pp. 346–360, Sep. 2017.

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

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

[30] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[31] H. E. Tasli, A. Gudi, and M. D. Uyl, "Remote PPG based vital sign measurement using adaptive facial regions," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1410–1414.

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

[33] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 172–175, 2nd Quart., 2002.

[34] W. Chen and D. McDuff, "DeepPhys: Video-based physiological measurement using convolutional attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 356–373.

[35] X. Niu, S. Shan, H. Han, and X. Chen, "RhythmNet: End-to-end heart rate estimation from face via spatial-temporal representation," *IEEE Trans. Image Process.*, vol. 29, pp. 2409–2423, 2020.

[36] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision With the OpenCV Library*. Newton, MA. USA: O'Reilly Media, 2008.

[37] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 532–539.

[38] H. Lee, A. Cho, S. Lee, and M. Whang, "Vision-based measurement of heart rate from ballistocardiographic head movements using unsupervised clustering," *Sensors*, vol. 19, p. 3263, Jul. 2019.

[39] J.-P. Lomaliza and H. Park, "Improved heart-rate measurement from mobile face videos," *Electronics*, vol. 8, no. 6, p. 663, Jun. 2019.

[40] *Polar H7 Chest Belt Heart Rate Sensor*. Accessed: Apr. 12, 2020. [Online]. Available: https://support.polar.com/en/support/H7_heart_rate_sensor

JEAN-PIERRE LOMALIZA received the M.S. degree in electronic engineering from Pukyong National University, Busan, South Korea, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include augmented reality and computer vision.

HANHOON PARK received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Hanyang University, Seoul, South Korea, in 2000, 2002, and 2007, respectively. From 2008 to 2011, he was a Postdoctoral Researcher with the NHK Science and Technology Research Laboratories, Tokyo, Japan. In 2012, he joined the Department of Electronic Engineering, Pukyong National University, Busan, South Korea, where he is currently a Professor. His current research interests include augmented reality, human–computer interaction, and affective computing.

MARK BILLINGHURST received the Ph.D. degree in electrical engineering from the University of Washington, in 2002. He was the Director of the HIT Laboratory NZ, University of Canterbury, from 2002 to 2015, developing innovative computer interfaces. He is currently a Professor of human computer interaction with the University of South Australia, researching empathic computing. He has produced more than 400 publications in areas, such as augmented reality, multimodal interaction, and mobile interfaces. He was elected as a Fellow of the Royal Society of New Zealand. In 2013, his work in augmented reality received the IEEE VR Technical Achievement Award.

• • •