Mobile Remote Photoplethysmography: A Real-Time Realization Perspective

Hooseok Lee
Kyung Hee University
Hoon Ko
Kyung Hee University
Heewon Chung
Kyung Hee University
Yunyoung Nam
Soonchunhyang University
Sangjin Hong
SUNY-Stony Brook University
Jinseok Lee (gonasago@khu.ac.kr)
Kyung Hee University

Research Article

Keywords: algorithm, rppg, instantaneous, dataset, remote, sensors

Posted Date: October 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-950943/v1

Mobile Remote Photoplethysmography: A Real-Time Realization Perspective

Hooseok Lee, M.S[1,†], Hoon Ko, M.S[1,†], Heewon Chung, M.S[1,†], Yunyoung Nam, Ph.D[2], Sangjin Hong, Ph.D[3], and Jinseok Lee, Ph.D[1,*]

[1]Department of Biomedical Engineering, Kyung Hee University, Yongin, Republic of Korea

[2]Department of Computer Science and Engineering, Soonchunhyang University, Asan, Republic of Korea

[3]Department of Electrical Engineering, SUNY-Stony Brook University, NY, USA

[*]gonasago@khu.ac.kr

†these authors contributed equally to this work.

Abstract

Remote photoplethysmography (rPPG) sensors have attracted a significant amount of attention as they enable the remote monitoring of instantaneous heart rates (HRs) and thus do not require any additional devices to be worn on fingers or wrists. In this study, we mounted rPPG sensors on a robot for active and autonomous instantaneous HR (R-AAIH) estimation. Subsequently, we proposed the algorithm providing accurate instantaneous HRs, which can be performed in real time with vision and robot manipulation algorithms. By simplifying the extraction of facial skin images using saturation (S) values in the HSV color space, and selecting pixels based on the most frequent S value on the face image, we achieved reliable HR assessment. The results of the proposed algorithm using the R-AAIH were evaluated by rigorous comparison with the results of existing algorithms on the UBFC-RPPG dataset (n = 42). Our algorithm exhibited an average absolute error (AAE) of 0.71 beats per minute (bpm). The developed algorithm is simple and the processing time is less than 1 s (275 ms for an 8-s window). The algorithm was further validated on our own dataset (BAMI-RPPG dataset [n = 14]) with an AAE of 0.82 bpm.

Introduction

Remote photoplethysmography (rPPG) sensors have attracted much attention because of their capacity to measure instantaneous heart rates (HRs) without any contact with the human skin. The rPPG sensor uses a camera for face detection and records facial skin images representing changes in the arterial blood volume between the systolic and diastolic phases of the cardiac cycle. Thus, the sensors enable remote monitoring of instantaneous HRs and thus do not require any device to be worn on the skin such as on the finger [1-3] or wrist [4-6]. However, these convenient monitoring capabilities have not led to successful commercialization via FDA approval because they provided lower accuracy than contact-PPG. They respond more sensitively to human movement and ambient light change. Recently, rPPG sensors have been mainly applied to a webcam connected to a computer, and related studies provided various algorithms providing high accuracy [7-15]. However, such a stationary approach makes it difficult to measure instantaneous HR during daily life activities, and it makes practical application difficult in various medical fields. We believe that, for rPPG sensors to be applied and utilized in various medical fields, it is necessary to acquire PPG signals without any space restrictions.

To extend the use of rPPG sensors, we proposed the algorithm providing accurate instantaneous HRs, which can be performed in real time with vision and robot manipulation algorithms. In this study, we mounted rPPG sensors on a robot for active and autonomous instantaneous HR (R-AAIH) estimation. This dynamic approach allows the robot to actively monitor instantaneous HRs. Furthermore, it enables active medical services, such as providing HR information to people with abnormal heart rates such as atrial fibrillation around the robot. As this study is the first step to realizing such active medical services, our R-AAIH navigates within a specific space while avoiding obstacles. While navigating, it recognizes a human face and detects facial skin images, which are then converted into rPPG signals for an instantaneous estimation of the HR. More specifically, R-AAIH involves six stages: simultaneous localization and mapping (SLAM), robot navigation, face detection, facial skin extraction, rPPG signal conversion, and HR estimation. The SLAM is the initial stage whereby the robot constructs or updates a map of an unknown environment while simultaneously keeping track of its position [16]. During the robot navigation stage, the robot determines its own spatial position and constructs a plan for a path toward a designated position [17]. Face detection involves the detection of faces using computer vision [18]. The facial skin extraction stage identifies those facial skin pixels that change according to the cardiac cycle. The stages in the rPPG signal conversion and HR estimation are used to compute the HR value by the acquisition of PPG signals obtained from the facial skin pixels over time.

In this study, we focused on computation complexity as well as accuracy for real-time implementation. We emphasize that the aim of this study was not to obtain a full-fledged optimized robot system, but to describe the implementation of all the six required stages and focus on the development of an achievable real-time system for providing instantaneous HR values per second using low-power hardware. In particular, we optimized the facial skin image stage by selecting pixels based on the most frequent saturation (S) value on the facial image, which resolved one of the obstacles of real-time operation. In the facial skin extraction stage, a face landmark-based approach has been widely described [15,19-21]. This approach involves the recognition of the geometric structure of faces in images and obtaining a canonical alignment of the face based on translation, scale, and rotation. The resultant face landmark networks propose a facial skin area in a variety of different forms such as a rectangle-, a bottom face-, and polygon-shapes, and provide high accuracy for estimating instantaneous HRs. However, the face landmark-based approach leads to heavy computation complexity, which may make real-time processing difficult when computing HR values per second when the robot is connected through a laptop personal computer. We first introduce our developed R-AAIH and propose a simplified facial skin extraction stage with limited accuracy degradation, which allows accurate instantaneous HR estimation in real-time.

Methods

Robot system description

Fig. 1 shows our device designed for an active and autonomous instantaneous HR estimation using a robot equipped with a rPPG sensors. The complete system involves six stages of simultaneous localization and mapping (SLAM), robot navigation, face detection, face skin extraction, rPPG signal conversion, and HR estimation. Fig. 1(a) shows our robot developed using a turtlebot2 framework (YUJIN ROBOT, Incheon, Korea). Within the framework, a three-dimensional (3D) camera (ASTRA PRO, ORBBEC, Michigan, USA) is mounted on the device for the SLAM and robot navigation stages, and a web camera (Logitech BRIO, Switzerland) is used for the remaining stages. Both cameras are operated through a laptop computer (TFG175, Hansung, Seoul, Republic of Korea) with AMD Ryzen 5 3400G, 3.70GHz processor.

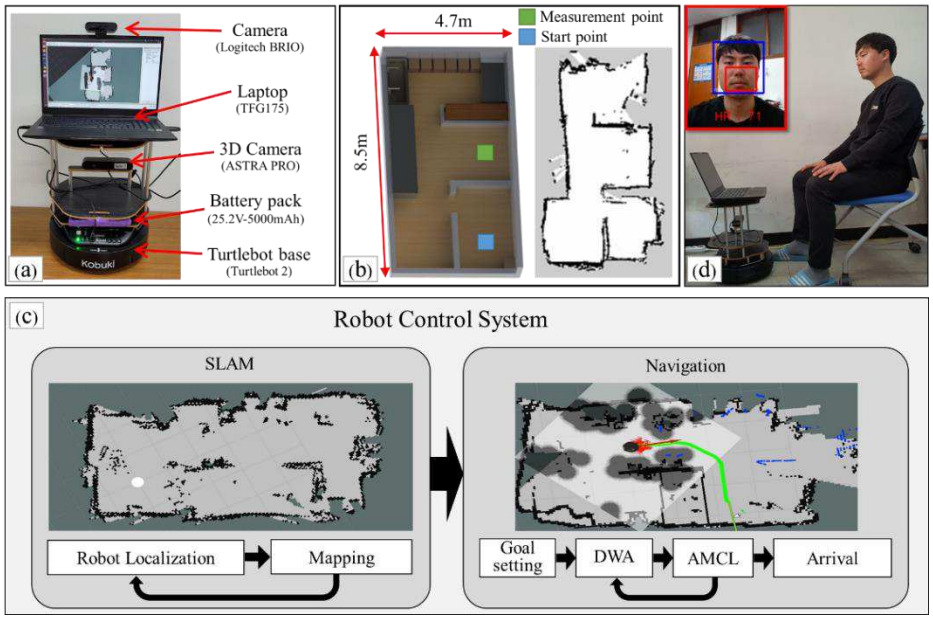

Figure 1. An overview of the proposed system for instantaneous HR estimation using a rPPG mounted on a robot; (a) The robotic device is based on a turtlebot2 framework with a 3D camera, a laptop and a webcam; (b) SLAM and navigation; (c) Real-time instantaneous HR estimation via face detection, face skin image extraction, rPPG acquisition; (d) Algorithm flow chart for SLAM and navigation.

The device is provided with an initial map that indicates the designated start/destination coordinates (Fig. 1(b), left), the robot then performs the SLAM and constructs a map of its surroundings based on data from the 3D camera (Fig. 1(b), right). In addition, the robot localizes its position within the mapped environment [16,22]. The SLAM iterates the mapping and localization data (Fig. 1(c), left) with respect to the initial map. In this study, we used a factored solution, the FastSLAM, which estimates the robot position using a particle filter and updates the map using an extended Kalman filter [23].

Once the SLAM is completed, the robot navigates to the designated start position, and continues until it reaches the designated destination. The navigation stage includes two iterative steps of localization and pathfinding (Fig. 1(c), right). For the localization step, we used the adaptive Monte Carlo localization (AMCL), known as particle filter localization [24-26]. The algorithm uses a particle filter to determine the distribution of likely states to define where the robot is initially localized combined with a posterior particle density estimation function. With each movement of the robot, the device updates the particle distribution to predict its new state (position and velocity), and the particles are resampled via recursive Bayesian estimation based on the measurements obtained (depth information) via a 3D camera. For the path-finding step, we used a dynamic window approach (DWA) to efficiently generate the trajectory for subsequent movement [27]. The DWA is an online collision avoidance strategy for mobile robots and is derived directly from the dynamics of the robot, which is normally designed to adapt to the constraints imposed by limited velocities and accelerations of the robot.

While a robot navigates, it searches for a human in its surroundings by detecting faces via the web camera. For real-time face detection, we used a deep neural network (DNN) based single shot scale-invariant face detector (S3FD) [28]. The S3FD uses a scale-equitable framework with a wide range of anchor-associated layers and a series of reasonable anchor scales to handle different facial sizes. More specifically, the architecture consists of a truncated VGG-16 network with extra convolutional layers, detection convolutional layers, normalization layers, predicted convolutional layers, and multi-task loss

layer. The detection layers are associated with specific anchor scales ranging from 16 to 512, which enable the robot to detect different facial sizes. The DNN was trained using 12,880 images of the WIDER FACE training set [29], and the trained model achieved a state-of-the-art performance on most common face detection benchmarks such as Annotated Faces in the Wild (AFW) [30], PASCAL Face [31], Face Detection Dataset, and the Benchmark (FDDB) [32]. Once a face is detected, the robot stops the navigation temporarily and performs instantaneous HR estimation via rPPG acquisition. For HR estimation, face skin images are extracted from the detected face image to remove non-pulsatile information. Finally, the face skin images are converted into a rPPG signal, which provides a real-time HR value (Fig. 1(d)). The instantaneous HR estimation lasts for 1 min, after which the robot navigates to the original destination and repeats the searches for a human in its surrounding. The process is repeated until the robot reaches the destination.

Problems with real-time rPPG acquisition

In our designed robot system, video images are acquired at 30 frames per second (fps). The key aspect of our study is to ensure that all the steps including those of face detection, skin image extraction, rPPG signal conversion, and HR estimation are realizable in real-time, Thus, to acquire a 1-s rPPG, we need to perform 30 face detections and 30 skin image extractions using 30 video images within a second. Regarding the rPPG window length, a short window can increase time-resolution while a long window improves SNR. However, the gain in performance from using a longer window come at the price of an increased latency. In this study, for real-time HR estimation, we used an 8-s window rPPG segment per second (i.e., an 8-s window with a 1-s shift), similar to the parameters used in previously reported algorithms [6,33]. Then, the following question arises: is it possible to acquire the rPPG signal and compute instantaneous HR measurements within a second on a single CPU? Most previous studies have focused on the accuracy of HR estimation using rPPG [15,20,21,34,35]. However, such an algorithm when applied to a robot such as the system we designed should focus not only on accuracy but also on the computation complexity. Although it is beyond the scope of this study, via accurate and real-time instantaneous HR estimation, a robot can perform additional real-time applications such as heart rate variability (HRV) analysis [36], atrial fibrillation diagnosis [37], and cardiac rehabilitation [38,39] based on instantaneous HRs values.

Face skin extraction with relative saturation value range (RSVR)

The detected face image may sometimes include hair and/or background other than the facial skin (Fig. 2(a), yellow rectangle). Because the hair and background images do not contain any pulsatile information, the face skin extraction stage is one of the most important stages that helps acquire a clean rPPG signal. To extract the only face skin area (pixels) as a region of interest (ROI), we first converted the detected face images (inside the rectangle area) to HSV color space images and obtained the histogram of the converted saturation (S) values. The S value histogram was then applied with a median filter of length 5 (Fig. 2(b)). In the histogram, we denoted the most frequent S value on the face image by $hist_{max}$. In addition, we defined the width to be considered as the face skin region centered on a $hist_{max}$ by the TH_{range}. Regarding the pixel value, we denoted the S value in the (i, j) pixel from the kth frame image as S_k^{ij}. The pixels for the facial skin are satisfied by the following condition as:

$$hist_{max} - TH_{range}/2 < S_k^{ij} < hist_{max} + TH_{range}/2 \quad (1)$$

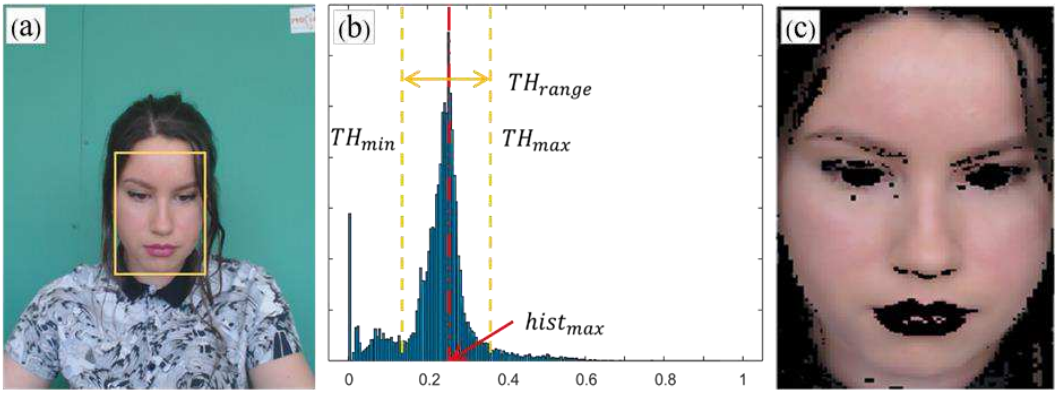

Figure 2. (In the public dataset of the UBFC-rPPG) Face skin extraction; (a) detected face image as a yellow rectangle may include hairs and some background other than the face skin, (b) The histogram of S values distributed around the center value $hist_{max}$ corresponding to the most frequent S value on the face image, (c) The resultant face skin images after applying the relative saturation (S) value range.

Similar to the notation of S_k^{ij}, we denote the red, green and blue values in the (i, j) pixel from the original image at the kth frame by R_k^{ij}, G_k^{ij} and B_k^{ij}, respectively. In addition, we denoted the red, green, and blue values corresponding to face skin image at the kth frame by $R(f)_k^{ij}$, $G(f)_k^{ij}$ and $B(f)_k^{ij}$, respectively. Then, R_k^{ij}, G_k^{ij} and B_k^{ij} can be expressed as:

$$\begin{bmatrix} R(f)_k^{ij} \\ G(f)_k^{ij} \\ B(f)_k^{ij} \end{bmatrix} = \begin{cases} \begin{bmatrix} R_k^{ij} & G_k^{ij} & B_k^{ij} \end{bmatrix}^T, & \begin{matrix} hist_{max} - TH_{range}/2 < S_k^{ij} \\ < hist_{max} + TH_{range}/2 \end{matrix}, \\ [0 \quad 0 \quad 0]^T, & otherwise \end{cases} \tag{2}$$

To find the optimum value of TH_{range}, we set the value as a relative one proportional to the size of the value $hist_{max}$ as follows:

$$TH_{range} = \alpha \cdot hist_{max}$$

where α is a constant. In this study, we set $\alpha = 0.2$. Note that we investigated the effects of a constant value α in the Results. In addition, we investigated the method of choosing the S feature over hue (H), value (V), red (R), green (G), and blue (B) for the face skin extraction in Results. Fig. 2(c) shows the resultant face skin image obtained. Recently, Boccignone et al. argued that facial skin extraction requires an adaptive threshold technique because each face has its own features [40]. In this study, because the relative S value range (RSVR) extracts the face skin pixels based on different of $hist_{max}$ values for each image, the resultant successive image pixels over time can represent the pulsatile component of the cardiac cycle under different conditions (i.e., ambient light and/or different subjects). Furthermore, the RSVR based facial skin extraction can significantly reduce the computation complexity when compared to current state-of-the-art methods [15,19-21].

rPPG conversion and HR estimation

Given the face skin image of $[R(f)_k^{ij} \quad G(f)_k^{ij} \quad B(f)_k^{ij}]^T$, we averaged all pixels corresponding to face skin for each of three channels as:

$$\begin{bmatrix} \bar{R}(f)_k \\ \bar{G}(f)_k \\ \bar{B}(f)_k \end{bmatrix} = \left[\frac{1}{W \cdot H - C_k^{otherwise}} \cdot \sum_{i=1}^{W} \sum_{j=1}^{H} \begin{bmatrix} R(f)_k^{ij} \\ G(f)_k^{ij} \\ B(f)_k^{ij} \end{bmatrix} \right] \tag{3}$$

where $\bar{R}(f)_k$, $\bar{G}(f)_k$ and $\bar{B}(f)_k$ represent the averaged pixel value for each channel. W and H are the width and height of the detected face image (rectangle area). $C_k^{otherwise}$ is the number of pixels outside the optimal range $hist_{range}$in the rectangle area for each channel. Then, the averaged pixel values for each channel can be arranged according to the image frame as:

$$\begin{aligned} R_{1:N} &= [\bar{R}(f)_1, \bar{R}(f)_2 \quad \cdots \quad \bar{R}(f)_N] \\ G_{1:N} &= [\bar{G}(f)_1, \bar{G}(f)_2 \quad \cdots \quad \bar{G}(f)_N] \\ B_{1:N} &= [\bar{B}(f)_1, \bar{B}(f)_2 \quad \cdots \quad \bar{B}(f)_N] \end{aligned} \tag{4}$$

where N is the total number of image frames. Then, the 8-s red channel data corresponding to 240 samples can be expressed as

$$R_{s+1:s+240} = [\bar{R}(f)_{s+1}, \bar{R}(f)_{s+2} \quad \cdots \quad \bar{R}(f)_{s+240}] \tag{5}$$

where $s+1$ represents the starting image frame. Note that green and blue channel data can be expressed similarly (i.e. $G_{s+1:s+240}$ and $B_{s+1:s+240}$). Based on $R_{s+1:s+240}$, $G_{s+1:s+240}$, and $B_{s+1:s+240}$, we applied the chrominance-based (CHROM) 35, and derived the rPPG signal $S_{s+1:s+240}$ as follows:

$$S_{s+1:s+240} = 1.5R_{s+1:s+240} + G_{s+1:s+240} - 1.5B_{s+1:s+240} \tag{6}$$

Using this approach, we acquired an 8-s window rPPG signal every second. For each window signal, we applied a fourth-order Butterworth bandpass filter (BPF) with cutoff frequencies of 0.4 and 4 Hz. The range of approximately 40–200 bpm covers the HRs of the subjects of all ages, both at rest and during high-intensity physical activity 41-43. The filtered signal was then normalized to a zero mean with a unit variance. Subsequently, we estimated the power spectral density (PSD) of the filtered signal using Welch's method 15, where the segment was divided into 8 sub-segments with 50% overlap, and each sub-segment was windowed with a Hamming window. Finally, we found the maximum power frequency f_{HR}(Hz), and estimated the instantaneous HR as $HR_{est}(t) = 60 \cdot f_{HR}$ bpm.

UBFC-rPPG dataset

For the training dataset, we used the publicly released UBFC-RPPG [44], which is specifically designed for remote HR measurement tasks. It contains one-minute long 42 videos from 42 different subjects. The videos were recorded using a Logitech C920HD Pro camera with an fps of 30 and resolution of 640 × 480 pixels in an uncompressed 8-bit RGB format. Each subject was made to sit in front of a camera, roughly 1 meter away from it. The subjects were required to play a time-sensitive mathematical game, which caused variations in their HRs. The video recorded natural movements of the subjects, including the rigid and non-rigid ones. During video recording, a transmitted pulse oximeter (CMS50E)-based PPG

signal was simultaneously measured on a finger to obtain a reference HR $HR_{true}(t)$. Similar to the acquisition of the $HR_{est}(t)$, we computed $HR_{true}(t)$ using an 8-s PPG signal obtained via the transmitted pulse oximeter device of the PowerLab 8/35 (ADInstrument, Sydney, Australia).

BAMI-rPPG dataset

For the testing dataset, we collected one-minute long videos from 14 different subjects. A total of 10 males and 4 females, with an average age of 29.21±2.36 years, were involved in the creation of this dataset. This study was approved by the institutional review board of Wonkwang University in Korea. All the participants provided written informed consents. All methods were performed in accordance with the relevant guidelines and regulations. The BAMI-RPPG dataset was built using our designed robot navigation system as described in Fig. 1. Each participant was randomly positioned in an indoor environment measuring 8.5 m × 4.7 m, and our robot navigated searching for a human using face detection after SLAM. When a face was detected, the face images were recorded, and the facial skin area was extracted with the relative S value range, which was then applied to the CHROM method for rPPG acquisition. Every second, the final 8-s rPPG was filtered, and the HR was estimated using PSD. During the HR estimation, a transmitted PPG signal was obtained using the finger-type oxygen saturation device for the reference HR, $HR_{true}(t)$.

Evaluation and metrics

We first investigated how to select the S feature over hue (H), value (V), red (R), green (G), and blue (B) values for the facial skin extraction in the training dataset, UBFC-RPPG (n = 42). In addition, we investigated the effects of parameter α by varying it from 0.1 to 0.5 with an interval of 0.1. Next, the performance of the proposed algorithm was evaluated in terms of accuracy and computation complexity. For performance comparison, we first compared our algorithm with ICA 15, POS 34, and CHROM 35 only. Next, we compared the performances when ICA, POS, and CHROM was each applied to the landmark-based face skin extraction. Regarding the landmark approach, we investigated the performance from the rectangle-15, bottom face-20 and polygon-21-face based methods as shown in Fig. 3. Furthermore, we validated our proposed algorithm on the testing dataset, BAMI-RPPG (n = 14). For validation, we also compared the performance of our algorithm with state-of-art methods to determine accuracy and computation complexity.

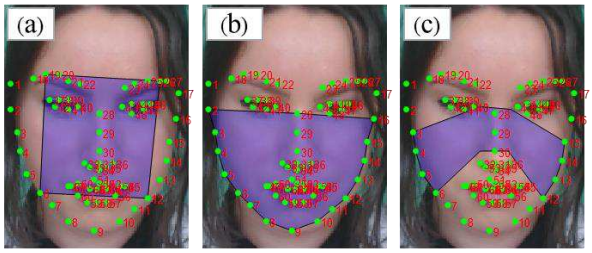

Figure 3. Landmark-based face skin extraction: (a) a rectangular face region of interest (ROI) [15], (b) a bottom face ROI [20], and (c) a polygonal face ROI [21].

The accuracy of the algorithm was evaluated by calculating the absolute error (AE) of its estimation:

$$AE(t) = |HR_{est}(t) - HR_{true}(t)| \quad (7)$$

where $HR_{true}(i)$ is the true HR (bpm) in the i^{th} window. The overall evaluation of HR estimation was performed on the basis of the absolute value of the AEs (AAE) (bpm), average of the relative AEs (ARE) (%):

$$AAE = \frac{\sum_{t=1}^{N} AE(t)}{N} \quad (8)$$

$$ARE = \frac{\sum_{i=1}^{N} \frac{AE(t)}{HR_{true}(t)} \times 100}{N} \quad (9)$$

where N is the total number of windows used for HR estimation.

To determine computation complexity, we investigated the computation time for all stages of face detection, face skin extraction, rPPG acquisition, and HR estimation, and evaluated whether the entire process was achievable in real-time in our designed robot system. Over the entire process, the robot needs to perform 30-time face detections and 30-time face skin extractions, one rPPG conversion, and one HR calculation within a second. Thus, we defined the processing time within a second (PTOS) (ms) s:

$$PTOS\ (ms) = 30 \cdot (T_{fd} + T_{fse}) + T_{rc} + T_{he} \quad (10)$$

where T_{fd}, T_{fse}, T_{rc} and T_{he}denote the computation time for face detection, face skin extraction, rPPG conversion, and HR estimation, respectively.

Results

Results using the UBFC-rPPG dataset

Using the UBFC-RPPG dataset, our proposed algorithm was evaluated for a total of 2184 windows corresponding to 42 1-min. video files. Fig. 4 shows a representative example of face skin extraction based on different features (H, S, and V) and different values of α (0.1 to 0.5). The result shows that S feature preserves more face skin images when α=0.2. Including HSV features, we further investigated RGB features for all 42 subjects. Table 1 summarizes the resultant AAE values according to the features (H, S, V, R, G, and B) and the parameter α (0.1 to 0.5) for face skin extraction. Among all possible features, S provided the lowest AAE value of 0.71 bpm when α=0.2.

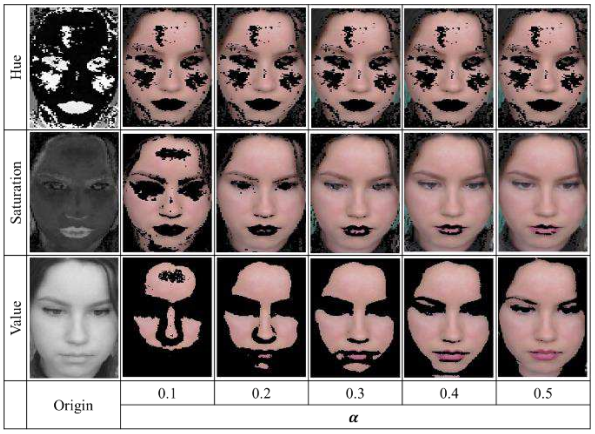

Figure 4. Landmark-based face skin extraction: (a) a rectangular face region of interest (ROI) [15], (b) a bottom face ROI [20], and (c) a polygonal face ROI [21].

Table 1. AAE values for face skin extraction according to H, S, V, R, G, and B features and parameter α using the UBFC-RPPG dataset.

Dataset	α	HSV			RGB		
		H	S	V	R	G	B
UBFC-RPPG	0.1	1.94	0.73	1.22	1.23	1.15	1.29
	0.2	1.63	**0.71**	1.04	0.98	1.07	0.99
	0.3	**1.43**	0.82	0.87	0.96	0.92	0.94
	0.4	1.46	0.94	**0.85**	0.93	0.92	**0.88**
	0.5	2.57	0.96	0.86	**0.93**	**0.89**	0.89

Table 2 compares the characteristics of AAE, ARE, and PTOS. With regard to accuracy, when ICA, POS, and CHROM were used without face skin extraction, the AAE values were 2.09 bpm, 1.26 bpm, and 1.13 bpm, respectively. When the landmark- based facial skin analysis was applied, the accuracy was enhanced by decreasing AAE values to 0.79 bpm. However, the computation time, PTOS, significantly increased up to 80,181 ms on an ADM Ryzen 5 3400G CPU at 3.70 GHz personal computer when the landmark-based face skin was applied. Conversely, our method provided not only low AAE and ARE values of 0.71 bpm and 0.75%, but also low PTOS of 275 ms, which is achievable in real-time.

Table 2. Comparison of the performances of various methods in terms of accuracy (AAE and ARE) and computation time (PTOS): UBFC-RPPG (42 datasets).

Face skin extraction	None			Landmark [19]									**Our method**
				Rectangle [15]			Bottom face [20]			Polygon face [21]			
rPPG conversion	ICA [15]	POS [34]	CH-ROM [35]	ICA [15]	POS [34]	CH-ROM [35]	ICA [15]	POS [34]	CH-ROM [35]	ICA [15]	POS [34]	CH-ROM [35]	
AAE (bpm) ARE (%)	2.09. (2.07)	1.26 (1.28)	1.13 (1.19)	2.62 (2.66)	0.79 (0.82)	0.78 (0.82)	0.79 (0.82)	0.79 (0.82)	0.79 (0.82)	1.05 (1.07)	0.91 (0.94)	0.80 (0.84)	**0.71 (0.75)**

PTOS (ms)	245	244	244	80,181	80,180	80,180	80,181	80,180	80,180	80,181	80,180	80,180	**275**
CPU Info.	AMD Ryzen 5 3400G at 3.70GHz												

Results using the BAMI-rPPG dataset

Our proposed algorithm with the H feature and α=0.2 was applied to the test dataset (BAMI-RPPG). The overall performance for 14 subjects is summarized in Table 3, where we compare the landmark-based approach to our algorithm in terms of AAE, ARE and POTS. The results show that our algorithm provides a low AAE of 0.82 bpm and an ARE of 1.12%. The other landmark-based approach provided an AAE ranging between 0.77 bpm and 2.03 bpm, whereas the ARE ranged between 1.04% and 2.39%. With respect to accuracy, our algorithm showed similar superior results even when compared to the landmark-based approach, which required high computation complexity.

Table 3. Performance comparison of various methods in terms of accuracy (AAE and ARE) and computation time (POTS): independent testing dataset – BAMI-RPPG (n = 14).

subjects	Face skin extraction	Landmark [19]									Our method
		Rectangle [15]			Bottom face [20]			Polygon face [21]			
	rPPG acquisition	ICA [15] AAE (bpm) ARE (%)	POS [34] AAE (bpm) ARE (%)	CH-ROM [35] AAE (bpm) ARE (%)	ICA [15] AAE (bpm) ARE (%)	POS [34] AAE (bpm) ARE (%)	CH-ROM [35] AAE (bpm) ARE (%)	ICA [15] AAE (bpm) ARE (%)	POS [34] AAE (bpm) ARE (%)	CH-ROM [35] AAE (bpm) ARE (%)	
1		1.51 (2.11)	1.37 (1.88)	1.45 (1.99)	1.13 (1.55)	1.26 (1.73)	1.40 (1.92)	0.80 (1.15)	0.80 (1.15)	1.22 (1.73)	0.84 (1.21)
2		0.82 (1.28)	1.06 (1.63)	1.01 (1.56)	0.76 (1.19)	0.83 (1.28)	1.03 (1.60)	0.48 (0.75)	0.49 (0.76)	0.54 (0.84)	0.56 (0.86)
3		0.66 (0.87)	0.56 (0.74)	0.65 (0.86)	0.63 (0.83)	0.57 (0.76)	0.65 (0.86)	0.75 (0.99)	0.62 (0.81)	0.66 (0.87)	0.55 (0.73)
4		0.77 (1.09)	0.57 (0.82)	0.61 (0.86)	0.93 (1.32)	0.71 (1.02)	0.79 (1.13)	0.59 (0.84)	0.42 (0.59)	0.47 (0.67)	0.63 (0.89)
5		0.57 (0.81)	0.55 (0.79)	0.57 (0.81)	0.42 (0.59)	0.38 (0.53)	0.39 (0.55)	0.42 (0.60)	0.41 (0.58)	0.45 (0.64)	0.48 (0.69)
6		1.02 (1.09)	1.06 (1.14)	1.42 (1.49)	0.99 (1.05)	0.91 (0.99)	0.95 (1.02)	17.4 (18.7)	0.71 (0.77)	1.03 (1.11)	0.65 (0.70)
7		0.73 (0.84)	0.73 (0.84)	0.67 (0.78)	0.91 (1.01)	0.59 (0.68)	1.13 (1.30)	0.97 (1.12)	1.43 (1.69)	1.70 (1.99)	1.00 (1.18)
8		0.40 (0.54)	0.44 (0.61)	0.44 (0.60)	0.41 (0.56)	0.48 (0.66)	0.47 (0.64)	0.39 (0.53)	0.42 (0.57)	0.46 (0.63)	0.35 (0.48)
9		0.51 (0.67)	0.60 (0.79)	0.71 (0.93)	0.54 (0.70)	0.63 (0.83)	0.68 (0.88)	0.53 (0.70)	0.57 (0.75)	0.60 (0.78)	0.53 (0.70)
10		0.45 (0.58)	0.48 (0.63)	0.66 (0.87)	0.49 (0.64)	0.54 (0.70)	0.51 (0.67)	0.45 (0.58)	0.48 (0.62)	0.49 (0.63)	0.47 (0.61)
11		1.97 (3.13)	1.60 (2.53)	2.04 (3.23)	0.76 (1.18)	0.83 (1.28)	1.22 (1.88)	0.95 (1.47)	0.80 (1.23)	0.79 (1.22)	0.80 (1.24)
12		2.45 (3.75)	2.78 (4.25)	2.53 (3.87)	2.57 (3.88)	2.64 (4.05)	2.65 (4.04)	1.11 (1.72)	1.40 (2.21)	1.04 (1.61)	1.98 (3.04)
13		2.33 (3.07)	0.74 (0.99)	0.60 (0.79)	1.83 (2.40)	1.18 (1.54)	2.39 (3.10)	1.19 (1.55)	0.71 (0.93)	0.89 (1.16)	1.12 (1.45)
14		1.36 (1.67)	1.27 (1.57)	2.00 (2.38)	4.33 (5.35)	1.59 (1.96)	2.20 (2.69)	2.31 (2.81)	1.55 (1.90)	1.94 (2.37)	1.52 (1.85)
Mean		**1.11 (1.54)**	**0.99 (1.37)**	**1.10 (1.50)**	**1.19 (1.59)**	**0.94 (1.29)**	**1.18 (1.59)**	**2.03 (2.39)**	**0.77 (1.04)**	**0.88 (1.16)**	**0.82 (1.12)**
POTS		**80,181**	**80,180**	**80,180**	**80,181**	**80,180**	**80,180**	**80,181**	**80,180**	**80,180**	**275**
CPU Info		AMD Ryzen 5 3400G @ 3.70GHz									

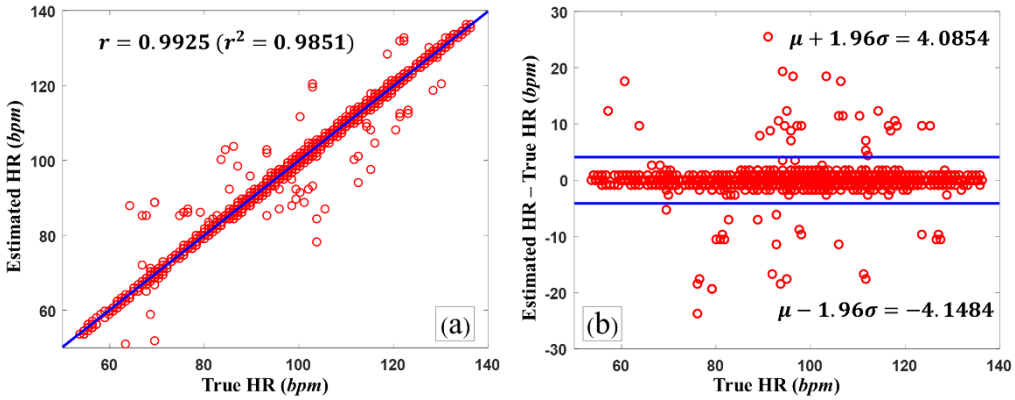

Figure 5. (a) Pearson correlations between estimated HRs and true HRs (r = 0.9925). (b) Bland-Altman plot(μ = -0.0315 and σ = 2.1005).

Fig. 5(a) shows the Pearson coefficients between the estimated valid and true HRs. The Pearson correlation coefficient of our model was 0.9925 ($r^2 = 0.9851$). Fig. 5(b) shows a Bland-Altman plot. The limit of agreement (LOA) lay between –4.05 bpm and 4.09 bpm (mean -0.0315 bpm, standard deviation 2.1005 bpm).

Discussion and Conclusion

We presented a system for an active and autonomous estimation of instantaneous HR using a rPPG mounted on a robotic device. Our proposed system makes it possible to measure HR daily life activities without space restrictions, and can be applied to various medical fields. For instance, it can be used for early detection of heart rate variability-related diseases such as asymptomatic atrial fibrillation 45 by actively monitoring instantaneous HRs. Throughout this paper, we focused on evaluating the HR estimation accuracy, which is one of the most important issues for rPPG. In order for our proposed system to be applied to reality, however, we should consider more complex environments. For the future study, we further investigated the performance when there are more than one subject in the designated space. Similar to the previous simulation, two participants were closely positioned in the indoor environment, and our robot navigated searching for a human using face detection after SLAM. When the two face were simultaneously detected, the face images were recorded separately, and each of facial skin area was extracted with the relative S value range, which was then applied to the CHROM method for rPPG acquisition. Fig. 6 shows one of the recorded video frames, where two faces were simultaneously detected and each instantaneous HR value was provided. We also presented the recorded video (Supplementary video data). This result shows that our proposed system can measure HRs from multiple subjects at the same time.

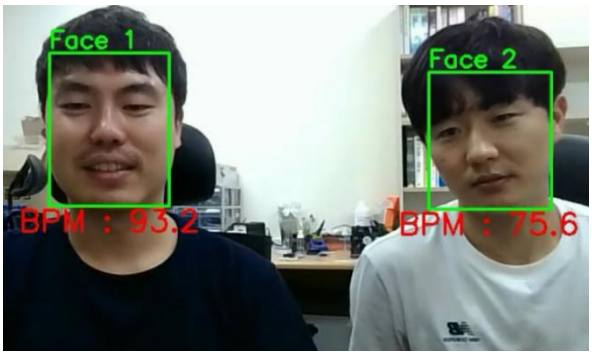

Figure 6. (Additional dataset) Simultaneous HR estimation from two subjects: ID is automatically assigned from one, and the face frames from each subject is converted to the instantaneous HR value (All two participants provided written informed consents for publication of identifying images in an online open-access publication).

For a real-time realizable algorithm, we proposed the Relative Saturation Value Range (RSVR), which effectively extracts the facial skin image for the performance enhancement of HR estimation and reduces computational complexity. However, the rPPG acquisition is based on non-contact reflectance; and thus, its signal-to-noise (SNR) is relatively low, especially under ambient light changes and movement artifacts. In both the datasets (UBFC-RPPG and BAMI-RPPG) used in this study, the detected face was relatively immobile, and the ambient light was relatively constant; thus, we were able to obtain accurate HR estimation results. Under the high motion of the detected face and/or the high change of an ambient light, the ROI of face skin leads to dynamic changes, which results in a low-quality rPPG signal and inaccurate HR estimation.

Table 4. HR estimation performance comparison between the proposed method and the FSM method according to the noise type.

Noise Types	Proposed		Proposed + FSM		
	AAE (bpm)	ARE (%)	AAE (bpm)	ARE (%)	VHR (%)
Head movement	2.79 ± 5.17	3.89 ± 7.19	0.56 ± 0.47	0.81 ± 0.67	63.96
Lighting interference	4.11 ± 9.06	5.57 ± 12.19	0.89 ± 1.93	1.26 ± 2.73	72.97

Values are reported as means ± standard deviations.

To investigate performance under these conditions, we performed additional experiments. We recorded a 1-min video with the subject moving their head rapidly and another 1-min video in which the subject was placed under highly variable ambient light conditions. For the HR reference, a transmitted PPG signal was obtained using a finger-type oxygen saturation device. Under those conditions, our algorithm provided AAEs as high as 2.79 bpm and 4.11 bpm, respectively, and the AREs increased by 3.89% and 5.57%, respectively, as summarized in Table 4. As a possible solution, we applied the recently introduced finite state machine framework [6], which automatically eliminates inaccurate estimates based on four states, namely: stable, recovery, alert, and uncertain. Each second the FSM framework evaluates its own state based on the estimated results and also evaluates the signal quality. A stable state indicates that the estimated HR is highly likely to be accurate and it is thus declared valid. A recovery state indicates that the estimated HR is somewhat likely to be accurate with the need to explore a possible transition to a stable state. An alert state indicates that the estimated HR is somewhat likely to be

inaccurate. An uncertain state indicates that the estimated HR is highly likely to be inaccurate. The FSM framework transits from one state to another every second in response to the estimation accuracy indicators, namely the crest factor (CF) and xthe HR change between consecutive windows. Details of the framework are presented in [6]. Accordingly, the FSM automatically validates the estimation results and ignores inaccurate estimation results caused by extremely low SNRs in rPPG signals. Table 4 shows that the AAE values decreased to 0.56 bpm and 0.89 bpm, respectively. However, the valid HR rate (VHR) (%) as a percentage of valid results among all the windows were 63.96% and 72.97%, indicating 36.04% and 27.03% of the estimated results were ignored. Fig. 7 compares the estimated HR results using our proposed method and when our method is combined with FSM framework.

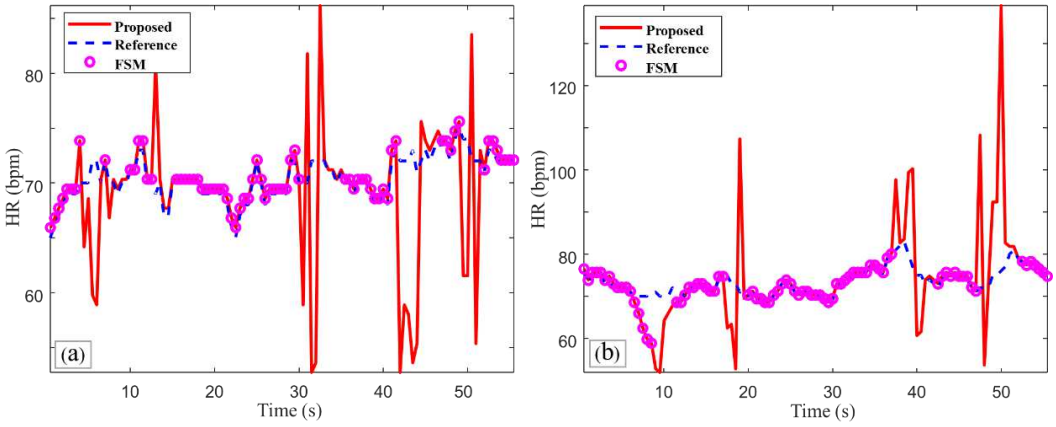

Figure 7. Comparison of the estimated HR results using our proposed method and when our method is combined with the FSM framework. HR estimated using the proposed method when noise is detected and following noise removal through the FSM: (a) under the rapid head movement condition, (b) under changes in ambient lighting conditions.

When our proposed algorithm is combined with the FSM framework, the PTOS was 276 ms, which is still achievable in real-time, as summarized in Table 5. However, the FSM framework discards the estimation results, which may prevent acquisition of continuous HR-related physiological information such as for HRV analysis and atrial fibrillation diagnosis. In future studies, we will further investigate how we can improve acquisitions in the presence of fast head movements and under variable ambient light conditions without any loss of instantaneous HR information. In addition, we have shown the simultaneous HR estimation from multiple subjects. In the future work, we will continue to investigate the dynamic issues such as occlusion and ID assignment with tracking.

Table 5. Computation time for each stage including FSM framework.

Types		Proposed method			
	face detection	face skin extraction	rPPG acquisition	HR estimation	FSM
Cycle time (ms)	8.11	1.02	0.46	0.72	1.12
Cycle Counts during one second	30	30	1	1	1
PTOS (ms)	243.3	30.6	0.46	0.72	1.12

Acknowledgements

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning (grants 202012B04 and NRF-2020R1A2C1014829)

Author contributions

HL, HC and HK developed the robot navigation system and remote PPG sensors. YN and SH performed the validation of the simulation and its results. JL conceived of the study, participated in the study's design and coordination and wrote the initial manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data Availability

All data generated or analyzed during this study are included in this published article.

References

1 Scully, C. G. *et al.* Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering* **59**, 303-306 (2011).

2 Abay, T. Y. & Kyriacou, P. A. Reflectance photoplethysmography as noninvasive monitoring of tissue blood perfusion. *IEEE Transactions on Biomedical Engineering* **62**, 2187-2195 (2015).

3 Kasbekar, R. S. & Mendelson, Y. Evaluation of key design parameters for mitigating motion artefact in the mobile reflectance PPG signal to improve estimation of arterial oxygenation. *Physiological measurement* **39**, 075008 (2018).

4 Chung, H., Ko, H., Lee, H. & Lee, J. Deep Learning for Heart Rate Estimation From Reflectance Photoplethysmography With Acceleration Power Spectrum and Acceleration Intensity. *IEEE Access* **8**, 63390-63402 (2020).

5 Lee, J., Chung, H. & Lee, H. Multi-mode particle filtering methods for heart rate estimation from wearable photoplethysmography. *IEEE Transactions on Biomedical Engineering* **66**, 2789-2799 (2019).

6 Chung, H., Lee, H. & Lee, J. Finite state machine framework for instantaneous heart rate validation using wearable photoplethysmography during intensive exercise. *IEEE journal of biomedical and health informatics* **23**, 1595-1606 (2018).

7 Benedetto, S. *et al.* Remote heart rate monitoring-Assessment of the Facereader rPPg by Noldus. *PloS one* **14**, e0225592 (2019).

8 Artemyev, M., Churikova, M., Grinenko, M. & Perepelkina, O. Robust algorithm for remote photoplethysmography in realistic conditions. *Digital Signal Processing* **104**, 102737 (2020).

9 Laurie, J., Higgins, N., Peynot, T. & Roberts, J. Dedicated exposure control for remote photoplethysmography. *IEEE Access* **8**, 116642-116652 (2020).

10 Rouast, P. V., Adam, M. T., Chiong, R., Cornforth, D. & Lux, E. Remote heart rate measurement using low-cost RGB face video: a technical literature review. *Frontiers of Computer Science* **12**, 858-872 (2018).

11 Macwan, R., Benezeth, Y. & Mansouri, A. Remote photoplethysmography with constrained ICA using periodicity and chrominance constraints. *Biomedical engineering online* **17**, 1-22 (2018).

12 Chaichulee, S. *et al.* Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 266-272 (2017).

13 Bousefsaf, F., Maaoui, C. & Pruski, A. Automatic selection of webcam photoplethysmographic pixels based on lightness criteria. *Journal of Medical and Biological Engineering* **37**, 374-385 (2017).

14 Fouad, R., Omer, O. A. & Aly, M. H. Optimizing remote photoplethysmography using adaptive skin segmentation for real-time heart rate monitoring. *IEEE Access* **7**, 76513-76528 (2019).

15 Poh, M.-Z., McDuff, D. J. & Picard, R. W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express* **18**, 10762-10774 (2010).

16 Durrant-Whyte, H. & Bailey, T. Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine* **13**, 99-110 (2006).

17 DeSouza, G. N. & Kak, A. C. Vision for mobile robot navigation: A survey. *IEEE transactions on pattern analysis and machine intelligence* **24**, 237-267 (2002).

18 Viola, P. & Jones, M. J. Robust real-time face detection. *International journal of computer vision* **57**, 137-154 (2004).

19 Bulat, A. & Tzimiropoulos, G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). *Proceedings of the IEEE International Conference on Computer Vision*, 1021-1030 (2017).

20 Li, X., Chen, J., Zhao, G. & Pietikainen, M. Remote heart rate measurement from face videos under realistic situations. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4264-4271 (2014).

21 Niu, X., Han, H., Shan, S. & Chen, X. Continuous heart rate measurement from face: A robust rppg approach with distribution learning. *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 642-650 (2017).

22 Bailey, T. & Durrant-Whyte, H. Simultaneous localization and mapping (SLAM): Part II. *IEEE robotics & automation magazine* **13**, 108-117 (2006).

23 Montemerlo, M., Thrun, S., Koller, D. & Wegbreit, B. FastSLAM: A factored solution to the simultaneous localization and mapping problem. *Aaai/iaai* **593598** (2002).

24 Thrun, S., Fox, D., Burgard, W. & Dellaert, F. Robust Monte Carlo localization for mobile robots. *Artificial intelligence* **128**, 99-141 (2001).

25 Fox, D. KLD-sampling: Adaptive particle filters. *Advances in neural information processing systems*, 713-720 (2002).

26 Pfaff, P., Burgard, W. & Fox, D. Robust monte-carlo localization using adaptive likelihood models. *European robotics symposium 2006*, 181-194 (2006).

27 Fox, D., Burgard, W. & Thrun, S. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine* **4**, 23-33 (1997).

28 Zhang, S. *et al.* S3fd: Single shot scale-invariant face detector. *Proceedings of the IEEE international conference on computer vision*, 192-201 (2017).

29 Yang, S., Luo, P., Loy, C.-C. & Tang, X. Wider face: A face detection benchmark. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5525-5533 (2016).

30 Zhu, X. & Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. *2012 IEEE conference on computer vision and pattern recognition*, 2879-2886 (2012).

31 Yan, J., Zhang, X., Lei, Z. & Li, S. Z. Face detection by structural models. *Image and Vision Computing* **32**, 790-799 (2014).

32 Jain, V. & Learned-Miller, E. Fddb: A benchmark for face detection in unconstrained settings. 5 (UMass Amherst technical report, 2010).

33 Lee, H., Chung, H. & Lee, J. Motion artifact cancellation in wearable photoplethysmography using gyroscope. *IEEE Sensors Journal* **19**, 1166-1175 (2018).

34 Wang, W., den Brinker, A. C., Stuijk, S. & de Haan, G. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering* **64**, 1479-1491 (2016).

35 De Haan, G. & Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering* **60**, 2878-2886 (2013).

36 Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O. & Karjalainen, P. A. Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine* **113**, 210-220 (2014).

37 Lee, J., Nam, Y., McManus, D. D. & Chon, K. H. Time-varying coherence function for atrial fibrillation detection. *IEEE Transactions on Biomedical Engineering* **60**, 2783-2793 (2013).

38 Chung, H., Lee, H., Kim, C., Hong, S. & Lee, J. Patient-provider interaction system for efficient home-based cardiac rehabilitation exercise. *IEEE Access* **7**, 14611-14622 (2019).

39 Lee, H. *et al.* Dedicated cardiac rehabilitation wearable sensor and its clinical potential. *PloS one* **12**, e0187108 (2017).

40 Boccignone, G. *et al.* An Open Framework for Remote-PPG Methods and their Assessment. *IEEE Access* **8**, 216083-216103 (2020).

41 Tanaka, H., Monahan, K. D. & Seals, D. R. Age-predicted maximal heart rate revisited. *Journal of the american college of cardiology* **37**, 153-156 (2001).

42 Gellish, R. L. *et al.* Longitudinal modeling of the relationship between age and maximal heart rate. *Medicine and science in sports and exercise* **39**, 822-829 (2007).

43 Chung, H., Lee, H. & Lee, J. State-dependent Gaussian kernel-based power spectrum modification for accurate instantaneous heart rate estimation. *PloS one* **14**, e0215014 (2019).

44 Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A. & Dubois, J. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters* **124**, 82-90 (2019).

45 Majos, E. & Dabrowski, R. Significance and management strategies for patients with asymptomatic atrial fibrillation. *Journal of atrial fibrillation* **7**, 1169 (2015).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- rppgtwopeople.mp4
- Supplementarylegends.docx