Non-Contact Physiological Parameters Extraction Using Facial Video Considering Illumination, Motion, Movement and Vibration

Hamidur Rahman, Mobyen Uddin Ahmed, and Shahina Begum, *Member, IEEE*

***Abstract*—*Objective:* In this paper, four physiological parameters, i.e., heart rate (HR), inter-beat-interval (IBI), heart rate variability (HRV), and oxygen saturation (SpO2), are extracted from facial video recordings. *Methods:* Facial videos were recorded for 10 min each in 30 test subjects while driving a simulator. Four regions of interest (ROIs) are automatically selected in each facial image frame based on 66 facial landmarks. Red-green-blue color signals are extracted from the ROIs and four physiological parameters are extracted from the color signals. For the evaluation, physiological parameters are also recorded simultaneously using a traditional sensor "cStress," which is attached to hands and fingers of test subjects. *Results:* The Bland Altman plots show 95% agreement between the camera system and "cStress" with the highest correlation coefficient $R = 0.96$ for both HR and SpO2. The quality index is estimated for IBI considering 100 ms R-peak error; the accumulated percentage achieved is 97.5%. HRV features in both time and frequency domains are compared and the highest correlation coefficient achieved is 0.93. One-way analysis of variance test shows that there are no statistically significant differences between the measurements by camera and reference sensors. *Conclusion:* These results present high degrees of accuracy of HR, IBI, HRV, and SpO2 extraction from facial image sequences. *Significance:* The proposed non-contact approach could broaden the dimensionality of physiological parameters extraction using cameras. This proposed method could be applied for driver monitoring application under realistic conditions, i.e., illumination, motion, movement, and vibration.**

***Index Terms*—Ambient illumination, driver monitoring, motion, movement, non-contact, physiological parameters, vibration.**

I. Introduction

DETECTION of the cardio-vascular pulse wave traveling through the body is referred to as plethysmography. The word 'photoplethysmography' (PPG) is widely used in camera system and composed of two words, 'photo' and 'plethysmography', where the word 'photo' refers to the use of light and 'plethysmography' comes from the Greek word 'plethysmos', which means increase [1]. For the camera system, the term PPG is known as remote PPG which was first introduced by Hertzman [2]. The principle concept of camera- based, non-contact technology is that ambient light reflected from the skin is modulated based on the absorption spectrum of heamoglobin. In the cardiovascular system heart pumps blood, and during each heartbeat, new blood flow occurs throughout the body through blood vessels. This blood circulation creates minute color variations in skin called micro-blushes, which cannot be seen by human eyes [3].

Physiological parameters are extracted from the color variation of the facial skin in lab situations, i.e., in sitting positions without considering any movements [4]–[15]. The two factors that affect the color values are blood volume variations caused by the cardiac pulse and (temporal) environmental illumination variations during the video recording [16]. Another way of introducing artifacts is motion, including movement and vibration. Motion artifacts are introduced through several actions, such as the variation of region of interest (ROI) in face region due to arbitrary motion of head as well as changes in facial expression [17]. Two types of motion artifacts, random motion artifacts and periodic motion artifacts, are imposed in image sequences due to motion, movement and vibration. Existing non-contact methods are applying RGB color facial video, which is suitable for laboratory environments or where the illumination of source light is constant [18]–[20].

In this article, four physiological parameters, HR, IBI, HRV and SpO2, are extracted in driving situations under challenging conditions of illumination variation, motion, movement and vibration. This proposed study is part of a national project named 'SafeDriver: A Real Time Driver's State Monitoring and Prediction System', in which one of the goals was to monitor drivers in real time using a non-contact-based approach, i.e., SafeDriver camera. Again, in driving situations, the extraction of physiological parameters based on facial images is challenging because the images can be contaminated easily due to motion, movement, vibration and illumination compared to the lab settings. Therefore, this study focuses on a physiological parameter extraction approach based on facial images under a variety of challenging situations. One RGB camera and one IR camera have been used simultaneously to record facial video in a simulator where the subject drives a simulated truck. Four regions of interest are automatically detected in each frame of the subject's

Manuscript received May 8, 2018; revised October 7, 2018, January 5, 2019, and March 8, 2019; accepted March 25, 2019. Date of publication May 15, 2019; date of current version December 23, 2019. *(Corresponding author: Hamidur Rahman.)*

H. Rahman is with the School of Innovation, Design and Engineering, Malardalen University, 722 20 Västerås, Sweden (e-mail: hamidur.rahman@mdh.se).

M. U. Ahmed and S. Begum are with the School of Innovation, Design and Engineering, Malardalen University.

Digital Object Identifier 10.1109/TBME.2019.2908349

facial image based on 66 facial landmarks. The colour signals are extracted from four ROIs simultaneously for each image frame under different illumination conditions. To remove illumination artefacts, RGB signals are converted into Lab colour space, and for motion artefacts caused by motion, movement and vibration, a modified least mean square (mLMS) adaptive filter is used. The results show that the proposed method achieves high degrees of accuracy with respect to reference measurements.

The rest of the paper organizes as follows: Section I.A describes state-of-the-art literature study, Section II focuses on materials and methods, Section III presents results and statistical significance test is added in Section IV. Then a discussion is presented in Section V and finally conclusion is drawn in Section VI.

A. State-of-the-Art

For a camera-based system, an ROI is selected from any part of the body that is not occluded. RGB colour signals are generated from each ROI, and physiological parameters are extracted from the colour signals. The literature shows that the ROI is selected from different parts of the body, such as facial skin [10], [14], [16], [18], [20]–[24], skin of hand or neck [25]–[27], finger [11], [28], [29] and eyes [30], [31]. However, facial image is widely used as an input for camera-based physiological parameter systems because it is convenient and flexible. In 1995, the first remote, non-contact pulse oximetry and camera images were investigated for monitoring health using only a colour camera to measure remote PPG signals [27]. But their approach does not report quantitative results; only a graph of heartbeats is presented but no correlation with reference signals (i.e., simultaneous ECG) are reported. Again, in 2005, another method was introduced for the measurement of the computer user's emotional state using the facial thermal image using a thermal camera [32]. Here, the experiment was conducted by 12 users, and the authors found an interesting result: user stress is correlated with increased blood flow in the frontal vessel of the forehead.

In 2006, Takano *et al*. showed that RR and HR can be acquired simultaneously using a digital CCD camera with normal ambient light [24]. Here, the images of a part of the subject's facial skin are captured consecutively, and the changes in the average image brightness of the region of interest (ROI) are measured for a short time. Their system can detect HR for a certain period of time, but the efficiency is unknown. However, the term 'brightness' in the article is confusing because physiological parameters are extracted from colour variation in ROI instead of brightness. Later, in 2007, Garbey *et al*. developed a contact-free measurement of cardiac pulse based on the analysis of thermal images [26]. In their article, ROI is selected from facial and hand skin. The authors used the temperature of the vessel, which is modulated by pulsative blood flow and directed to the recovery of frequency of the component signal with the highest energy content [26]. Again, in 2008, by Verkruysse *et al*. [1], a digital camera was used to extract HR and RR from facial video recorded in ambient light. Their system is able to extract HR and RR from 30 seconds to a few minutes. However, they did not consider illumination variation in their experiment. In 2009, Banitsas *et al*. used smart phones to record PPG signals from users' fingers and were able to extract HR [33]. In other experiments, HR and RR were extracted by a smart phone camera using a user's finger [6], [11], [18].

In recent years, camera-based systems have been further improved. In 2011, Poh *et al*. at the Affective Computing Research Group in MIT proposed an algorithm to extract underlying source signals from the R, G, and B colour bands [22]. They used ICA for blind source separation and extracted HR and HRV from 1-minute colour facial video recording under a normal sitting position. In the same year, several other experiments took place, such as [16] and [34]. However, state-of-the-art methods are good for laboratory applications in normal sitting positions. In real applications, e.g., driver monitoring, environmental illumination changes frequently, and dynamic illuminance changes can cause artefacts in colour signals. A motivation has been observed in 2014 by Xiaobi *et al*. to reduce environmental illumination artefacts from RGB colour signals using a normalized mean least square (NLMS) adaptive filter [16]. They achieved good correlation and agreement between the proposed method and the reference sensor under illumination variations. However, no motivation is observed regarding how their proposed method can be used taking into account movement and vibration. Jeanne *et al*. presented an IR camera-based heart rate monitoring system. The system has been developed for laboratory settings without considering the dynamic light environment, vibration and motion [3].

Non-contact camera-based systems need to consider another important issue: motion artefacts caused by motion, movement and vibration. In the literature, several techniques are used for removing motion artefacts from images, such as periodic moving-average filter methods, FFT methods, SVD methods, wavelet denoising methods and adaptive filtering methods [35]. A spatial restoration algorithm was introduced using a non-linear interpolation technique for eliminating motion artefacts [36]. This algorithm preserves the edges surrounding the artefact area of the image sequences and is mostly used in biomedical applications. The Authors in [37] presented an algorithm using convolutional neural network (CNN) to detect motion artefacts automatically in MRI images. Recently some works [16], [17], [38], [39], considering motion artefacts have been reported to extract physiological parameters using cameras without taking illumination and vibration into consideration.

Several motivations are also observed to extract physiological parameters from driver's facial video recordings. In 2014, Zhang *et al*. developed a webcam-based noncontact system for monitoring the physiological information of drivers using a simulator [40]. However, the authors did not explain how motion artefacts and illumination variation have been considered. Additionally, their results do not show correlation and agreement with the reference sensor. In the same year, Guo *et al*. showed a similar approach to monitoring drivers' HRV continuously under real world driving circumstances [20]. The authors applied the independent vector analysis (IVA) technique to extract HR from driver's facial video, but they failed to overcome motion and illumination artefacts. Again in [41], the authors presented

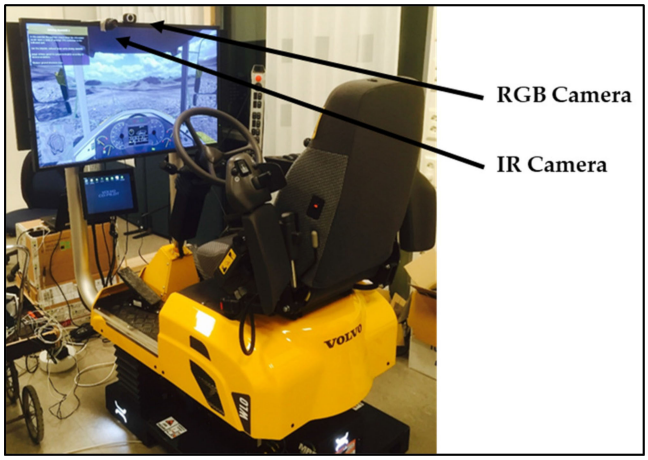

Fig. 1. Volvo Constriction Equipment Simulator used for data collection.

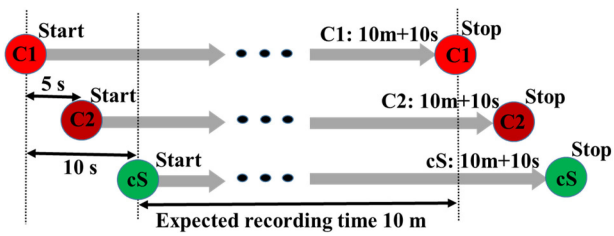

Fig. 2. Time Synchronization of RGB camera (C1), IR camera (C2) and cStress sensor (cS).

a multiset canonical correlation analysis (MCCA) approach to extract HR and HRV from drivers' facial videos both in laboratory setting and real road driving. They have shown that their approach achieved better performance than the ICA-based approach taking motion artefacts into account. However, the authors did not take into account illumination or vibration.

II. Materials and Methods

A. Data Collection

Thirty subjects aged between 20 and 50 years including 24 males and 6 females with mixed skin colours participated in the study. All participants were healthy without any sleeping and motion sickness problems. They also agreed to avoid alcohol for 72 hours before arrival and to not smoke or consume caffeine for 3 hours before the experiment. The experiment was conducted using a Volvo construction equipment (VCE) simulator[1] which is generally used for bucket filling of construction sites, as shown in Fig. 1.

Due to the driving simulator, omnidirectional motion and vibration occur while driving. Additionally, the illumination of the environment was changed manually by switching the room light 'on' and 'off', and the amount of light was recorded by a lux meter.[2] A downhill scenario in a deconstruction environment was selected, as it simulated heavy movement and vibration.

The total experimental session was approximately thirty minutes long, where the first ten minutes was used for experimental setup, i.e., signing the consent letter, giving instructions about the experiment and attaching reference sensors. In the second ten minutes, each subject became familiar with the simulator, i.e., driving a truck, and the last ten minutes consisted of the actual driving experiment. In the actual experiment, for the first five minutes, the facial video of the driver was recorded under light conditions (i.e., light is on), and for the last five minutes, the video was recorded in dark conditions (i.e., light is off). Here, the highest amount of light was 340 lux when the light was on, and the lowest amount of light was 35 lux when the light was off. Additionally, 'vibSensor'[3] was used to record the vibration of the simulator. The average vibration of the simulator was observed as 5 Hz during data collection. Notably, the vibration range of a passenger car is 1-3 Hz under normal road conditions [42].

To record the facial video of the drivers, two different cameras were used: an RGB camera (Logitech HD webcam C525) and an IR camera (integrated illumination 850 nm including bandpass filter to block daylight, field of view is approx. 30 degrees), which are placed on top of the simulator display. The frame rate for both cameras is 30 fps (frame per second). In total, for 30 test subjects, 60 video films are recorded, each 10 minutes long; 30 recorded videos using an RGB camera and 30 recorded videos using an IR camera are obtained for facial video recording. For the evaluation, a reference sensor system, 'cStress'[4] was attached with the driver's body, and physiological parameters, HR, IBI and SpO2, were recorded during driving. The two cameras and the reference sensors were connected to a single HP laptop computer (Intel Core i7-4810 MQ, CPU 2.8 GHz, Windows 7, OS 64 bit), and were synchronized manually. Three different desktop applications for the two cameras and the sensor system were opened to record the respective video and physiological parameters. An oral 'START' command was given to the driver to start driving, and immediately, the 'video record' button of the RGB camera (C1) was clicked on to record the facial video of the driver. Then, after five seconds, the 'video record' button of the IR camera was clicked on, and after ten seconds, the 'record' button of the cStress sensor was clicked on to record HR, IBI and SpO2. The driver was to continue the driving task until he was given a 'STOP' command. The 'stop' button of each camera and cStress sensor was clicked 10 minutes and 10 seconds after clicking the 'start' button. Therefore, the duration of each recording by RGB camera, IR camera and cStress sensor was 10 minutes and 10 seconds. To synchronize the time for all the systems, this extra 10 seconds was deleted from each recording data: the first 10 seconds of the video recording from the RGB camera, the first and last 5 seconds of the video recording from the IR camera and the last 10 seconds of the cStress data.

[1] https://www.volvoce.com/united-states/en-us/services/volvo-services/productivity-services/training-simulators/

[2] https://www.clasohlson.com/uk/Lux-Meter/32-7361

[3] https://itunes.apple.com/us/app/vibsensor-accelerometer-recorder-vibration-analysis/id932854520?mt=8

[4] http://stressmedicin.se/neuro-psykofysilogiska-matsystem/cstress-matsystem/

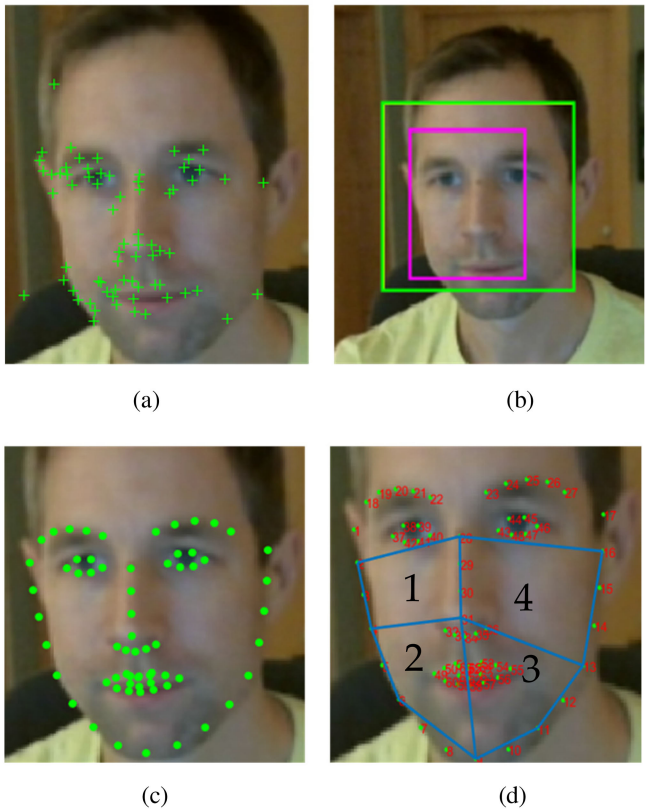

Fig. 3. (a) Facial features selection for *M1* (b) ROI selection for method *M1* (c) facial landmarks selection for M2 (d) four ROI selection for M2.

B. ROI Selection

Traditionally, facial regions from the entire image are detected using the Viola and Jones algorithm [43]. This algorithm uses adaBoost classifier to select a small number of visual features shown in Fig. 3(a). In the first image frame, a boosted cascade classifier is used to identify the coordinates of the face location. A rectangular box is plotted to define facial region using the coordinates of facial region. However, the box contains little non-facial skin, which is unnecessary for camera-based physiological parameter extraction techniques. Therefore, the ultimate ROI is further defined by another rectangular box inside the first box, which is equivalent to 60% of the width and 80% of the height of the first box shown in Fig. 3(b). ROI is detected automatically for every image frame, which is computationally expensive. To speed up the ROI selection, the face is tracked in every image frame using the Kanade-Lucas-Tomasi (KLT) face tracking algorithm. Facial tracking is performed based on selected features in the first image frame. Here, MATLAB function 'detectMinEigenFeatures' is used for feature selection [44].

In traditional methods, facial features are selected based on edges and corners in facial regions. The number of selected features and their locations vary in each frame due to motion, movement and vibration while video recording. In this paper, a multiple ROI selection method is proposed. First, 66 facial landmarks are selected that are pre-defined facial key points. These landmarks are automatically detected in the image frame using the DRMF method [45] and are numbered from 1 to 66 shown in Fig. 3(c). Then, four ROIs are detected where the first ROI is selected using landmarks 2, 4, 31 and 28; the second ROI is selected using landmarks 4, 6, 9 and 31; the third ROI is selected using landmarks 9, 11, 13 and 31 and the fourth ROI is selected using landmarks 13, 16, 28 and 31 which is shown in Fig. 3(d). In each frame, average RGB values are extracted from all four ROIs simultaneously, which are used as sources of color signals. However, depending on the face position due to head movement to the left/right, one or two ROIs are ignored. Note that throughout this article, the traditional method for selecting one ROI is denoted by *M1* and four ROI selection using facial landmarks method is denoted as *M2*.

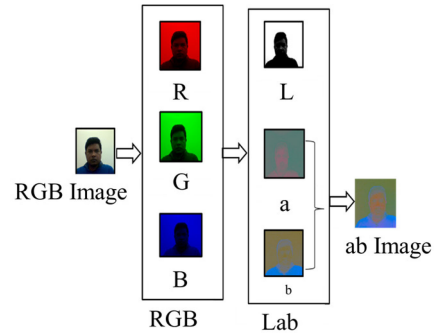

Fig. 4. Transformation RGB color space into Lab color space.

C. Illumination, Motion, Movement and Vibration

Considering illumination, we have applied the concept of our previous article [46] in which RGB color signals were converted into Lab color space. The Lab color space consists of three independent signals, L, a and b, where L represents the lightness of the images and a (red/green) and b (yellow/blue) represent the combination of other color channels. These two-color channels have been separated from the Lab color space. This ab signal is a pure color signal, and physiological parameters have been extracted using this signal. This transformation of RGB into Lab color space is visualized in Fig. 4.

Due to motion and movement, random motion artifacts are imposed into images and produce blurry images. As a result, facial landmarks selection is hindered due to blurry image and ROI detection becomes uncertain. In this article, motion artifacts have been reduced in two ways. Firstly, the blurry image is reconstructed using the wiener restoration filter G using equation 1 below [47].

$$G\left(\mathrm{k},1\right)=\frac{\mathrm{H}^{*}\left(\mathrm{k},1\right)\mathrm{S}_{x}\left(\mathrm{k},1\right)}{\left|\mathrm{H}\left(\mathrm{k},1\right)\right|^{2}\mathrm{S}_{x}\left(\mathrm{k},1\right)+\mathrm{S}_{u}\left(\mathrm{k},1\right)} \tag{1}$$

Where S_x and S_u are the power spectrum and noise respectively for the pixel (k, l) where k and l indicate row and column of the image frame. H is the transfer function of the impulse response of the fraction of $\frac{S_x}{S_u}$ having a positive time solution under the inverse Laplace transform. G can filter out the noise from the blurry image frame to provide an estimate of the underlying image of interest. The algorithm is optimal in a sense of least

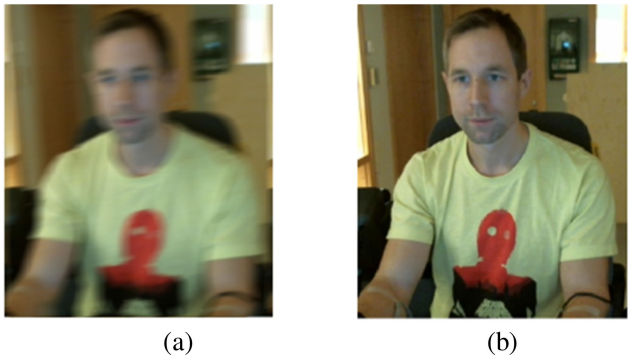

Fig. 5. (a) Blur image (b) Reconstructed image.

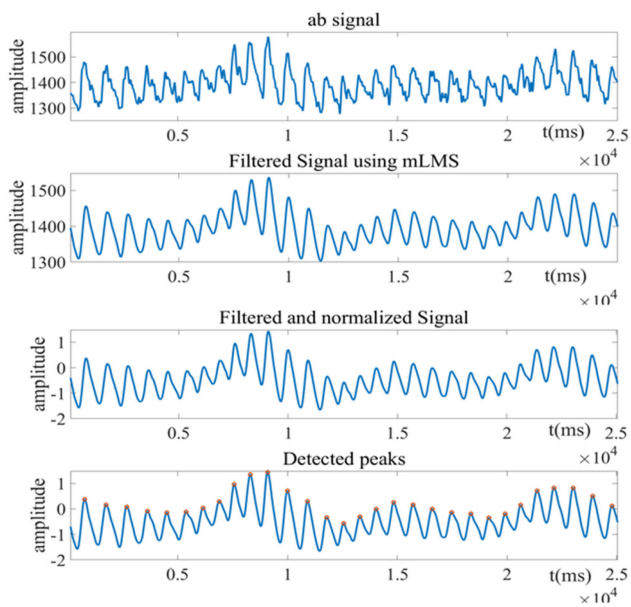

Fig. 6. Signal Processing: (a) ab signal with motion artifacts (b) reconstructed signal using mLMS filter (c) normalized signal and detected peaks.

mean square error (MSE) between the estimated and the true images. The blurry image and reconstructed image are shown in Fig. 5.

Due to vibration, the color signals still contain a small amount of motion artifacts. Because HR is extracted based on detected major peaks, it is essential to detect all true peaks from the ab signal. In this paper, modified LMS (mLMS) adaptive filter is used, as it shows better performance than other adaptive filters [48]. This filter has the ability to adjust its impulse response to filter out the motion artifact from the signal. The motion artifacts due to vibration and the filtered signal is shown in Fig. 6.

D. Parameter Extraction

Four physiological parameters i.e., HR, IBI, HRV and SpO2 are extracted from facial video recordings by both RGB and IR camera. HR, IBI and HRV are extracted based on detected peaks from color signals. Fig. 7 shows flowchart of HR, IBI and HRV extractions workflow.

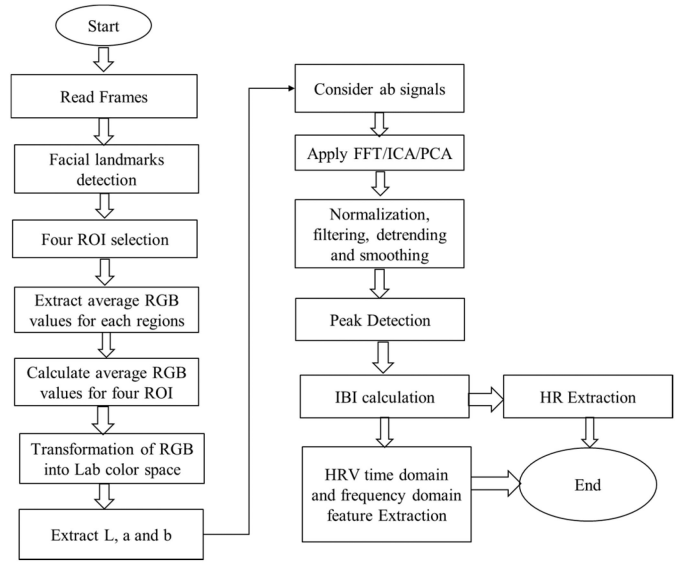

Fig. 7. HR, IBI and HRV extraction workflow.

The filtered ab signals in Fig. 6 are normalized for peak detection. All major peaks are detected using MATLAB function 'findpeaks'. Here, each peak represents one heart beat denoted by small red circle. IBI is time intervals between two consecutive peaks and IBI time series is generated using all time intervals of consecutive peaks. HR is calculated based on the number of heart beats per minute. HRV features both in time domain and frequency domain are calculated using IBI time series which is discussed elaborately in our previous article. The calculation procedures of HR, IBI and HRV are presented in our previous articles [49], [50] and [51] respectively.

Oxygen Saturation is extracted according to the following equation used in [4].

$$\mathrm{S_pO_2} = \mathrm{A} - \mathrm{B}\begin{pmatrix}\mathrm{SD_{red}}\\ \mathrm{M_{red}}\\ \mathrm{SD_{blue}}\\ \mathrm{M_{blue}}\end{pmatrix} = A - B\begin{pmatrix}\mathrm{SD_{red}}\\ \mathrm{SD_{blue}}\end{pmatrix}\cdot\begin{pmatrix}\mathrm{M_{blue}}\\ \mathrm{M_{red}}\end{pmatrix} \tag{2}$$

Here,

SD_{red} = Standard deviation of red color values of image frames;

SD_{blue} = Standard deviation blue color color values of image frames;

M_{red} = mean value of red color values of image frames;

M_{blue} = mean value of blue color values of image frames.

A and B are linear value coefficients.

E. Approach and Methods for Evaluation

HR, IBI, HRV and SpO2 are extracted for 30 test subjects from both the RGB and IR camera videos and compared with cStress measurements. Here, 8 out of the 10 minutes of sample data are considered for the experiment, where the first and last minute of sample data are excluded to obtain homogeneous data sets. Data sets are *divided into two environments*: *E1*, an environment with higher illumination, and *E2*, an environment

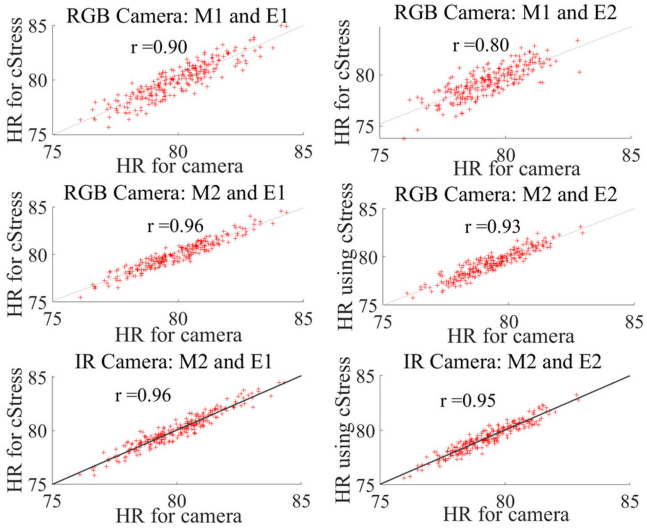

Fig. 8. Scatter plot and correlation co-efficient E1: with lower light illumination (left) and E2: with higher light illumination (right).

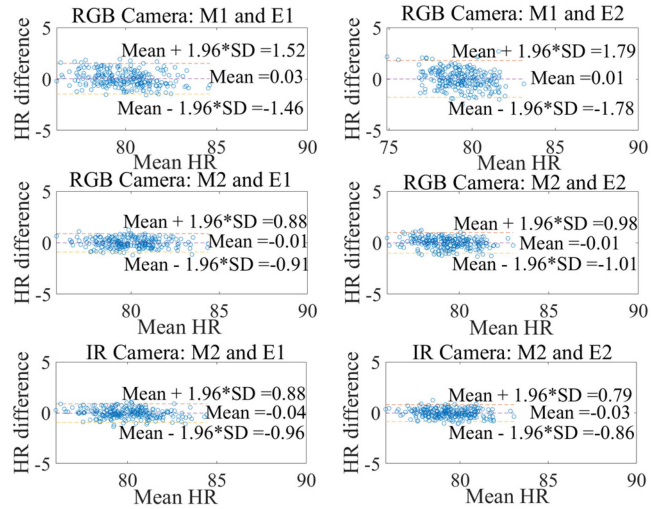

Fig. 9. Bland Altman plot for HR, E1: with lower light illumination (left) and E2: with higher light illumination (right).

with lower illumination. Additionally, data in both environments were contaminated with other complex factors, e.g., motion, movement and vibration. *Two methods* have been used to compare: *M1*, a method with a single ROI selection approach (i.e., State-of-the-art), and *M2*, a method with multiple ROI selection approach (i.e., proposed). Several statistical parameters such as mean error (ME), root mean squared error (RMSS) and standard deviation of error (SDE), are calculated comparing methods *M1* and *M2*, environments E1 and E2 and cameras RGB and IR. Again, plots are presented based on the correlation coefficient and Bland Altman to compare the reference sensor and camera systems using the two methods. IBI is evaluated using a Poincare' plot to visualize significance between two systems [54]. The Poincaré plot is a type of nonlinear analysis used to quantify self-similarity based on IBIs versus the previous IBI. Thus, pairs of successive IBI form an attractor in the Poincare' plot.

III. Results

This section presents statistical and graphical evaluations on HR, IBI, HRV and SpO2 using two cameras, cStress, two environments and two methods.

A. HR

Correlation coefficient 'r' between the reference system cStress and the camera systems using the two methods are calculated for both environments and presented in Fig. 8. Here, considering both RGB and IR camera the correlation coefficient increases 6% for proposed *method M2* in E1 than state-of-the-art method. However, for RGB camera, the correlation coefficient decreases in E2 for both *methods M1 and M2*. Interestingly, correlation coefficients for IR camera are almost same for both methods in both environments.

To show the agreement between the two systems considering two methods and two environments the Bland Altman plot [52] is applied and are presented in Fig. 9. Here, x-axis represents mean HR of two systems and y-axis represents HR differences between two systems. For 95% agreement, first mean difference of two systems are calculated and then the upper and lower limit of agreements are defined as mean $\pm 1.96 \times$ SD where SD is standard deviation of HR difference.

TABLE I
Error Estimation of HR for RGB and IR Camera
(All Values Are BPM)

Camera	Statistical Parameter	*M1* E1	*M1* E2	*M2* E1	*M2* E2
RGB	*ME*	4.04	4.52	2.49	2.51
	RMSE	1.17	1.31	0.64	0.65
	SDE	4.2	4.7	2.57	2.59
	Mean + 1.96 × SD	1.52	1.79	0.88	0.98
	Mean	0.03	0.01	-0.01	-0.01
	Mean - 1.96 × SD	-1.46	-1.78	-0.91	-1.01
IR	*ME*	4.84	5.2	3.2	3.16
	RMSE	2.62	2.64	2.04	1.96
	SDE	5.52	5.85	3.8	3.73
	Mean + 1.96 × SD			0.88	0.79
	Mean			-0.04	-0.03
	Mean - 1.96 × SD			-0.96	-0.86

The summary of the Bland Altman plot is shown in Table I.

Again, Table I presents estimation of various errors such as mean error (ME), root mean squared error (RMSE) and standard division of error (SDE) comparing between RGB camera and cStress and IR camera and cStress for both methods *M1* and *M2* considering both environment E1 and E2. It is observed that ME, RMSE and SDE are decreased in each case which is about (40–50%).

B. IBI

To evaluate IBI, the R-peak quality index (QI) is calculated based on the cumulative frequency (i.e., time difference) of IBI between the cStress reference system and camera systems. QI is

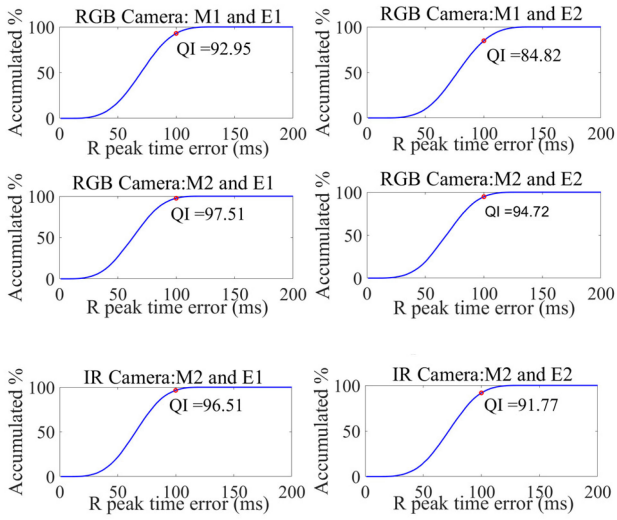

Fig. 10. Quality Index (%) for IBI considering two methods, two environments and two cameras.

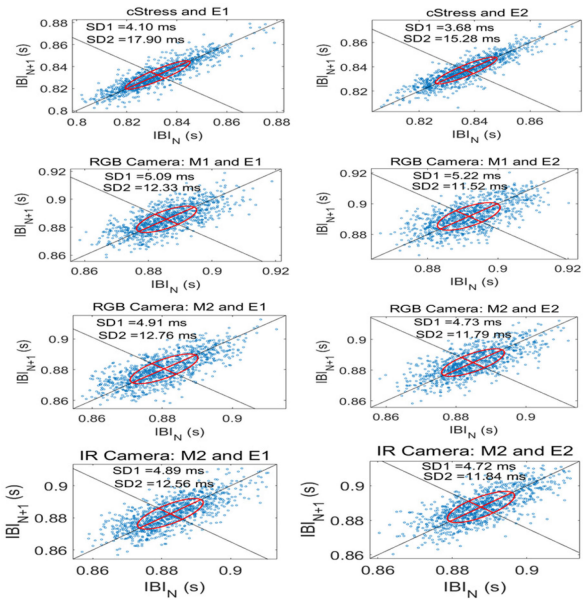

Fig. 11. Poincare' plot for IBI for RGB and IR camera in both environment.

presented as cumulative percentage, (i.e., cumulative frequency ÷ n) × 100, where n is the total number of observations [53]. Fig. 10 presents the QI for the two methods, the two environments and the two cameras. Here, the X-axis represents IBI differences (ms) of two systems, the y-axis represents the accumulated percentage and the QI is calculated using 100 ms as the R-peak error. It can be observed in Fig. 10 that the QI values increase with our proposed method (M2) for both environments; however, it slightly decreased using the IR camera.

Fig. 11 shows the Poincare' plot taking into account the environments, the two cameras, the two methods and cStress. Here, points above the line of identity indicate a longer IBI than the preceding IBI, and points below the line of identity indicate a shorter IBI than the preceding. Standard deviations along the line of identity (*SD2*) and perpendicular to the line of identity (*SD1*) represent the magnitude of the major and minor axes of the ellipse, respectively [54].

The absolute difference (SD1) between cStress and RGB using *method M1* in E1 is 0.99 ms and E2 is 1.54 ms; and *method M2* in E1 is 5.57 ms and E2 is 3.76 ms. *The absolute difference (SD2)* between cStress and RGB using *method M1* in E1 is 0.81 ms and E2 is 1.05 ms; and *method M2* in E1 is 5.14 ms and E2 is 3.49 ms. It can be observed considering both environments, *method M2* is closer to cStress for the value of *SD1* and *SD2*. Similarly, the absolute difference (SD1) between cStress and IR camera using *method M2* in E1 is 0.79 ms and E2 is 1.04 ms; and *absolute difference (SD2)* between cStress and IR camera using *method M2* in E1 is 5.34 ms and in E2 is 3.44 ms. The results indicate that *method M2* shows similar performance for IR camera.

TABLE II
HRV TIME DOMAIN FEATURES (ALL VALUES ARE MILLISECONDS (ms))

Time Domain Parameters		E1			E2		
		cStress	*M1*	*M2*	cStress	*M1*	*M2*
RGB	meanNN	832.34	885.7	880.5	837	891.7	884.9
	SDNN	52.74	44.06	44.34	50.16	42.96	43.11
	RMSSD	29.55	39.11	37.73	26.23	39.17	37.38
	SDSD	23.36	25.08	24.26	20.3	24.96	23.88
IR	meanNN	832.34	889.5	882.5	837	893.4	887.7
	SDNN	52.74	43.46	43.99	50.16	43.03	42.91
	RMSSD	29.55	39.18	37.65	26.23	39.22	37.72
	SDSD	23.36	25.29	24.3	20.3	25.02	24.16

TABLE III
HRV FREQUENCY DOMAIN FEATURES (ALL VALUES ARE MILLISECONDS (ms^2))

Freq. Domain Parameters		E1			E2		
		cStress	*M1*	*M2*	cStress	*M1*	*M2*
RGB	LF	79.51	89.66	88.64	80.9	91.23	89.87
	HF	9.28	10.27	10.16	9.14	10.18	10.05
	LF/HF	8.61	8.74	8.73	8.88	8.97	8.95
IR	LF	79.51	90.41	89.02	80.9	91.52	90.4
	HF	9.28	10.36	10.21	9.14	10.24	10.11
	LF/HF	8.61	8.74	8.73	8.88	8.94	8.95

C. HRV

HRV features in both the time and frequency domains are calculated for cStress, both methods, both environments and both camera systems. A detailed description of the HRV feature calculation is presented in our previous article [51]. Time domain features—1) meanNN = average of all beat-to-beat or NN intervals, 2) SDNN = standard deviations of all NN intervals, 3) RMSSD = root mean square of successive differences between adjacent NN intervals and 4) SDSD = standard deviation of successive differences between adjacent NN intervals—are calculated in millisecond (ms) and presented in Table II. Frequency domain features—1) LF = low frequency power (0.04–0.15 Hz), 2) HF = high frequency power (0.15–0.4 Hz) and 3) LF/HF Ratio = ratio between LF HF, are calculated and typically expressed in terms of "power", i.e., milliseconds squared (ms^2) and presented in Table III.

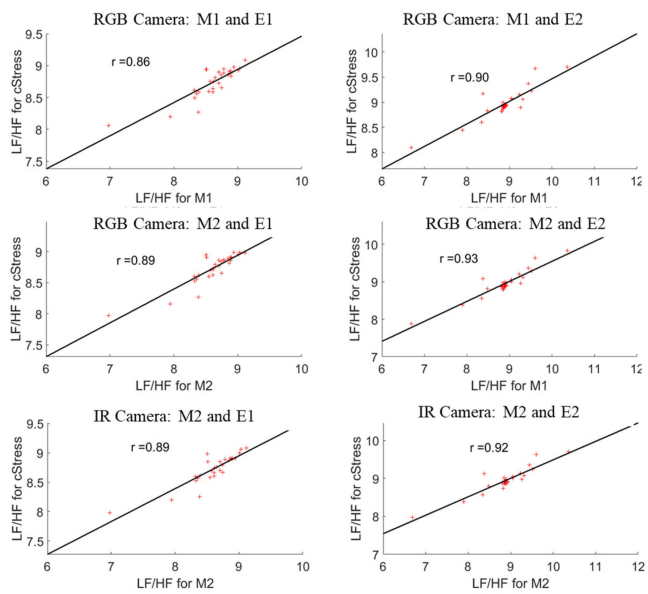

Fig. 12. Scatter plot of LF/HF and correlation co-efficient.

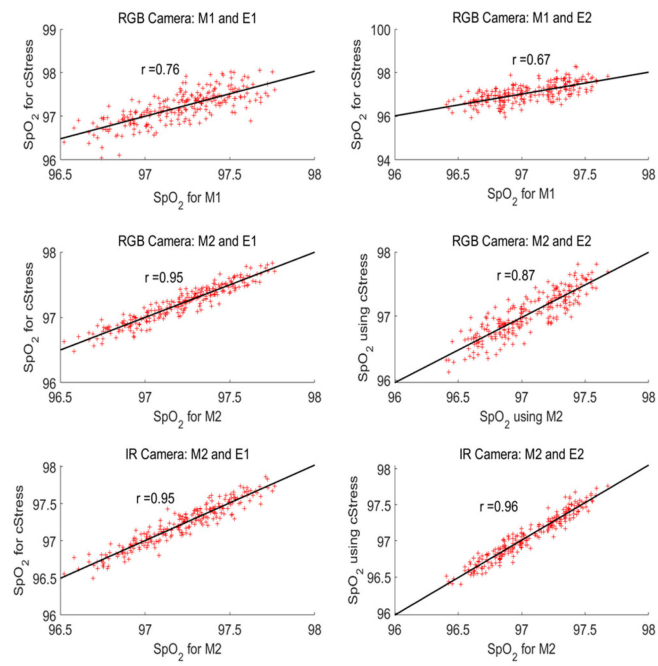

Fig. 13. Scatter plot and correlation co-efficient of SpO2 for both camera and environment.

The scatter plot and correlation coefficient for LF/HF determined for both cameras and cStress, both methods and both environments are shown in Fig. 12. Here, *method M2* shows better correlation coefficient than *method M1* in both environments. However, for the IR camera the correlation coefficient for method M2 is similar in E1 and 1% lower in E2 compared to the RGB camera.

D. SpO2

The Correlation coefficient 'r' between the reference system cStress and the camera systems using the two methods is calculated for both environments and presented in Fig. 13. Under the lower illumination condition, the correlation coefficient is significantly increased for the proposed method than the state-of-the-art method, which is which is similar to the IR camera system. Under the lower illumination in E2, the correlation coefficient decreases with methods 1 and 2 for the RGB camera and increases by 1% for the IR camera.

However, to show the agreement between the two systems, the Bland Altman plot is presented in Fig. 14 for both methods *M1* and *M2* in both environments. Here, the x-axis represents the mean SpO2 of the two systems, and the y-axis represents SpO2 differences between the two systems. For 95% agreement, the first mean difference of the two systems is calculated, and the upper and lower limits of the agreements are defined as the mean $\pm 1.96 \times$ SD, where SD is the standard deviation of the SpO2 difference. All the calculations are presented in Table IV.

Again, various errors such as ME, RMSE and SDE are calculated between cStress and both methods, which are presented in Table IV. In both environments, ME, RMSE and SDE decrease significantly for method M2. However, for the RGB camera, method M2 performs better in environment E1, and for IR camera, M2 performs better in environment E2.

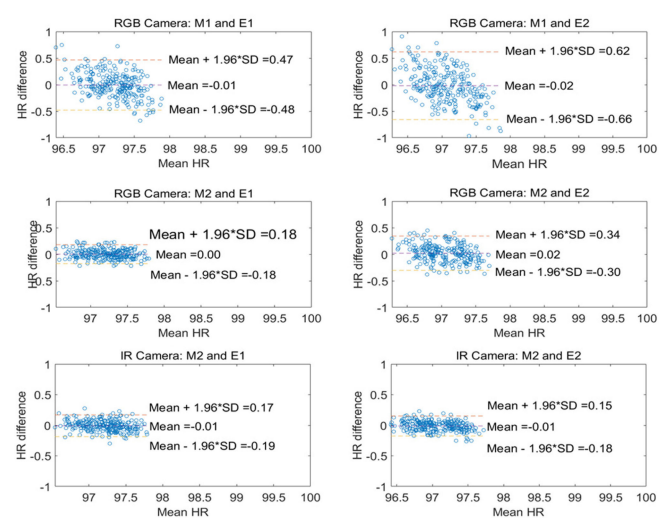

Fig. 14. Bland Altman plot for SpO2.

IV. Statistical Significance Test

A Statistical significance test, one-way analysis of variance (ANOVA1), is conducted for proposed method M2 comparing RGB and IR camera with respect to the cStress sensor. In the test, several statistical parameters—the sum of squares (SS) due to each source; degrees of freedom (df) associated with each source; mean squared error (MS), which is SS/df for each source of variation; F-statistic (F), the ratio of the mean squared errors; and probability of the significance test (P) are calculated for HR, IBI and SpO2 and presented in Table V. For HR, the results show that $P(F > 0.05) = 0.95$ in E1 and $P(F > 0.04) = 0.96$ in E2, which indicates that there are no significant differences between

TABLE IV
ERROR ESTIMATION OF SPO2 (ALL VALUES ARE %)

Camera	Statistical Parameter	*M1* E1	*M1* E2	*M2* E1	*M2* E2
RGB	ME	1	1.33	0.33	0.68
	RMSE	0.87	1.15	0.33	0.61
	SDE	1.32	1.76	0.47	0.92
	Mean + 1.96 × SD	0.47	0.62	0.18	0.34
	Mean	-0.01	-0.02	0.00	0.02
	Mean - 1.96 × SD	-0.48	-0.66	-0.18	-0.30
IR	ME	1.33	0.99	0.34	0.33
	RMSE	1.15	0.88	0.33	0.34
	SDE	1.75	1.32	0.47	0.47
	Mean + 1.96 × SD			0.17	0.15
	Mean			-0.01	-0.01
	Mean - 1.96 × SD			-0.19	-0.18

TABLE V
ANNOVA1 TEST RESULT FOR HR, IBI AND SPO2 (ALL VALUES ARE %)

Parameter	Environment	SS	df	MS	F	P>F
HR	E1	0.2	2	0.11	0.05	0.95
	E2	0.1	2	0.06	0.04	0.96
IBI	E1	2.1	2	1.06	0.01	0.99
	E2	9.2	2	4.62	0.04	0.96
SpO2	E1	0.1	2	0.01	0.09	0.92
	E2	0.1	2	0.01	0.09	0.51

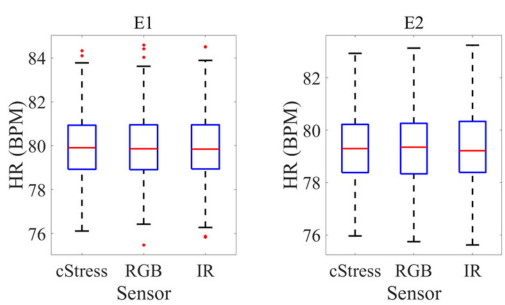

Fig. 15. ANOVA test for HR considering method M2 in E1 and E2.

cStress and camera systems (i.e., RGB and IR cameras) for method M2. Again, considering IBI and SpO2, the results show that there are no significant differences between the proposed method M2 and cStress.

For the visual comparison, box plots are drawn for method M2 between cStress, RGB and IR cameras in both environment E1 and E2. Fig. 15, Fig. 16 and Fig. 17 are presented below for HR, IBI and SpO2 respectively. For every box, the central mark indicates the median, the bottom edge indicates the 1st percentile (25%) and top edge indicates the 3rd percentile (75%). The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '.' symbol.

V. DISCUSSION

Considering the simulator environment in this study, the results obtained for four physiological parameters demonstrate

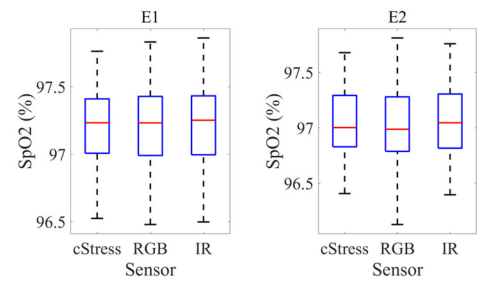

Fig. 16. ANOVA test for SpO2 considering method M2 in E1 and E2.

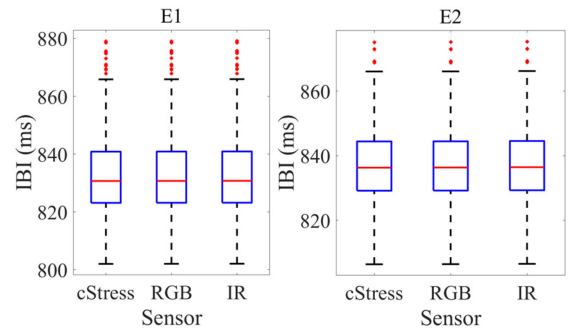

Fig. 17. ANOVA test for IBI considering method M2 in E1 and E2.

that there is a strong correlation and agreement between the proposed method and the reference cStress sensor. The correlation coefficient of HR for method M2 under environment E1 is 0.96 for the RGB and IR cameras; under environment E2, the correlation coefficients are 0.93 for the RGB camera and 0.95 for the IR camera (Fig. 8). To our knowledge, only a few attempts have been made to monitor non-contact HR using IR camera such as [26] and [12]. However, the correlation coefficient for their proposed method is unknown. For the RGB camera, several authors find a correlation coefficient for non-contact HR extraction in a normal sitting position. Although [22] finds a correlation coefficient of 1.0 for the normal sitting position, no correlation coefficient is reported taking into account illumination, motion, movement, vibration. In [22], RMSE for the lab situation is 1.24 and [16] finds a correlation coefficient of 1.27 taking illumination into account, whereas in our proposed method, the RMSE is 0.64 and 0.65 in environment E1 and E2, respectively (Table I).

Considering the agreement between the camera system and the reference sensor, [10] shows that the 95% limit of agreement is −4.55 to 4.44. In our case, the 95% limit of agreement is −0.91 to 0.88 (Table I), which represents an improvement to the state-of-the-art value. Again, IBI was evaluated using QI and Poincare' plots. Considering environment E1, the obtained highest QI value is 97.51% for the RGB camera and 96.51% for the IR camera based on the threshold value, i.e., the difference between two consecutive QIs. (Fig. 10). The results show that the proposed method for both the RGB and the IR cameras provides better IBI values in higher illumination. Both QI and the Poincare' plot present similar results. However, we could

not compare our result for IBI with the state-of-the-art method, as, to the best of our knowledge, we did not find any literature to evaluate IBI using QI and a Poincare' plot.

Considering HRV, the literature shows that the correlation coefficient of LF/HF is 0.88 [22] in normal lab situations and 0.64 in driving situations [41]. In our case, the correlation coefficients for the RGB camera are 0.89 and 0.93 and for the IR camera are 0.89 and 0.92 in environments E1 and E2, respectively (Fig. 12). Considering SpO2, the correlation coefficients for the RGB camera are 0.95 and 0.87 and for the IR camera are 0.95 and 0.96 in E1 and E2, respectively (Fig. 13). The results show that the proposed method for the RGB camera works better in E1, whereas the IR camera shows better performance in E2.

VI. Conclusion

Four physiological parameters, HR, IBI, HRV and SpO2, are extracted in a simulator environment that takes into account illumination, motion, vibration and movement. To extract colour signals, four ROIs were selected based on 66 facial landmarks for each facial image. Subsequently, extracted physiological parameters are quantified and compared with the reference sensor by various statistical, graphical and geometrical methods both in light and dark conditions. Bland Altman plots are used to see the 95% agreement between the non-contact and reference sensors, and the highest correlation coefficient achieved for both HR and SpO2 is 0.96. For the inter-bit interval, the accumulated percentage or quality index (QI) is estimated considering a 100 ms R-peak error, and the achieved QI is 97.5%. Both time domain and frequency domain features are calculated for heart rate variability (HRV), and the highest correlation coefficient achieved is 0.93. The results show that high degrees of accuracy have been achieved compared to the reference measurements. Again, one-way analysis of variance (ANOVA1) is applied between the proposed method and reference sensor cStress for the parameters HR, IBI and SpO2. The ANOVA1 test indicates that there is no significant difference between the proposed method and the existing sensor system. In the future, the proposed approach should be evaluated in a real-road driving environment and in other complex challenging situations, such as high temperature; a completely dark/bright environment; unusual movements; and facial occlusion by hands, sunglasses, scarves, beards, etc.

Acknowledgment

The authors would like to acknowledge the Swedish Knowledge Foundation (KKS), Swedish Governmental agency for innovation Systems (VINNOVA), Volvo Car Corporation (VCC), The Swedish National Road and Transport Research Institute (VTI), Karolinska Institute (KI), Autoliv AB, Hök instrument AB, Anpassarna AB and Prevas AB for their support of the research projects. A bug thanks to VTI for their Car simulator for data collection and the test persons for their support and cooperation. Many thanks to reviewers for their comments.

References

[1] W. Verkruysse *et al.*, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, pp. 21434–21445, 2008.
[2] A. Hertzman and C. Spealman, "Observations on the finger volume pulse recorded photoelectrically," *Amer. J. Physiol.*, vol. 119, pp. 334–335, 1937.
[3] V. Jeanne *et al.*, "Camera-based heart rate monitoring in highly dynamic light conditions," in *Proc. Int. Conf. Connected Veh. Expo.*, 2013, pp. 798–799.
[4] A. K. Kanva *et al.*, "Determination of SpO and heart-rate using smartphone camera," in *Proc. Int. Conf. Control Instrum. Energy Commun.*, 2014, pp. 237–241.
[5] W. Karlen *et al.*, "Estimation of respiratory rate from photoplethysmographic imaging videos compared to pulse oximetry," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1331–1338, Jul. 2015.
[6] R. C. Peng *et al.*, "Extraction of heart rate variability from smartphone photoplethysmograms," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–11, 2015.
[7] M. A. Hassan *et al.*, "Heart rate estimation using facial video: A review," *Biomed. Signal Process. Control*, vol. 38, pp. 346–360, 2017.
[8] U. Bal, "Non-contact estimation of heart rate and oxygen saturation using ambient light," *Biomed. Opt. Express*, vol. 6, pp. 86–97, 2015.
[9] J. Kranjec *et al.*, "Non-contact heart rate and heart rate variability measurements: A review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, 2014.
[10] M. Z. Poh *et al.*, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, pp. 10762–10774, 2010.
[11] T. I. Papon *et al.*, "Non-invasive heart rate measuring smartphone applications using on-board cameras: A short survey," in *Proc. Int. Conf. Netw. Syst. Security*, 2015, pp. 1–6.
[12] F. Bousefsaf *et al.*, "Remote assessment of physiological parameters by non-contact technologies to quantify and detect mental stress states," in *Proc. Int. Conf. Control Decis. Inf. Technol.*, 2014, pp. 719–723.
[13] L. He *et al.*, "A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 2088–2091.
[14] D. N. Tran *et al.*, "A robust real time system for remote heart rate measurement via camera," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2015, pp. 1–6.
[15] H. Rahman *et al.*, "SmartMirror: An embedded non-contact system for health monitoring at home (Oct 2016)," in *Proc. 3rd Int. Conf. IoT Technol. HealthCare*, Västerås, Sweden, Oct. 18–19, 2016, pp. 133–137.
[16] X. Li *et al.*, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4264–4271.
[17] C. Park and H.-J. Choi, "Motion artifact reduction in PPG signals from face: Face tracking & stochastic state space modeling approach," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 3280–3283.
[18] W. J. Jiang *et al.*, "Real-time quantifying heart beat rate from facial video recording on a smart phone using Kalman filters," in *Proc. IEEE 16th Int. Conf. E-Health Netw. Appl. Serv.*, 2014, pp. 393–396.
[19] K. Sungjun *et al.*, "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.*, 2012, pp. 2174–2177.
[20] G. Zhenyu *et al.*, "Physiological parameter monitoring of drivers based on video data and independent vector analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 4374–4378.
[21] D. Datcu *et al.*, "Noncontact automatic heart rate analysis in visible spectrum by specific face regions," in *Proc. 14th Int. Conf. Comput. Syst. Technol.*, Ruse, Bulgaria, 2013, pp. 120–127.
[22] M.-Z. Poh *et al.*, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
[23] M. Lewandowska *et al.*, "Measuring pulse rate with a webcam—A noncontact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, 2011, pp. 405–410.
[24] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, pp. 853–857, 2007.
[25] K. Humphreys *et al.*, "Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry," *Rev. Sci. Instrum.*, vol. 78, 2007, Art. no. 044304.

[26] M. Garbey *et al.*, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1418–1426, Aug. 2007.

[27] G. D. Costa, "Optical remote sensing of heartbeats," *Opt. Commun.*, vol. 117, pp. 395–398, 1995.

[28] J. B. Bolkhovsky *et al.*, "Statistical analysis of heart rate and heart rate variability monitoring through the use of smart phone cameras," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 1610–1613.

[29] K. Banitsas *et al.*, "A simple algorithm to monitor HR for real time treatment applications," in *Proc. 9th Int. Conf. Inf. Technol. Appl. Biomed.*, 2009, pp. 1–5.

[30] A. Parnandi and R. Gutierrez-Osuna, "Contactless measurement of heart rate variability from pupillary fluctuations," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interact.*, 2013, pp. 191–196.

[31] K. Hung and Y.-T. Zhang, "Preliminary investigation of pupil size variability: Toward non-contact assessment of cardiovascular variability," in *Proc. 3rd IEEE/EMBS Int. Summer School Med. Devices Biosens.*, 2006, pp. 137–140.

[32] C. Puri *et al.*, "StressCam: Non-contact measurement of users' emotional states through thermal imaging," in *Proc. Conf. Hum. Factors Comput. Syst.*, Portland, OR, USA, 2005, pp. 1725–1728.

[33] S. Suzuki *et al.*, "Development of non-contact monitoring system of heart rate variability (HRV)—An approach of remote sensing for ubiquitous technology," in *Ergonomics and Health Aspects of Work With Computers*, B.-T. Karsh, Ed. Berlin, Germany: Springer, 2009, pp. 195–203.

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

[35] S. R. Yadhuraj *et al.*, "GUI creation for removal of motion artifact in PPG signals," in *Proc. 3rd Int. Conf. Adv. Comput. Commun. Syst.*, 2016, pp. 1–5.

[36] A. Rareş *et al.*, "Image sequence restoration in the presence of pathological motion and severe artifacts," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, pp. IV-3365–IV-3368.

[37] K. Meding *et al.*, "Automatic detection of motion artifacts in MR images using CNNS," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 811–815.

[38] L. Feng *et al.*, "Motion artifacts suppression for remote imaging photoplethysmography," in *Proc. 19th Int. Conf. Digit. Signal Process.*, 2014, pp. 18–23.

[39] J. M. Cho *et al.*, "A preliminary study on photoplethysmogram (PPG) signal analysis for reduction of motion artifact in frequency domain," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci.*, 2012, pp. 28–33.

[40] Z. Qi *et al.*, "Webcam based non-contact real-time monitoring for the physiological parameters of drivers," in *Proc. IEEE 4th Annu. Int. Conf. Cyber Technol. Autom. Control Intell. Syst.*, 2014, pp. 648–652.

[41] H. Qi *et al.*, "Non-contact driver cardiac physiological monitoring using video data, in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, 2015, pp. 418–422.

[42] R. S. Barbosa, "Vehicle vibration response subjected to longwave measured pavement irregularity," *J. Mech. Eng. Autom.*, vol. 2, pp. 17–24, 2012.

[43] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, pp. 747–747.

[44] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision (IJCAI)," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.

[45] A. Asthana *et al.*, "Robust discriminative response map fitting with constrained local models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3444–3451.

[46] H. Rahman *et al.*, "Non-contact heart rate monitoring using lab color space, in *Proc. 13th Int. Conf. Wearable Micro Nano Technol. Pers. Health*, Crete, Greece, May 29–31, 2016.

[47] R. Gonzalez and R. Woods, *Digital Image Processing*. Reading, MA, USA: Addison-Wesley, 1992.

[48] M. Z. Islam *et al.*, "Performance comparison of modified LMS and RLS algorithms in de-noising of ECG signals," *Int. J. Eng. Technol.*, vol. 2, pp. 466–468, 2012.

[49] H. Rahman *et al.*, "Real time heart rate monitoring from facial RGB Color video using webcam, in *Proc. 29th Annu. Workshop Swedish Artif. Intell. Soc.*, Malmö, Sweden, 2016.

[50] H. Rahman *et al.*, "Non-contact physiological parameters extraction using camera, in *Proc. 1st Workshop Embedded Sens. Syst. Health Through Internet Things*, Oct. 2015, pp. 448–453.

[51] H. Rahman *et al.*, "Vision-based remote heart rate variability monitoring using camera, in *Proc. 4th Int. Conf. IoT Technol. HealthCare*, Angers, France, Oct. 24–25, 2017, pp. 10–18.

[52] J. Martin Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, pp. 307–310, 1986.

[53] M. U. Ahmed *et al.*, "Quality index analysis on camera-based R-peak identification considering movements and light illumination," in *Proc. 15th Int. Conf. Wearable Micro Nano Technol. Pers. Health*, Gjøvik, Norway, 2018, pp. 84–92.

[54] P. W. Kamen *et al.*, "Application of the poincar plot to heart rate variability: A new measure of functional status in heart failure," *Internal Med. J.*, vol. 25, pp. 18–26, 1995.