IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 27, 2019, accepted June 8, 2019, date of publication June 12, 2019, date of current version June 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922304

Optimizing Remote Photoplethysmography Using Adaptive Skin Segmentation for Real-Time Heart Rate Monitoring

R. M. FOUAD[1,2], OSAMA A. OMER[1,2], AND MOUSTAFA H. ALY[3]

[1]Electrical Engineering Dept., Faculty of Engineering, Aswan University, Aswan 81542, Egypt.
[2]College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Aswan 81511, Egypt
[3]College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria 21913, Egypt

Corresponding author: Osama A. Omer (omer.osama@aswu.edu.eg).

ABSTRACT Choosing a proper Region of Interest (ROI) for Remote Photoplethysmography (rPPG) is essential and a challenging first step, and it has a direct effect on the accuracy and reliability of the overall heart rate (HR) algorithm. Non-skin areas have no contribution to the HR information; however, few works have tackled the issue of non-skin pixels included in the ROI. First, this paper considers improving the quality of the rPPG signal by filtering out non-skin pixels included within the ROI. The feasibility of employing skin segmentation for ROI definition is demonstrated. Then, this technique is compared with our previous real-time rPPG-based method. Moreover, we explore the effect of extracting the HR from three ROIs using signal fusion. Second, we give a comprehensive account of the examined methods in our algorithm for face detection, face tracking, skin detection, and blind signal separation. Finally, we compare our rPPG measurements with ground truth values obtained from a commercial pulse oximeter. Based on the simulation results, the proposed algorithm significantly improves the quality of the rPPG technique.

INDEX TERMS Heart rate, remote photoplethysmography (rPPG), unobtrusive monitoring, skin segmentation, real-time.

I. INTRODUCTION

Remote measurement of human vital signs has a promising potential for clinical and non-clinical scenarios. It provides a comfortable way for determining subjects' physiological parameters for diagnosis and regular checking. Quite recently, a considerable attention has been paid to remote Photoplethysmography (rPPG), which is a camera-based contactless acquisition method of the human cardiac pulse. A remarkable feature of this emerging method is the ability to estimate the vital signs without any physical contact with the subject, which eliminates any inconvenience during the assessment process. This striking feature opens the door for employing rPPG technology in non-clinical applications, such as home health monitoring, driver monitoring, fitness-cardio training, video surveillance, and face anti-spoofing.

Traditional contact methods for measuring HR comprise Electrocardiography (ECG) and Pulse Oximetry. ECG offers the most accurate HR measurements, but it requires attaching medical electrodes to the subject. Similarly, pulse oximeter, which is based on Photoplethysmography (PPG), must be attached to a body part, such as toes or an earlobe. These two methods are robust and cost-efficient; however, both have several drawbacks because both require direct contact with the subject. Attaching medical sensors and electrodes to the subject is inconvenient, as it may cause pain and stress. The aforementioned disadvantages of contact methods accentuate the need for contactless HR measuring methods.

Lately, there have been numerous studies investigating contactless HR measuring methods; including methods based on Doppler effect [1], thermal imaging [2], and piezoelectric measurements [3]. These techniques have been examined in literature, but with limited applications due to size, cost and complexity. Most recent studies as well as current work focus on camera based HR monitoring methods. These methods are widely considered to be very efficient techniques for measuring the HR remotely, and many researchers have proven its feasibility. Moreover, these techniques are very promising

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

in terms of safety, cost, health, reliability, and computational efficiency.

Camera-based HR measuring methods are divided into two types: color-based methods (rPPG), which extracts the HR from the slight color changes happening in the skin due to heartbeats [4]; and motion-based methods (Ballisto-cardiography (BCG)), which measures the HR from the tiny body motions accompanying the cardiac activity [5]. The rPPG-based methods offer encouraging results for both static and dynamic scenarios. It allows measuring the HR from multiple persons simultaneously. Additionally, the HR can be extracted from several Region of Interests (ROIs) (e.g., forehead and cheeks), however these ROIs are required to be clear to allow accurate HR measurements. The major drawback of this technique is that PPG signal, which contains information about the HR, is not the same for different ROIs. Furthermore, poor illumination and skin tone variation affect its accuracy. On the other hand, because the BCG-based methods rely on the mechanical effects due to cardiac activity, it has the advantage of extracting the HR from unclear or not visible ROIs, and it is robust against varying illumination and skin colors. Nevertheless, this method suffers when the subject is not steady. Each of these methods (Color-based and Motion-based) has its pros and cons under various circumstances, resulting in challenging practical problems.

A series of recent studies have proven the feasibility of applying the camera-based HR monitoring methods in cars for drivers monitoring [6]–[8]. Several techniques have been proposed to apply it in practical setting, some focusing on improving the method itself, and others on finding the most reliable setting for accurate measurements. In short, the literature pertaining to using camera-based methods strongly agrees that these methods are very promising, but they are still in their infancy, meaning that a further research is required.

Here, our paper is focusing on developing a robust HR measuring system for the drivers monitoring, where the challenges such as motion artifacts, and poor illumination are the main concerns. Within the framework of these criteria, this paper focuses on refining the quality of the extracted cardiovascular wave by using skin segmentation for defining the ROI, which enables us to estimate the HR from skin pixels only.

The main contributions of this paper are represented as:

1- The adaptive skin detection for ROI definition is investigated. It is then compared to our previous real-time rPPG-based method, in which face detection only was used to define the ROI [9].
2- Furthermore, we examine the effect of segmenting the face skin into three ROIs, and then extract the HR from these three ROIs by fusing their resultant HR signals.
3- The effect of using different methods for face detection, face tracking, skin detection, and blind signal separation is examined. Subsequently, we compare these methods to decide which one achieves better performance.
4- Finally, we compare our rPPG measurements with ground truth values obtained from a commercial pulse oximeter.

The remainder of the paper is organized as follows: Section II explains origin and fundamental principles of rPPG. The related works are reviewed in Section III. The proposed method is discussed in Section IV. Experimental results are presented and discussed in Section V. Section VI summarizes the results of this work and draws the main conclusions.

II. BACKGROUND

A. PHOTOPLETHYSMOGRAPHY (PPG)

The origin of rPPG can be traced back to PPG, a method firstly introduced in 1937 by Hertzman and Spealman [10]. In broad terms, PPG can be defined as a non-invasive optical method for detecting the blood volume pulse (BVP). Arterial blood light absorption differs from that of the surrounding biological tissues at specific optical wavelengths [11], [12]. Moreover, blood volume in capillaries under the skin changes periodically with heartbeats. PPG takes the advantage of this phenomenon to extract the BVP from the light reflected from the skin, which is modulated according to the small blood volume variations in capillaries.

Based on how the light source and photodetector are configured, PPG operates in two possible modes: reflectance mode and transmission mode, where the usage of the later one is limited to specific areas, such as fingertip [13].

B. REMOTE PHOTOPLETHYSMOGRAPHY (rPPG)

1) ALGORITHMIC BASICS OF rPPG

In previous studies, many approaches have been adopted in rPPG algorithms. These are:

i. **Blind Source Separation (BSS)-based method:** This approach can be divided into ICA-based methods [14], and PCA-based methods [15]. In this approach, the BVP signal is recovered by splitting the mixed RGB signals into independent source signals.
ii. **Chrominance-based method (CHROM):** In this method, a linear addition of the chrominance-signals is performed, and a standard skin tone is used to white balance the image [16].
iii. **BVP signature based method:**This method is robust to motion artifacts. It separates color variations, due to heart rate, from the noise produced by motion artifacts. This separation is done by utilizing the blood volume pulse signature in different optical wavelengths [17].
iv. **Spatial Subspace Rotation (2SR):**This is a novel method introduced in [18]. In this technique, the BVP signal is recovered by measuring spatial subspace time-rotation of the skin pixels.

The main difference between each of these rPPG approaches is how red, green and blue signals are combined to extract the BVP signal. Our rPPG system is based on Blind Source Separation (BSS) method.

We use the rPPG mathematical model defined in [19] to understand the basic principles of rPPG algorithms. This mathematical model enables us to define and analyze different problems and limitations regarding rPPG in details.

2) SKIN REFLECTION MODEL

Consider rPPG measuring system setup, which consists of three main parts: light source (i.e., ambient light), camera and a human skin with pulsating blood. The light source is assumed to have varying intensities with constant optical properties. Moreover, the light varying intensities are dependent on the distance between the camera and the skin.

Based on the dichromatic model, the light reflected from the skin can be modeled as follows [19]:

$$\mathrm{C_k}\,(t) = I\,(t)\,.\,(\mathrm{V_s}\,(t) + \mathrm{V_d}\,(t)) + \mathrm{V_n}\,(t)\,, \tag{1}$$

where:

- $\mathrm{C_k}\,(t)$: The RGB signals (ordered in one column).
- $I(t)$: the light intensity.
- $\mathrm{V_s}(t)$: the specular reflection.
- $\mathrm{V_d}\,(t)$: the diffuse reflection.
- $\mathrm{V_n}(t)$: camera quantization noise.

$\mathrm{C_k}(t)$ represents the RGB values of the k-th pixel in the monitored skin area; the level of light intensity is denoted by $I(t)$. Light intensity is dependent on light source and the distance between camera, monitored skin area, and the light source. The light intensity is modulated by the specular reflection $\mathrm{V_s}(t)$ and the diffuse reflection $\mathrm{V_d}(t)$. $\mathrm{V_n}(t)$ is the quantization noise, which results from sensors of the camera. It represents the undesired random color and brightness changes in camera images. All variables in Eq. (1) are time dependent because of motion artifacts and the pulsating nature of the blood.

The specular reflection light is reflected from the skin surface, and it does not penetrate skin tissues. Therefore, the specular reflection light does not contain any information about the BVP signal; it is similar to the light that is reflected from a mirror. Moreover, the spectral composition of specular reflection light and the light source are identical. $\mathrm{V_S}(t)$ can be written as follows:

$$\mathrm{V_s}\,(t) = \mathrm{u_s} \cdot (s_0 + s(t)), \tag{2}$$

where:

- $\mathrm{u_s}$: the light spectrum's unit color vector.
- s_0: the stationary component of specular reflection.
- $s(t)$: the varying component of specular reflection due to motion artifacts.
- The diffuse reflection light is the light which penetrates the skin, then it is reflected and scattered inside skin tissues. Thus, diffuse reflection light contains a useful information about the BVP signal.

$V_d(t)$ can be written as following follows:

$$V_d\,(t) = \mathrm{u_d} \cdot d_0 + \mathrm{u_p} \cdot p\,(t)\,, \tag{3}$$

where:

- $\mathrm{u_d}$: unit vector of the skin color.
- d_0: represents the motionless light reflection strength.
- $\mathrm{u_p}$: the strength of relative pulsatile in RGB signals.
- $p(t)$: the BVP signal.

By substituting Eq. (2) and Eq. (3) into Eq. (1), we reach:

$$\mathrm{C_k}(t) = I\,(t) \cdot \left(\mathrm{u_s} \cdot (s_0 + s(t)) + \mathrm{u_d} \cdot d_0 + \mathrm{u_p} \cdot p\,(t)\right) + \mathrm{V_n}\,(t)\,, \tag{4}$$

The motionless components in specular reflection and diffuse reflection are added up together to form one motionless skin reflection component as follows:

$$\mathrm{u_c} \cdot c_0 = \mathrm{u_s} \cdot s_0 + \mathrm{u_d} \cdot d_0, \tag{5}$$

where:

- $\mathrm{u_c}$: The unit vector of the skin color reflection.
- c_0: the reflection strength.

The light intensity $I(t)$ can be expressed as follows:

$$I\,(t) = I_0 + I_0 \cdot i\,(t)\,, \tag{6}$$

where:

- I_0: the stationary part of the light intensity.
- $i(t)$: the time varying part of the light intensity.

We can rewrite Equation 4 as follows:

$$\mathrm{C_k}\,(t) = I_0 \cdot (1 + i\,(t)) \cdot \left(u_c \cdot c_0 + u_s \cdot s(t) + u_p \cdot p\,(t)\right) + \mathrm{V_n}(t) \tag{7}$$

The time-varying components $i(t)$, $s(t)$, and $p(t)$ are zero-mean signals. It is worth mentioning that the specular reflection light is the dominant component among all other components. Therefore, skin areas, where all components are negligible to specular reflection light, have to be rejected.

From Eq. (7), we can conclude that the solution to rPPG problem is recovering the BVP signal $p\,(t)$ from the RGB signals $\mathrm{c_k}(t)$.

Given that the number of pixels, within the monitored skin area, is satisfactorily large, the camera quantization error effect can be eliminated by taking the average pixels' value within the monitored skin-tissue area. This step leads to:

$$\mathrm{C}\,(t) \approx I_0 \cdot (1 + i\,(t)) \cdot \left(u_c \cdot c_0 + u_s \cdot s\,(t) + u_p \cdot p\,(t)\right), \tag{8}$$

This yields $\mathrm{C}(t)$ without camera quantization error, however the camera quantization error cannot be neglected when the number of pixels, within the monitored skin area, is not large enough.

Equation (8) is expanded to:

$$\begin{aligned}\mathrm{C}\,(t) &= \mathrm{u_c} \cdot I_0 \cdot c_0 + \mathrm{u_s} \cdot I_0 \cdot s\,(t) + \mathrm{u_p} \cdot I_0 \cdot p\,(t) + \mathrm{u_c} \cdot I_0 \cdot c_0 \cdot i\,(t) \\ &\quad + \mathrm{u_s} \cdot I_0 \cdot s\,(t) \cdot i\,(t) + \mathrm{u_p} \cdot I_0 \cdot p\,(t) \cdot i(t),\end{aligned} \tag{9}$$

In Eq. (9), the DC components are much bigger compared to AC components. Therefore, Equation 9 is approximated to:

$$\mathrm{C}(t) \approx \mathrm{u_c} \cdot I_0 \cdot c_0 + \mathrm{u_c} \cdot I_0 \cdot c_0 \cdot i\,(t) + \mathrm{u_s} \cdot I_0 \cdot s\,(t) + \mathrm{u_p} \cdot I_0 \cdot p\,(t)\,, \tag{10}$$

This approximation confirms that C(t) is a linear combination of $i(t)$, $s(t)$, and $p(t)$. Thus, recovering BVP signal from C(t) can be done through a decomposition system, which recovers the source signals $i(t)$, $s(t)$, and $p(t)$ from the mixed signals in C(t). Based on this finding, Blind Source Separation (BSS) is considered as an ideal method for decomposing C (t) into source signals. For example, a suitable ICA technique [33], [66]–[69], can be used for this decomposing process.

Each video frame f_i can be written as follows:

$$f_i = \begin{bmatrix} R_i^{1,1} G_i^{1,1} B_i^{1,1} & \cdots & R_i^{N,1} G_i^{N,1} B_i^{N,1} \\ \vdots & \ddots & \vdots \\ R_i^{1,M} G_i^{1,M} B_i^{1,M} & \cdots & R_i^{N,M} G_i^{N,M} B_i^{N,M} \end{bmatrix} \quad (11)$$

where (M, N) is the frame size. Each video frame can be reduced to include only pixels within ROI as in $f_{ROI}(i)$, shown at the bottom of this page, where the ROI is defined by a bounding box with coordinates X and Y, and with height H and width W.

By taking the average pixel value within the ROI, f_{ROI} can be reduced as follows:

$$f'_{ROI}(i) = \begin{bmatrix} R_i \\ G_i \\ B_i \end{bmatrix} \quad (12)$$

Now, we can equate Eq. (10) with Eq. (13)

$$f'_{ROI}(j) = \mathrm{C}(j) = \begin{bmatrix} R_j \\ G_j \\ B_j \end{bmatrix} \approx \mathrm{u_c} \cdot I_0 \cdot c_0 + \mathrm{u_c} \cdot I_0 \cdot c_0 \cdot i(j) + \mathrm{u_s} \cdot I_0 \cdot s(j) + \mathrm{u_p} \cdot I_0 \cdot p(j) \quad (13)$$

Equation (13) represents the raw signals. Moreover, it relates the skin reflection model with the actual signals obtained from the video stream.

III. RELATED WORKS

In 2008, for the first time, Verkruysse *et al.* [20] proved the possibility of measuring heart rate remotely using a consumer level camera. Their technique is commonly referred to as rPPG, which is similar to traditional PPG, a method firstly introduced in 1937 by Hertzman and Spealman [10]. In their experiments, subjects were asked to stay stationary, and they recorded subjects' face using consumer level camera from several meters. Their basic system setup is illustrated in Figure 1. The recorded videos resolution was 640 × 480 with a frame rate of 30 fps. In this study, the forehead was chosen as ROI, and the raw RGB signals were computed in every frame by calculating the mean value of all pixels within the ROI. Then, FFT was used to analyze the computed signals

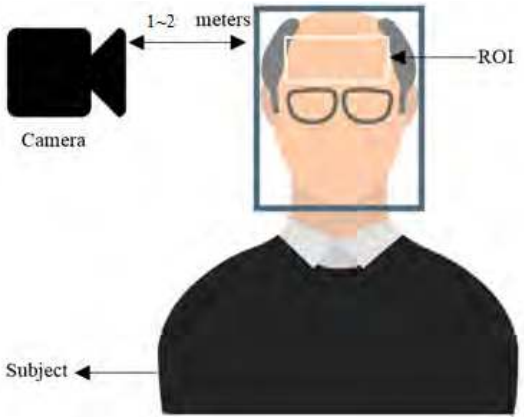

FIGURE 1. Basic setup of an rPPG system.

in frequency domain. In this study, it has been shown that the strongest plethysmographic signal can be found in the green channel, and this discovery agrees with the fact that hemoglobin has a better absorption to green light than red light and blue light [21]. However, red and blue channels also contain information about the heart rate. Verkruysse *et al.* made no attempt to utilize heart rate information from blue and red channels.

In 2011, Poh *et al.* [14] improved the system presented in [22]. Their implemented method was tolerant to motion artifacts, and it was capable of measuring heart rate with automatic face detection and tracking. They recorded subjects' face using a webcam from about 50 cm, and a 30-seconds moving analysis window was used. Their recorded videos resolution was 640 × 480 with 15 fps and a 24-bit RGB color space. One of their major advances is utilizing the three RGB channels to extract the cardiovascular heart wave.

In 2012, Kwon *et al.* [23] reproduced Poh's approach and developed the FaceBEAT application on a smartphone. In this study, they explained the potential that the reliable heart rate can be measured remotely by the facial video recorded using a smartphone camera. First, using the front facing-camera of a smartphone, facial video was recorded. They detected facial region on the image of each frame using face detection, and yielded the raw trace signal from the green channel of the image. To extract more accurate cardiac pulse signal, they applied independent component analysis (ICA) to the raw trace signal. The heart rate was extracted using frequency analysis of the raw trace signal and the analyzed signal from ICA. The main contribution in this study is proving the feasibility of using smartphones as video source and for computation.

In 2013, an approach using an ROI and neural-network-based skin detection was proposed [24] that allows for more

$$f_{ROI}(i) = \begin{bmatrix} R_i^{X,Y} G_i^{X,Y} B_i^{X,Y} & \cdots & R_i^{X+W,Y} G_i^{X+W,Y} B_i^{X+W,Y} \\ \vdots & \ddots & \vdots \\ R_i^{X,Y+H} G_i^{X,Y+H} B_i^{X,Y+H} & \cdots & R_i^{X+W,Y+H} G_i^{X+W,Y+H} B_i^{X+W,Y+H} \end{bmatrix}$$

accurate measurement. Additional areas such as the neck and arms are included in the ROI in this algorithm.

In 2013, there was a comparative study for linear and nonlinear techniques for BSS. In this study, it was found that Laplacian eigenmap produces gave best results [25]. Other studies [26] addressed the problem of moving subjects with respect to the light source. In this study, authors argued that an optimal fixed combination of band passed RGB channel signals can be found based on a ratio of normalized color signals when assuming ''standardized'' skin, thus eliminating noise derived from specular reflection. A deficiency of this approach was that it excluded BSS from the algorithm's design. Furthermore, Sun *et al.* [27] proposed to investigate the feasibility of remote assessment of HR, RR, and HRV by applying a time-frequency representation method to the video recordings of the subjects' palm regions. All videos were recorded at a rate of 200 fps under the resting conditions to minimize motion artifacts. The authors demonstrated that 200-fps iPPG system could provide a closely comparable measurement of HR, RR, and HRV to those acquired from contact PPG references.

In 2014, Li *et al.* also aimed at reducing noise [28] by using an adaptive filter to reduce noise from illumination changes using background illumination as a reference. There are other algorithms for facial landmark detection that have been proposed [28] which applied discriminative response map fitting (DRMF) [29] after face detection. Researchers working with facial landmark points used these points to define more exact and robust ROIs. Using nine landmark points, [28] defined a region that includes the cheeks and no background pixels, similar to [30] and [31], which defined a region that includes the forehead and the area below the eyes. The location of the ROI can be updated frame by frame without having to re-detect the ROI.

Using BSS, the component selection, have recently been optimized [32] different machine learning methods to extract HR from features Similarly, [33] used support vector regression (SVR) to extract the HR from frequency domain features. The deficiencies in this study is that it has no detailed on ROIs.

In 2015, a different approach was used [34], the facial region was divided into many small ROIs that yielded an array of signals from the green channel, each of which was later combined using a weighted average based on a goodness metric. Similarly, the researchers in [35] stochastically selected an array of points and combined them using an importance-weighted Monte Carlo approach.

Tran *et al.* [36] proposed a real-time rPPG-based system using consumer-level camera. They combined face detection with skin detection for accurate construction of ROI. Then, Linear Discriminant Analysis (LDA) was used to extract heart rate signal from the three RGB channels. Another real-time rPPG-based mobile application was developed in [37]. Instead of implementing face detection and ROI localization in every frame as in [22], this method used LK optical flow feature [38] to track the ROI.

In 2016, Rumi ski [39] demonstrated the possibility of estimating HR from rPPG signals in the YCrCb (YUV) space using both ICA-based and PCA-based methods. The experimental results showed that the best HR estimation performance can be achieved by applying PCA to the V channel obtained from a forehead. Another rPPG studies, aiming at eliminating the impact of illumination variations [40]. The framework using EEMD followed by a multiple-linear regression model was later employed to evaluate HR for reducing the effects of ambient light changes. The deficiency in this study is utilizing the only green channel to extract the cardiovascular heart wave.

In 2017 [41], it was shown that at the HR computation stage, the frequency domain methods are not capable to detect instantaneous heartbeat changes, and are not as robust as time domain methods according to Poh *et al.* Supervised learning methods are mainly applied at this stage. There is one recent paper applying auto-regression to extract BVP signals after the video processing [41], but there is no end-to-end usage of supervised learning methods. Later, Qi *et al.* [42] proposed a novel method for noncontact HR measurement by exploring correlations among facial subregion data sets via JBSS. The testing results on a large public database also demonstrated that the proposed JBSS method outperformed previous ICA-based methodologies.

In 2018, the two influential studies on rPPG [21], [22] have motivated a growing number of publications and progressions in this area, which was indicated in [43]. This research provided a more complete and comprehensive understanding of rPPG, facilitate further development of rPPG, and inspire numerous potential applications in healthcare.

In [44], a novel framework for remote Photoplethysmography pulse extraction on compressed videos is proposed. In [45], the general framework is used to design rPPG algorithms for specific situations. Moreover, authors provided the parameters which can be adjusted to increase the accuracy.

In the literature, numerous names and abbreviations of this method have evolved, such as rPPG [26], camera-based PPG (cbPPG) [16], non-contact PPG (ncPPG) [46], distance PPG (DistancePPG) [34], pulse camera (PulseCam) [47], videoplethysmography (VPG) [39], imaging PPG (iPPG/IPPG) [49], PPG imaging (PPGi/PPGI) [49], and remote imaging PPG (RIPPG) [50]. In this paper, we will use the term rPPG.

IV. METHODOLOGY

As shown in the previous section, a variety of rPPG methods have been reported in the literature, even though most previous works have based their methods on a common framework [18], [26], [35], [48]. This framework comprises of: Defining ROIs through detection and tracking, extracting the raw signals, processing and filtering raw signals, combining RGB channels to find the cardiac wave, and finally extract the vital sign of interest. The major differences between the various rPPG methods lie in how the ROI is

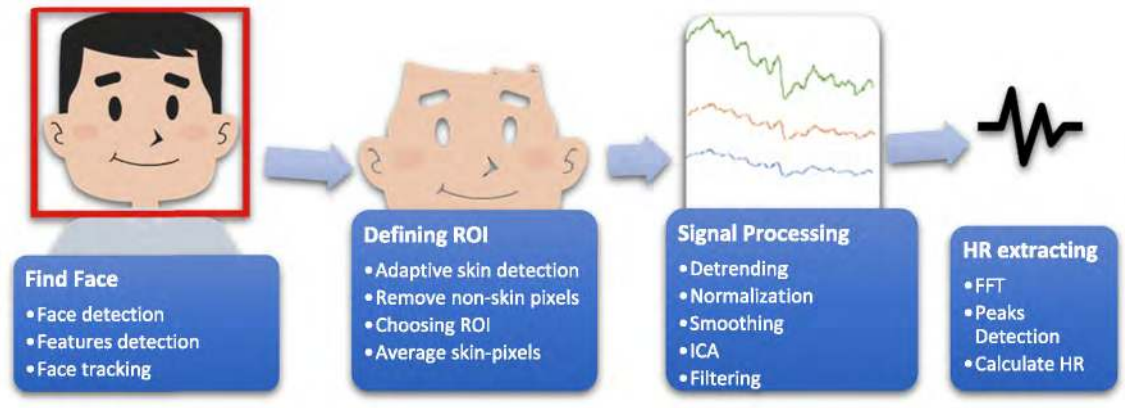

FIGURE 2. Block diagram of the proposed rPPG method.

defined and how the RGB signals are combined to estimate the cardiovascular signal.

Selecting a suitable ROI for the rPPG-based HR measuring is essential and challenging first step, and it has a direct impact on the accuracy and reliability of the overall algorithm. Non-skin areas have no contribution to the HR information; however, few researchers have addressed the issue of non-skin pixels included in the ROI [51], [52].

Our approach focuses on refining the quality of the extracted cardiovascular wave using skin segmentation for defining the ROI, which enables us to estimate the HR from skin pixels only. The block diagram of the proposed rPPG method is shown in Fig. 2. We can divide our method into three main parts: raw signals extraction, signal processing, and HR extracting, as follows. Figure 3 shows the flowchart of the proposed HR monitoring system.

A. RAW SIGNALS EXTRACTION

The aim of this part is to constitute the temporal RGB signals which are used to estimate the rPPG signal in a later step. This has a significant impact on the quality of the overall rPPG method; extracting a clear signal with distinguishable pulsating rPPG wave, which directly affects the system accuracy and reliability.

Raw signal extraction is a critical step in the process of recovering cardiovascular wave. The raw signals are extracted from red, green, blue channels by calculating the mean of all pixels' values within the ROI in each frame over a 10-seconds sliding window.

The raw signals are extracted from the ROIs. So, the first step is to identify the ROIs. The ROIs are identified by the following procedure:

1) FACE DETECTION

Viola-Jones face detection technique is employed to automatically detect subject's face [53]. This step provides a bounding box coordinates defining the subject face.

2) FACE TRACKING

Implementing face detection at every frame consumes a lot of computational power resources. Moreover, it causes undesired noises because the output bounding box of the face is not the same in the successive frames. Instead of redetecting the face at every frame, we track the face using KLT tracker [54], in which only specific features of the face are tracked over the time.

3) ADAPTIVE SKIN DETECTION

Skin detection is performed on every frame to filter out non-skin pixels. This skin segmentation process is done using the algorithm proposed by Conaire *et al.* [55], which is a reliable method capable of efficiently detecting skin areas. This adaptive skin detection depends on maximizing joint information, where classifiers are added assuming that their errors are complementary, see Fig. 4 (middle).

4) CHOOSING ROI

Two ROI definitions are investigated in this work. The first ROI is all skin-pixels returned from the skin segmentation process, the second ROI has three parts, and it is obtained by extracting the skin-pixels contained in forehead, left cheek, and right cheek. The two ROI definitions are shown in Fig. 4 (right).

5) CONSTITUTE TEMPORAL rgb SIGNALS

The raw RGB signals are composed by calculating the average pixel value of the skin-pixels within the ROI region over time. The extracted raw signals are shown in Fig. 5.

B. SIGNAL PROCESSING

After extracting the raw signals, a number of signal processing techniques are employed such as: detrending, normalization, smoothing, and filtering.

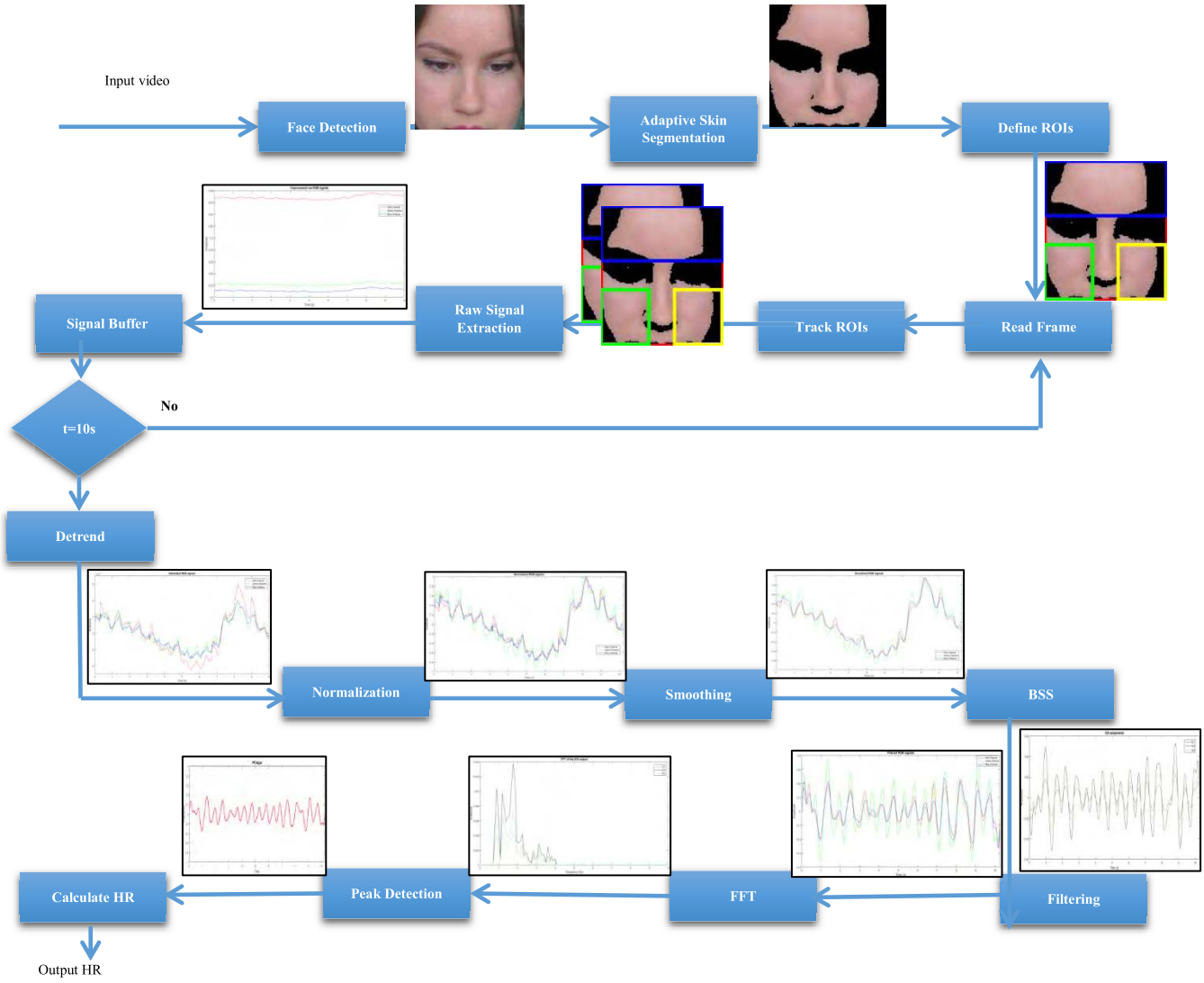

FIGURE 3. Flowchart of the proposed algorithm using adaptive ROIs.

FIGURE 4. Illustration of the employed skin detection method, (left) the detected face, (middle) the detected skin, and (right) the ROIs.

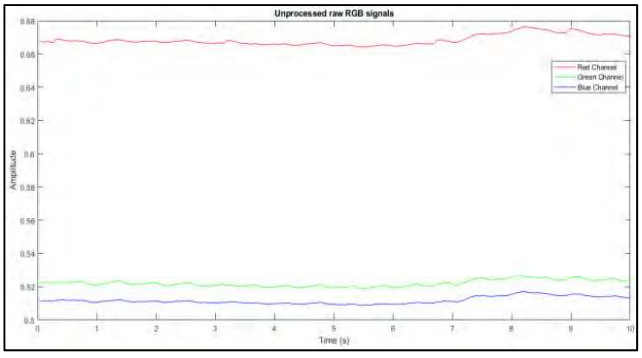

FIGURE 5. Unprocessed raw RGB signals.

1) DETRENDING

Due to the noises caused by changing of the environmental parameters, in this step, we remove the linear trends from the raw signal. The effect of the detrending process is shown in Fig. 6.

2) NORMALIZATION

We are only interested in the periodicity of the signal. Therefore, the raw signal is normalized by dividing the raw signal by its maximum absolute value. The output of this step is shown in Fig. 7.

3) SMOOTHING

In this step, the normalized RGB signals are smoothed using a sliding average filter. This step leaves out noise, and it

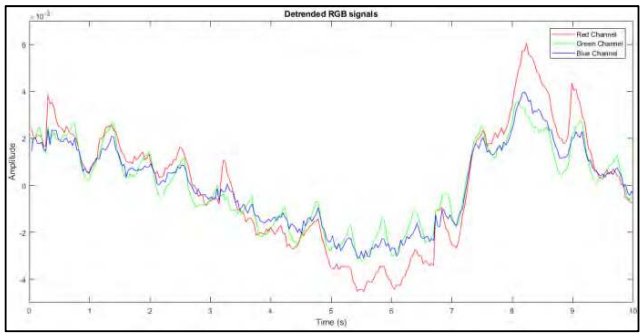

FIGURE 6. Detrended RGB signals.

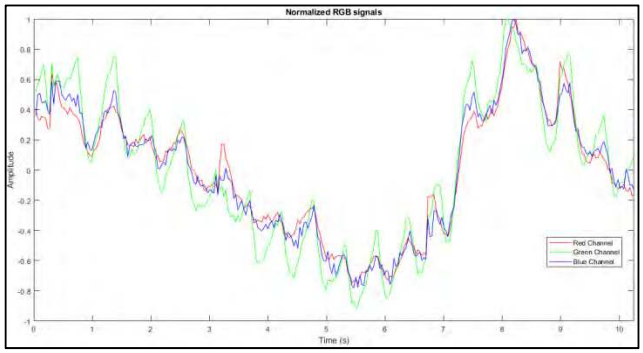

FIGURE 7. Normalized RGB signals.

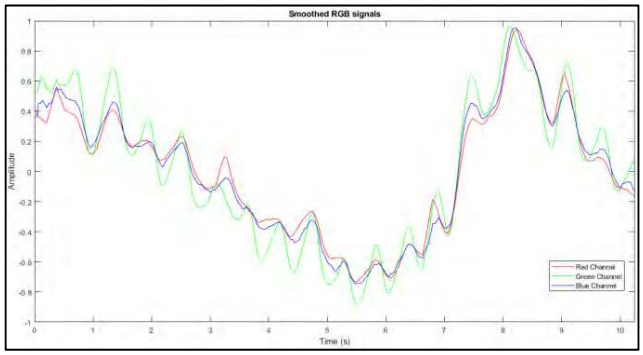

FIGURE 8. Smoothed RGB signals.

clarifies the signal. Therefore, it increases the robustness and flexibility of the analyses. The output of this step is shown in Fig. 8.

4) BLIND SIGNAL SEPARATION

The RGB signals contains information about the HR in mixed components. Therefore, ICA is used to recover the source signals from these mixed signals. In our system, FastICA method is used to analyze the RGB signals to reveal the original source signals. Moreover, FastICA is an effective technique that can be utilized to eliminate noise artifacts. The extracted raw signal is composed of three single series, which were obtained from red, green, and blue channels. The goal of BSS is to recover the one-dimensional plethysmographic signal from the three raw signals. BSS is implemented to

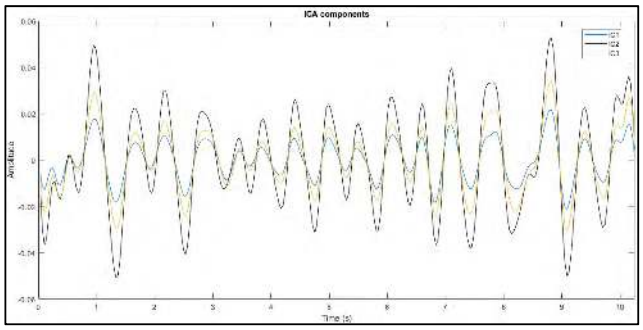

FIGURE 9. ICA components in time domain.

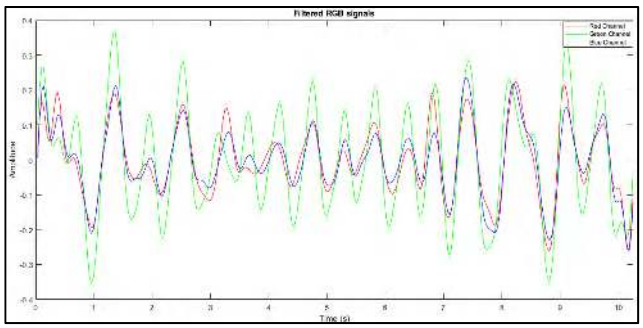

FIGURE 10. Filtered RGB signals in the time domain.

compute the optimal combination of the three raw signals. We use BSS algorithm [56] to get three independent source components from the raw signals. The three ICA components are compared, and the component with the highest periodicity is chosen for the next step, see Fig. 9.

5) FILTERING

The goal of this step is to increase the quality of the acquired plethysmographic signal by removing undesirable noise. Driving conditions include: non-stationary subject and varying illumination, these factors degrade SNR ratio. In this step, the raw signal is applied to a band-pass-filter with ideal behavior to eliminate high and low frequency noise. The filter removes components which exist outside the 0.7 Hz to 4 Hz frequency band. This band was commonly used in previous studies, and it corresponds to heart rate measurements between 42 and 240 bpm, see Fig. 10.

C. HEART RATE CALCULATION

In this part, the spectrum of the resulting ICA components is obtained by using FFT method. The peaks in the ICA components power are determined, and the index frequency of the highest peak corresponds to the HR frequency. This is done by the following sequence.

1) FAST FOURIER TRANSFORM (FFT)

In this step, the spectrum of the chosen ICA component is calculated by using FFT algorithm [57], see Fig. 11.

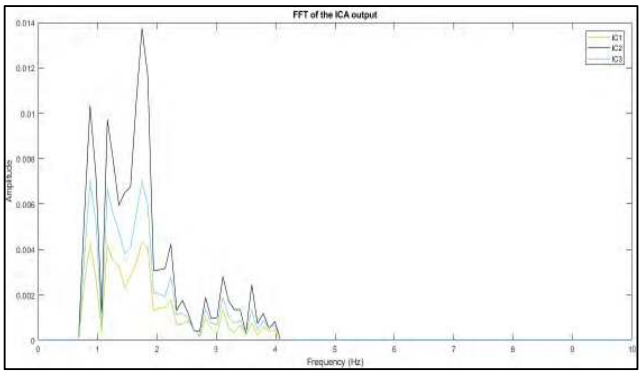

FIGURE 11. FFT of the ICA output.

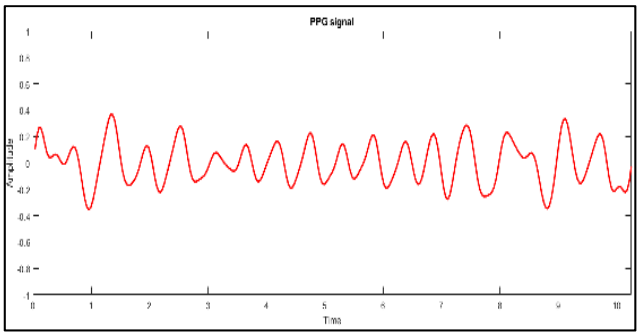

FIGURE 12. Final BVP signal.

2) PEAKS DETECTION

The highest peaks in the signal's power spectrum are obtained in the frequency domain with a sliding Fourier window, which enables varying heart rate estimation. The index frequency of the highest peak is the estimate of the heart rate frequency, see Fig. 12.

3) CALCULATE HEART RATE

The heart rate is calculated by multiplying the estimated heart rate frequency by 60, this yield a value between 42 and 240 bpm.

V. PARAMETERS OPTIMIZATION OF THE PROPOSED rPPG-BASED HEART RATE MONITORING SYSTEM

The main focus of this section is to optimize the developed rPPG-based method in the previous section. With the aim of improving the proposed method, we examine the effect of using different methods for face detection, face tracking, skin detection, and blind signal separation. Subsequently, we compare these methods to decide which one achieves better performance.

A. FACE DETECTION METHODS: VIOLA-JONES vs. LIAO

Face detection is a vital first step in the process of extracting the rPPG signal. Using robust and reliable face detection method plays an important role in determining the accuracy of the overall system. The inability to detect the face at a given time directly leads to failure in defining the ROI, and estimating the HR at this specific period of time. Driving conditions is a challenging environment, in which subject moves unpredictably and lightening varies frequently. Therefore, detecting the face can be hard in such situation; and using a reliable face detection method is a must. In this subsection, two face detection methods are investigated and compared to each other, namely Viola-Jones method [53] and Liao method [36].

1) VIOLA-JONES FACE DETECTION METHOD

Baker and Matthews [54] introduced a machine learning based method for detecting objects very fast. Their technique accomplishes decent detection rates in a very short time, making it ideal for real-time applications. Viola-Jones technique is one of the most widely used methods in detecting subjects' faces, noses, eyes, mouth, or upper body. This technique is able to process images rapidly. Thanks to its novel image representation "Integral linage", which enables quick features detection. Moreover, Viola and Jones developed a machine learning procedure that extracts only small number of features from a greater set and provides competent classifiers. Another important feature of this method is merging complex classifiers in a cascade to quickly remove background areas and focus more on object-like areas.

2) LIAO FACE DETECTION Method

Monkaresi *et al.* [32] developed a robust and fast unconstrained face detection method. This method was designed to address challenges such as position changing and face blockages. For the first time, Normalized Pixel Difference (NPD), which is a novel feature of image, was introduced. NPD is calculated from the pixel values based on Weber Fraction. Additionally, NPD is unchanging with scale, and the original image can be reconstructed from it. Liao *et al.* showed that NPD can be computed from a look up table, leading to a rapid face detection method. Moreover, they proved that their method can provide the state-of-the-art accuracy in unconstrained conditions.

B. FACE TRACKING METHODS: KANADE-LUCAS-TOMASI (KLT) Vs. CAMSHIFT

Implementing face detection at every frame consumes a lot of computational power resources, limiting the real-time application of this method. Moreover, it causes undesired noises because the output bounding box of the face is not the same in the successive frames. Instead of redetecting the face at every frame, we track the face over the time to track the face bounding box. Furthermore, face tracking methods eliminate the effect of small head pose variations on the bounding box, this yields a more robust and stable coordinates of the face. Therefore, face tracking method must be able to update the bounding box coordinates accurately, efficiently, and rapidly. Face tracking process consists of three main parts: Face detection, extracting trackable facial points, and tracking these facial points. This subsection compares the results of the

proposed method using each of the following face tracking methods: Kanade-Lucas-Tomasi (KLT) and CAMShift.

1) KANADE-LUCAS-TOMASI (KLT) FACE TRACKING METHOD

The first examined method is based on the KLT algorithm, in which several feature points are tracked over the frames [59]. At the initialization process of our algorithm, the face is detected using Viola-Jones, and the output of this step is a bounding box containing the face of the subject. This bounding box is the input to the KLT algorithm which next extracts a set of trackable points of the subject's face. These points can be efficiently tracked over the time, and if the movements are not severe, redetecting the face is not needed. In our system, we redetect the face if the number of trackable points is less than 10 points.

2) CAMSHIFT FACE TRACKING METHOD

The second method is based on CAMShift algorithm, wherein the skin color is the tracked feature over the frames [60]. After the face detection step, the face can be tracked using features which are invariant to subject's movements. In this tracking method, the skin color of the subject is the tracked feature. What is interesting about tracking using skin color is the ability to track the face as long as any part of skin is included in the examined region; however, it sometimes suffers when skin-color like objects are near the face. Firstly, the examined video frame is converted to HSV color space and then skin color is determined from the Hue channel. Secondly, the skin color is tracked using a histogram tracker. Since nose has the most precise measurement of the skin color, we extract the Hue channel from only skin-pixels located within the nose region, which are next employed to initialize the histogram-based tracker.

C. SKIN DETECTION METHODS: CONAIRE vs. RGB-H-CbCr

Principally, rPPG is based on extracting the heart rate from the periodic color variations in the skin. For that reason, the skin pixels are the only effective contributors to the information about the cardiovascular wave. A key limitation of the method developed in [9], [14], and [22] is the inability to differentiate between skin pixels and non-skin pixels. This flaw was addressed in the present work; furthermore, the experimental results confirmed the significance of including only skin pixels in the ROI. The main function of skin detection techniques is to identify skin regions in the image frame, which enable us to remove non-skin pixels from the ROI. In our algorithm, skin detection is applied only to the bounding box of the face, which is obtained from the face detection step. So, the most suitable skin detection method for our system should be able to filter out the following areas: background, eyes, eyebrows, hair, facial hair, and any other non-skin areas. This sub-section compares the results of the proposed rPPG algorithm using each of the following skin detection methods: Conaire skin detection method [55] and RGB-H-CbCr skin detection method [61].

1) CONAIRE SKIN DETECTION METHOD

Conaire *et al.* [55], introduced an efficient adaptive machine-learning classification method based on exploiting the mutual agreement among two independent information sources. This adaptive skin detection method depends on maximizing joint information, where classifiers are added assuming that their errors are complementary. The major objective of this method was to overcome the shortcomings of traditional classification methods, which are: time-consuming training process, and using overgeneralized or too small data set for training the classifier. This method takes the advantage of employing two data sources, and their redundant data to generate an adaptive classifier on different data. Additionally, Conaire *et al.* proved the robustness of their method in learning varying skin models.

2) RGB-H-CbCr SKIN DETECTION METHOD

The second investigated skin detection method is based on RGB-H-CbCr skin color model. Cardoso [62], introduced a new color model for the skin called RGB-H-CbCr which is dedicated to face detection. This method recognizes skin areas through a group of bounding rules that are originated from the distribution of skin color in a training data set. The used model employs the redundant hue and chrominance channels with the RGB channels to differentiate between skin pixels and non-skin pixels.

D. BLIND SOURCE SEPARATION (BSS) METHODS

The RGB signals contain information about the HR in a mixed pattern. Therefore, we are allowed to simply define the task of the rPPG system as the process of extracting the cardiovascular wave from the mixed signals in the RGB channels. This decomposition process can be done through BSS. Therefore, BSS is used to recover the source signals from the mixed signals. In this approach, the BVP signal is recovered by splitting the mixed RGB signals into separated source signals. In this subsection, different BSS methods are used to analyze the RGB signals to reveal the original source signals. Then, their results, based on our proposed rPPG algorithm, are introduced and compared to decide which one achieves the finest performance.

1) JOINT APPROXIMATION DIAGONALIZATION OF EIGENMATRICES

The Joint Approximation Diagonalization of Eigenmatrices (JADE) is an Independent Component Analysis (ICA) algorithm firstly introduced by Cardoso in [62]. JADE decomposes mixed signals into their original source signals by means of the 4^{th} order moments. It achieves a good numerical performance by including all the cumulants of 2^{nd} and 4^{th} order, and implementing rapid optimization through joint diagonalization. The 4^{th} order moments represent the non-Gaussianity, which defines the independence between the source signals. The most striking feature of JADE over other ICA methods is that it is based on matrix calculation, similar

to PCA. Other ICA methods are based on optimization process. Therefore, their results may vary depending on the following factors: the chosen starting point and the path tracked by the searching technique for optimization. JADE has been widely employed in a variety of applications, where real-data processing is required, such as mobile communication, radars, and biomedical signals, as in our application.

2) FASTICA

FastICA is an Independent Component Analysis (ICA) algorithm. Due to its efficiency and speed, FastICA is one of the most widely used method for ICA; this computationally powerful algorithm was developed by Stone [64]. FastICA exploits a rapid fixed-point iterative method to determine projections that maximize the non-Gaussianity of the signals. For ICA, fixed-point iteration methods are very fast compared to the traditional gradient descent methods. Also, FastICA can be used as an exploratory data analysis tool by performing projection pursuit. Another significant feature of this algorithm is the possibility of parallel execution, making it very suitable for real-time applications.

Independent Component Analysis (ICA) can be described as a special case of BSS. In ICA, the source signals are assumed to be non-Gaussian and, it is described as a linear mixture of statically independent components, i.e.

$$x = M \cdot s \tag{14}$$

For 3-dimension and time dependent vectors,

$$[x_1(t), x_2(t), x_3(t)] = M \cdot [s_1(t), s_2(t), s_3(t)] \tag{15}$$

$$[s_1(t), s_2(t), s_3(t)] = M^{-1}.[x_1(t), x_2(t), x_3(t)] \tag{16}$$

where x is the vector of the mixed signals, s is the vector of source signals, and M is the matrix maximizing the statistical independence of the source signals [64].

There are many ICA techniques available in the literature. The remarkable difference between each of these ICA methods is how the unique feature of the single source signal is identified.

3) KERNEL ICA (kICA)

Kernel ICA (kICA) is an efficient technique for ICA, which was firstly proposed by Bach and Jordan [65]. kICA guesses source signals through optimization of a general variance contrast function, that is founded on kernel Hilbert space. In this contrast function, mutual information is considered as a representation of arithmetical independence. In other words, kICA depends on minimizing the contrast function based on Kernel notions. Therefore, by applying the contrast function to a set of mixed data, we can extract components that are independent as much as possible.

4) SECOND ORDER BLIND IDENTIFICATION (SOBI)

Second Order Blind Identification (SOBI) is another ICA method, proposed by Belouchrani *et al.* [66], for temporally correlated signals. SOBI is founded on mutual diagonalization of an arbitrary group of covariance matrices. The remarkable features of this method are as followings: firstly, it only depends on second-order statistics of the examined data sources; secondly, in contrary to higher order techniques, SOBI allows the decomposition of a set of Gaussian signals; thirdly, the use of numerous covariance matrices significantly improves the robustness of the algorithm.

5) PCA

Lewandowska *et al.* [15] argued that PCA [67] can yield an effective performance as ICA. Therefore, we experimented our method using PCA to compare it with ICA, and determine which one achieves a better performance.

Principle Component Analysis (PCA) is another BSS method; it is a famous technique for extracting features and projecting data. PCA achieves a dimensionality reduction via eliminating redundant variables and projecting the data according to the significant variables, which have the largest contribution towards data variance.

The PCA can be modeled by considering the following transformation:

$$Y = U^T X \tag{17}$$

where Y is the vector that includes the mixed signals, X is the original source signals whose dimension is determined by the number of samples and the number of source signals, and U is the orthonormal matrix whose columns represents the principal components. These component columns are the eigenvectors of the covariance matrix [64]

$$C_X = \frac{1}{N} X X^T \tag{18}$$

VI. RESULTS AND DISCUSSION

A. DATA SETS

In order to verify the validity of our rPPG method, we use a dataset composed of 45 videos [68], these videos show object connected with A CMS50E pulse oximeter to get the ground truth PPG and the objects were asked to sit still. The distance between the object and the camera while recording the video from data set is 1-2 m. Each video is almost 2 minutes long and is recorded with a low cost webcam (Logitech C920 HD pro) at 30 frames per second with a resolution of 640×480 in uncompressed 8- bits RGB format. All subjects are recorded using ambient light.

The length of the sliding window plays a critical role in deciding the best achievable accuracy. The best possible achievable measurements accuracy for a given window length is given by: $df = \frac{60}{T}$, where T is the length of the window. From this relation, the expected accuracy for T=10 s is 6 bpm. Increasing window length improves the accuracy, but it is slower and it requires higher complexity; on the other hand, decreasing window length degrades the accuracy, but it requires less complexity and provides faster measurements, which improves system overall latency. 10-seconds window length provides a reasonable tradeoff between accuracy and complexity.

B. RESULTS

The results of the proposed system with different step variations are reported in this section.

1) COMPARISON BETWEEN THE FACE DETECTION METHODS

To determine which method yields a better performance for our algorithm, we tested our rPPG algorithm using each of the two methods on 5 subjects. The experimental result of this comparison is shown in Table 1. The resulting face bounding box of each of the two methods is shown in Fig. 13. The results of the experiments found clear superiority for Viola-Jones over Liao in detecting faces for our rPPG-based method. The overall this finding is in accordance with findings reported in the most previous studies in the literature. Consequently, we decided to employ Viola-Jones as the face detection technique in our system.

TABLE 1. Comparison between the chosen face detection methods.

RMSE (Total Error)	Viola-Jones (bpm)	Liao (bpm)
Subject 1	**6.0257**	6.1292
Subject 2	**15.4821**	15.6935
Subject 3	**16.45**	19.2038
Subject 4	6.3614	**3.9924**
Subject 5	**11.3376**	13.7888

FIGURE 13. The resulting bounding box of the Viola-Jones method (Left) and Liao method (Right).

2) COMPARISON BETWEEN SKIN SEGMENTATION METHODS

In order to verify the validity of the skin segmentation methods, we conducted our rPPG method using each of the two skin detection methods on 5 subjects. The experimental result of the comparison between these methods is shown in Table 2. The segmented output of each of the two methods is shown in Fig. 14. The results confirm that Conaire's method provides a better RMSE, and it outperforms the second investigated skin detection method. This was an anticipated finding because during the experiments, we reported that Conaire's method was very robust in distinguishing between skin and non-skin areas within the face area of the subjects, which is critical for our application. In general, the Conaire's method was the one that obtained the most robust results; thus, we have chosen it as the skin segmentation technique in our algorithm.

TABLE 2. Comparing the examined skin segmentation methods.

RMSE (Total Error)	Conaire (bpm)	RGB-H-CbCr (bpm)
Subject 1	**6.0257**	7.2584
Subject 2	**15.4821**	16.4827
Subject 3	**16.45**	22.5054
Subject 4	**6.3614**	6.7247
Subject 5	**11.3376**	13.3810

FIGURE 14. Skin segmentation using: Conaire method (Left) and RGB-H-CbCr method (Right).

TABLE 3. Summary of the experimental results for 5 subjects.

RMSE (Total Error)	Proposed	ROI without skin segm. [9]	Face [14]	Cropped [33]
Subject 1	**3.71**	5.095	3.91	11.25
Subject 2	**4.42**	11.07	8.6	15.6
Subject 3	**3.97**	6.06	7.8	8.9
Subject 4	3.1	**2.65**	2.9	8.86
Subject 5	**2.7**	4.30	11.37	15.51

3) COMPARISON BETWEEN ROIS DEFINITION METHODS

To assess the significance of employing skin segmentation, we compare between the proposed scheme with the conventional schemes [14], [33], where skin detection wasn't included. Moreover, we examine the effect of dividing skin-pixels into 3 regions by comparing between our two predefined ROIs, namely ROI 1 and ROI 2. The overall RMSE results for 5 subjects are summarized in Table 3. Also, the comparison between the proposed scheme and the conventional schemes [14], [33], in sense of HR measurement accuracy, is tabulated in Table 4. The ROIs definitions are shown in Fig. 15.

4) COMPARISON BETWEEN ROI TRACKING METHODS

To decide which tracking method provides an improved performance for our algorithm, we implemented our rPPG algorithm using each of the two methods on 5 subjects.

The set of the tracking points for KLT algorithm is shown in Fig. 16 (left). Also, the set of the tracking points for KLT

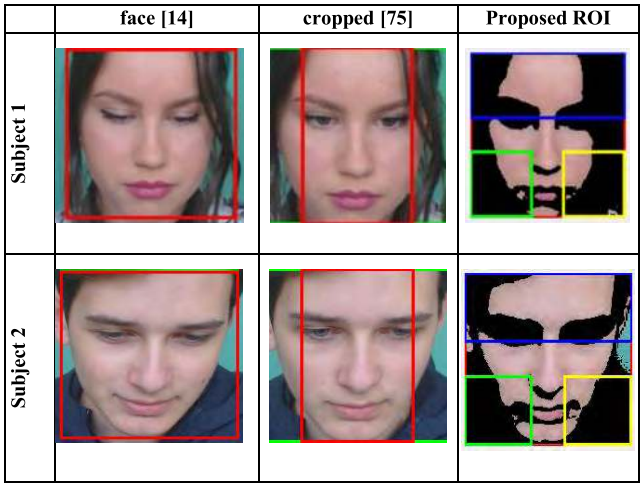

FIGURE 15. ROI definitions.

TABLE 4. Summary of the accuracy of different schemes for 5 subjects

Accuracy (%)	Proposed	ROI without skin segm. [9]	Face [14]	Cropped [33]
Subject 1	**98.63138**	98.39479	98.5995	96.73952
Subject 2	**98.51481**	96.80639	97.59684	94.62
Subject 3	**98.5898**	98.20617	97.8083	97.51239
Subject 4	98.7242	**98.78862**	98.75324	97.52382
Subject 5	**98.78162**	98.53519	96.69416	94.67546

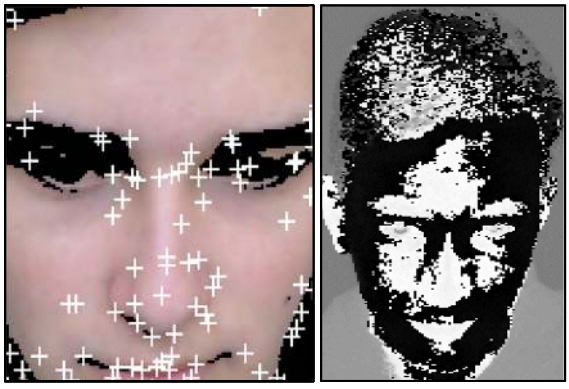

FIGURE 16. The detected trackable facial points for the KLT (left) and CamsShift (right).

algorithm are shown in Fig. 16 (right). The comparison of overall measurement results are summarized in Table 5. It can be seen that slightly superior results are achieved with KLT tracking method. We report that KLT is considered to be more robust for our system. Subsequently, we use it as the face tracking method in our system.

5) COMPARISON BETWEEN THE BSS METHODS

To assess each of the investigated BSS methods, and to determine which one yields a better performance for our algorithm, we tested our rPPG algorithm using: JADE, FastICA, kICA, SOBI, and PCA on 5 subjects. The experimental results of this comparison are shown in Table 6. Interestingly, the experimental results revealed that PCA achieves the best RMSE among all the examined BSS methods, which is in accordance with findings reported by Cardoso [67]. Subsequently, we report that PCA is the most suitable choice for the decomposition technique employed in our system.

TABLE 5. Comparing between KLT and CamShift.

RMSE (Total Error)	KLT (bpm)	CamShift (bpm)
Subject 1	6.0257	**5.6978**
Subject 2	15.4821	**14.9564**
Subject 3	**16.45**	17.6279
Subject 4	**6.3614**	11.0581
Subject 5	**11.3376**	17.6529

TABLE 6. Comparing different BSS methods.

RMSE	JADE (bpm)	FastICA (bpm)	kICA (bpm)	SOBI ICA (bpm)	PCA (bpm)
Subject 1	**5.0523**	6.0257	5.9997	5.8503	5.1862
Subject 3	15.5895	15.4821	**15.3127**	15.8436	15.5645
Subject 4	15.1628	16.45	13.1060	14.2612	**10.3854**
Subject 5	3.1051	6.3614	4.8306	6.7553	**2.6495**
Subject 8	10.1623	11.3376	8.2435	8.4804	**6.5861**

TABLE 7. Comparing between proposed scheme and the conventional schemes in sense of the computational time.

Data set video	Method	Comp.time (s/frame)
1	Cropped [33]	0.152
	Face [14]	0.184
	Proposed ROIs	**0.135**
2	Cropped [33]	0.132
	Face [14]	0.183
	Proposed ROIs	**0.128**

C. COMPUTATIONAL COMPLEXITY

The computational complexity of the proposed scheme is compared with the conventional schemes [14], [33] in this section. The computational complexity is measured as the execution time for each scheme. The execution times are measured on MATLAB R2014a platform using Intel®core(TM) i5-2500 Cpu 3.30 GHz. The computational time is shown in Table 7. From this table, it can be shown that thanks to the exclusion of the non-skin pixels from the computation, the overall computation time is reduced compared to the cases of whole face [14] and using cropped face [33]. This result can be justified by the number of the required operations for each scheme per iteration which is tabulated in Table 8. This table shows that the number of the required operations including additions, multiplications, subtractions and divisions is reduced in the proposed scheme compared to schemes [14] and [33].

D. DISCUSSION

Our attention is focused not only on extracting skin-pixels but also on which part of the facial skin to be used for

TABLE 8. The required number of operation for different methods per iteration.

Operation	Proposed ROIs	Face [14]	Cropped [33]
Addition	19	25	30
Subtraction	22	32	36
Multiplication	44	50	56
Division	33	40	42

signal extraction. The results demonstrate two findings. First finding is that the skin segmentation is an effective way to improve the accuracy of the rPPG technique by filtering out non-skin pixels, which is consistent with that found in [68]. The second finding is that extracting the rPPG signal from ROI with segmentation provides a better accuracy than extracting it from ROI without segmentation.

The second finding has some exceptions as in subject 4 in Tables 3 and 4. As has been reported in [15], [33], [50] and [69] that the rPPG signal is not homogeneously distributed on the face, and the best facial regions for rPPG signal extraction are the forehead and cheeks. The most likely explanation of this exception is that the effect of RGB quantization error, which cannot be neglected when the number of skin-pixels is small. We observed that ROI down sampling has a negative effect on the rPPG signal. This observation agrees with the fact that the strength of camera sensor noise is inversely proportional to the square root of the sum of pixels included in the averaging step. We expect that using a higher resolution camera may lead to a better signal extraction from ROI with segmentation.

E. PROS. AND CONS. OF THE PROPOSED SCHEME

The main advantages for the proposed scheme are represented in the accuracy of the HR monitoring, where it outperforms the conventional algorithms in [14] and [33]. Moreover, due to the calculation of the HR from the segmented skin pixels; the required computation time is less than the conventional methods.

On the other side, the main disadvantage of the proposed scheme is that it requires a segmentation step as a pre-processing.

VII. CONCLUSION

This work is undertaken to optimize rPPG for continuous and unobtrusive monitoring conditions. In this paper, adaptive skin detection for ROI definition in rPPG is examined, and then it is compared with our previous real-time rPPG-based method. The most obvious finding to emerge from this study is that segmenting facial skin improves the RMSE by nearly 50%. This finding enhances our understanding that only skin pixels contribute constructively to the rPPG signal.

Additionally, we inspected the effect of dividing the face skin into three ROIs, and then extracted the HR from these three ROIs through signal fusion. Our experiments reveal that extracting the HR from all the face yields a better RMSE than extracting it through signal fusion from different facial regions. Also, because the strength of camera sensor noise is inversely proportional to the square root of the sum of pixels included in the averaging step, we encourage future works, interested in employing signal fusion, to use higher resolution cameras which may result in a better signal extraction.

Moreover, we examined the effect of using different methods for face detection, face tracking, skin detection, and blind signal separation. Subsequently, we compared these methods to decide which one achieves better performance for our method. The evidence from these comparisons suggest that choosing suitable methods for BSS and skin detection plays a critical role in deciding the accuracy of the HR measurements.

Finally, we compared our rPPG measurements with ground truth values obtained from a commercial pulse oximeter. From the outcome of our investigation, it is possible to conclude that skin segmentation significantly improves the quality of rPPG signals.

REFERENCES

[1] S. S. Ulyanov and V. V Tuchin, "Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes," *Proc. SPIE*, vol. 1884, no. 1, pp. 160–168, 1993.
[2] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "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.
[3] K. Aminian, X. Thouvenin, P. Robert, J. Seydoux, and L. Girardier, "A piezoelectric belt for cardiac pulse and respiration measurements on small mammals," in *Proc. 14th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Paris, France, Oct./Nov. 2011, pp. 2663–2664.
[4] F. P. Wieringa, F. Mastik, and A. F. W. van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: A first step toward 'SpO_2 camera' technology," *Ann. Biomed. Eng.*, vol. 33, no. 8, pp. 1034–1041, 2005.
[5] J. E. Parra and G. da Costa, "Optical remote sensing of heartbeats," *Proc. SPIE*, vol. 4368, pp. 113–121, Aug. 2001.
[6] Q. Zhang, Q. Wu, Y. Zhou, X. Wu, Y. Ou, and H. Zhou, "Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers," *Measurement*, vol. 100, pp. 311–321, Mar. 2017.
[7] J. Kuo, S. Koppel, J. L. Charlton, and C. M. Rudin-Brown, "Evaluation of a video-based measure of driver heart rate," *J. Saf. Res.*, vol. 54, pp. 55–59, Sep. 2015.
[8] H. Rahman, S. Barua, and B. Shahina, "Intelligent driver monitoring based on physiological sensor signals: Application using camera," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst. (ITSC)*, Las Palmas, Spain, Sep. 2015, pp. 2637–2642.
[9] R. M. Fouad, A. Onsy, and O. A. Omer, "Improvement of driverless cars' passengers on board health and safety, using low-cost real-time heart rate monitoring system," in *Proc. 24th IEEE Int. Conf. Automat. Comput.* Newcastle upon Tyne, U.K.: Newcastle Univ., Sep. 2018, pp. 545–550.
[10] A. B. Hertzman and C. R. Spealman, "Observation on the finger volume pulse recorded photoelectrically," *Amer. J. Physiol.*, vol. 119, no. 1937, pp. 334–335, 1937.
[11] N. E. Almond and E. D. Cooke, "Observations on the photoplethysmograph pulse derived from a laser Doppler flowmeter," *Clin. Phys. Physiol. Meas.*, vol. 10, no. 2, pp. 137–145, 1989.
[12] H. Ugnell and P. Å. Öberg, "The time-variable photoplethysmographic signal; dependence of the heart synchronous signal on wavelength and sample volume," *Med. Eng. Phys.*, vol. 17, no. 8, pp. 571–578, 1995.
[13] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. 1–39, Feb. 2007.
[14] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
[15] M. Lewandowska, J. Rumi ski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a Webcam—A non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, 2011, pp. 405–410.

[16] D. Wedekind, H. Malberg, S. Zaunseder, F. Gaetjen, K. Matschke, and S. Rasche, "Automated identification of cardiac signals after blind source separation for camera-based photoplethysmography," in *Proc. IEEE 35th Int. Conf. Electron. Nanotechnol. (ELNANO)*, Kyiv, Ukraine, Apr. 2015, pp. 422–427.

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

[18] W. Wang, S. Stuijk, and G. de Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1974–1984, Sep. 2016.

[19] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 99, pp. 1479–1491, Jul. 2017.

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

[21] E. J. va Kampen and W. G. Zijlstra, "Determination of hemoglobin and its derivatives," *Adv. Clin. Chem.*, vol. 8, pp. 141–187, Jan. 1966.

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

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

[24] K.-Z. Lee, P.-C. Hung, and L.-W. Tsai, "Contact-free heart rate measurement using a camera," in *Proc. 9th Conf. Comput. Robot Vis.*, 2012, pp. 147–152.

[25] L. Wei, Y. Tian, Y. Wang, T. Ebrahimi, and T. Huang, "Automatic webcam-based human heart rate measurements using Laplacian eigenmap," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 281–292.

[26] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[27] Y. Sun, S. Hu, R. Kalawsky, S. Greenwald, and V. Azorin-Peris, "Non-contact imaging photoplethysmography to effectively access pulse rate variability," *J. Biomed. Opt.*, vol. 18, no. 6, p. 061205, Oct. 2012.

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

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

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

[31] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2593–2601, Oct. 2014.

[32] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014.

[33] Y. Hsu, Y.-L. Lin, and W. Hsu, "Learning-based heart rate detection from remote photoplethysmography features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4433–4437.

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

[35] B. Chwyl, A. G. Chung, J. Deglint, A. Wong, and D. Clausi, "Remote heart rate measurement through broadband video via stochastic Bayesian estimation," *J. Comput. Vis. Imag. Syst.*, vol. 1, p. 1, Oct. 2015.

[36] D. N. Tran, H. Lee, and C. Kim, "A robust real time system for remote heart rate measurement via camera," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.

[37] M.-C. Li and Y.-H. Lin, "A real-time non-contact pulse rate detector based on smartphone," in *Proc. Int. Symp. Next-Gener. Electron. (ISNE)*, 2015, pp. 1–3.

[38] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," Carnegie-Mellon Univ. Robot. Inst., Tech. Rep., Sep. 1981.

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

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

[41] M. Villarroel, J. Jorge, C. Pugh, and L. Tarassenko, "Non-contact vital sign monitoring in the clinic," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May/Jun. 2017, pp. 278–285.

[42] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomed. Signal Process. Control*, vol. 31, pp. 309–320, Jan. 2017.

[43] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Trans. Instrum. Meas.*, to be published.

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

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

[46] S. Hu, V. A. Peris, A. Echiadis, J. Zheng, and P. Shi, "Development of effective photoplethysmographic measurement techniques: From contact to non-contact and from point to imaging," in *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Sep. 2009, pp. 6550–6553.

[47] M. Kumar, J. Suliburk, A. Veeraraghavan, and A. Sabharwal, "PulseCam: High-resolution blood perfusion imaging using a camera and a pulse oximeter," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 3904–3909.

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

[49] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Aug. 2015, pp. 6398–6404.

[50] L. Feng, L. M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

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

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

[53] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Kauai, HI, USA, vol. 1, Dec. 2001, pp. I-511–I-518.

[54] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.

[55] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–6. doi: 10.1109/CVPR.2007.383448.

[56] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Procces.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[57] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3, May 1998, pp. 1381–1384.

[58] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 211–223, Feb. 2016.

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

[60] G. R. Bradski, "Real time face and object tracking as a component of a perceptual user interface," in *Proc. 4th IEEE Workshop Appl. Comput. Vis. (WACV)*, Oct. 1998, pp. 214–219.

[61] N. A. B. A. Rahman, K. C. Wei, and J. See, "RGB-H-CbCr skin colour model for human face detection," in *Proc. MMU Int. Symp. Inf. Commun. Technol. (M2USIC)*, 2006, pp. 90–96.

[62] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, Jan. 1999.
[63] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 5, no. 3, pp. 3917–3920, Apr. 1997.
[64] J. V. Stone, "Independent component analysis: An introduction," *Trends Cogn. Sci.*, vol. 6, no. 2, pp. 59–64, Feb. 2002.
[65] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, Jan. 2002.
[66] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
[67] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
[68] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognit. Lett.*, vol. 124, pp. 82–90, Jun. 2017.
[69] L. Feng, L.-M. Po, X. Xu, and Y. Li, "Motion artifacts suppression for remote imaging photoplethysmography," in *Proc. Int. Conf. Digit. Signal Process. (DSP)*, Aug. 2014, pp. 18–23.

R. M. FOUAD received the B.Sc. degree in electronics and communications engineering from the Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Aswan, Egypt, in 2014, the M.Sc. degree in electronics and communications engineering from Aswan University, Aswan, Egypt, in 2018. He is currently pursuing the Dr. Phil. degree in electrical and computer engineering with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada. His current research interests include 5G networks, digital communications, computer vision, embedded systems, and software defined radio

OSAMA A. OMER received the B.Sc. and M.Sc. degrees from South Valley University, in 2000 and 2004, respectively, and the Ph.D. degree from the Tokyo University of Agriculture and Technology, in 2009. He spent six month as a Postdoctoral Researcher with Medical Engineering Department, Luebeck University, Germany, three month as a Postdoctoral Researcher with Kyushu University, Japan, and six months as a Research and Development Scientist Engineer with NOKIA Research and Development Center, Tokyo, Japan, in 2008. His research interests include medical imaging, super-resolution, image/video coding, and wireless communications. He is currently an Associate Professor with the Electrical Engineering Department, Aswan University, and he is the Head of Electronics and Communications Department, Arab Academy of Science, Technology and Maritime Transport, South Valley branch.

MOUSTAFA H. ALY received the B.Sc., M.Sc., and Ph.D. degrees in communications from Alexandria University, Egypt, in 1976, 1983, and 1987, respectively. He is a Professor with the College of Engineering and Technology, Arab Academy for Science, Technology, and Maritime. He has published more than 250 papers and has supervised more than 130 M.S. and Ph.D. graduate students. His current research includes include optical fiber communications, free space optics, and optical networks.