IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 2, 2021, accepted August 15, 2021, date of publication August 19, 2021, date of current version September 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106046

Reduction of Motion Artifacts From Remote Photoplethysmography Using Adaptive Noise Cancellation and Modified HSI Model

DONGRAE CHO[1], (Member, IEEE), JONGIN KIM[1], (Member, IEEE), KWANG JIN LEE[1], (Member, IEEE), AND SAYUP KIM[2], (Member, IEEE)

[1]Deepmedi Research Institute of Integrated Technology, Deepmedi Inc., Seoul 06232, South Korea
[2]Digital Transformation Research and Development Department, Korea Institute of Industrial Technology, Ansan 15588, South Korea

Corresponding authors: Dongrae Cho (dongrae30@deep-medi.com) and Sayup Kim (sayub@kitech.re.kr)

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE), South Korea, as a product and service design of recliner chair, including smart life care service, under Grant 20007942.

ABSTRACT Remote photoplethysmography (rPPG) is a method to measure cardiac activities without any contact sensors. Non-contact sensors include radar, laser, and digital cameras, and there have been wide developments regarding the measurement of rPPG signals using continuous face frames. However, non-contact sensors are quite sensitive to the subject's motion, which causes motion artifacts. In this paper, two hypotheses are proposed: a) the motion artifacts are caused by unevenly reflected light due to the curvature of the subject's face; and b) melanin and residuals in the continuous face frames are time-varying values whenever the subject's movement is triggered. Adaptive noise cancellation based on recursive least square (ANS based on RLS) using the Lambert-Beer law and the hue–saturation–intensity (HSI) model were applied. The former is used for skin modeling, and the latter is used to reduce noises derived by the curvature of the face. Furthermore, the proposed algorithm is directly applied to two-dimensional continuous face frames and results in the rPPG signal and rPPG image, respectively. To evaluate proposed algorithm, two different experiments (e.g., static and dynamic situation) were conducted. Furthermore, in a study with 15 participants, the performances of heart rate estimation and heart rate variability (HRV) were evaluated by comparing the proposed method with previously developed methods. The results showed that a) the artifacts derived by head movement are efficiently removed, compared to previous methods; and b) rPPG images describing the spread of facial blood flow are acquired in real-time.

INDEX TERMS Adaptive noise cancellation (ANC), heart rate variability (HRV), HSI color model, photoplethysmography (PPG), recursive least square (RLS), remote sensing, skin modeling.

I. INTRODUCTION

Quantitative understanding of cardiac activity is an important part of individual health management in the modern society. Many medical devices have been developed to understand individual health and to acquire their accurate biosignals. The electrocardiogram (ECG) and photoplethysmography (PPG) are two such methods that can monitor cardiac activities without invasive procedures. Especially, heart rate variability (HRV), which is calculated from the change in the beat interval of ECG and PPG, is a useful tool to investigate physiological phenomena [1]. In general, since the ECG shape is derived by electrical changes in the heart, the HRV calculated using ECG is used as a reference signal. However, at least three channels are required to measure the ECG, and it is very sensitive to the magnetic field of the external environment. Alternatively, HRV calculated using PPG has been widely used since the cardiac activity can be monitored by only one PPG sensor attached to the index finger. In addition, many studies have demonstrated that HRVs calculated using PPG and ECG are not significantly different [2]–[4].

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

The shape of the PPG curve is induced by changes to the blood circulating through the index finger, and blood flow in the vessels is detected using the optical characteristics of the skin [5]–[7]. Those optical characteristics also apply to changes in facial skin color as well [7]–[9]. Several research groups have developed techniques to identify changes in blood movement caused by heart activities by measuring the changes in facial skin color using a webcam; the technology utilized to estimate the cardiac activities using only facial images is called remote-PPG (rPPG) [9]–[12]. The skin images (recorded by a webcam) capture the skin color and reflective light. Although a skin image seems to represent only the external surface of the human body, it also reflects the activity of the capillary circulation system.

There are two limitations in measuring accurate rPPG information. First, as the variation of blood flow is relatively small under the skin, it cannot be visible to the naked eye. However, the information of blood flow change is widely distributed in the pixels of the human skin image. Few pixels are not enough for tracking the blood flow changes. To estimate precise blood flow changes, sequential and stable two-dimensional (2D) images recorded using a video camera (with sufficient pixels of human face skin) are required. Second, the motion artifacts induced by face motion can interfere with an accurate estimation of rPPG information [13]. Since the motion artifacts induced by face motion can be mixed with pure rPPG information, a filtering process is required to extract it [14]–[16].

Our study is based on the two important assumptions. First, from the perspective of the face recording camera, the facial images acquired from webcam include not only the blood flow change but also unexpected signals such as shadow and uneven reflected light triggered by curvature of the face. Such unexpected signals can trigger the motion artifacts whenever the position of face is changed [7], [17]. Fig. 1 illustrates the process of face image acquisition when the angles (α and β) of the face are changed. To reduce the effects of irregular intensities, the hue–saturation–intensity (HSI) model, which is a transformation of the red–green–blue (RGB) model is applied [7]. In general, the intensity of the HSI model is related to the image brightness, it is necessary to change the uneven intensity of HSI to the evenly distributed intensity of HSI. Second, from the perspective of the face recording camera, the colors of the hemoglobin, melanin, and residual pigments of the facial skin are time-varying values influenced by the heart activity and continuous movements [5], [18]. In other words, if the noises derived from the melanin and residual pigments can be reduced, pure rPPG information can be obtained because only hemoglobin is related to the blood flow changes. Therefore, the goal of this study is to apply adaptive noise cancellation (ANC) based on recursive least squares (RLS) to each pixel [14], [15].

II. RELATED WORKS

Previous studies have aimed at estimating rPPG signals using the continuous face frames. Takano *et al.* proposed a method

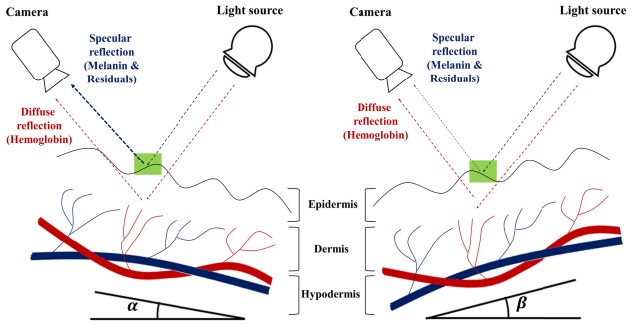

FIGURE 1. Face image acquisition based on the two different angles (α and β) of face skin. The square filled green represents the face skin target. The arrow thickness represents the intensity strength of the reflected light. Although the camera can capture the same face skin area, the reflected light differs according to the face curvature.

using a simple average of the image brightness [19]. It was possible to estimate the heart rate from the rPPG signal using the autoregressive spectral analysis as they assumed that the change of image brightness could reflect the blood flow alteration. Xu *et al.* proposed a human skin model and used the Beer–Lambert law to estimate the rPPG signal [20], [21]. They assume that the human skin consists of three layers (epidermis, dermis, and subcutaneous fat tissue), and the melanin and hemoglobin are distributed in the epidermis and dermis, respectively. Given that the melanin components are time-invariant, the absorption change should depend on the movement of hemoglobin and can be defined as a rPPG signal. Poh *et al.* represented RGB human skin images using a model of mixed sources [8]. To solve this problem, blind source separation (BSS) approach was applied. In the BSS approach, the independent component analysis (ICA) algorithm is used for extracting the independent components (ICs) from randomly mixed multivariate signals. In the ICs, the periodic signal can be defined as a rPPG signal. Wu *et al.* proposed the Eulerian video magnification method to identify changes in the human skin color from videos [22]. They applied the spatial–temporal filtering to the sequences of face video images for each pixel. Consequently, the rPPG image and the signal are acquired simultaneously. However, as this method focuses on rPPG images, the estimated rPPG signal can be distorted or delayed compared with the actual PPG signal. Wang *et al.* represented the Plane-Orthogonal-to-Skin (POS) algorithm defining a plane orthogonal to the skin-tone in the temporally normalized RGB space for a rPPG signal [12]. Since the signal extracted from POS algorithm is orthogonal to the plane of skin-tone, the signal is less sensitive to the skin-tone but more to the blood flow changes.

Recently, deep learning technology has been developed with advanced computer performance and huge cloud services. Earlier, deep learning was only used for image classification and analysis using the vest 2D images [23]–[26]. Researchers use deep learning not only for face recognition [27]–[29] but also for rPPG signal extraction [9], [30], [31]. Chen *et al.* and Yu *et al.* proposed DeepPhys

and rPPGNet, respectively [9], [32]. These deep learning models consisted of a convolutional neural network (CNN). DeepPhys can calculate the motion representation based on skin reflection model and train the difference between frames using only the attention mechanism [32]. rPPGNet uses pulse signal and labeled skin regions as the input data when learning and its performance is evaluated using video compression [9]. Lee *et al.* proposed the Meta-rPPG, which composed of a feature extractor and PPG estimator modeled on a CNN and long short-term memory (LSTM) [31]. Since the distributional changes in the face images are unpredictable, a transductive meta-learner is developed by obtaining unlabeled samples and constructing a self-supervised weight adjustment. Yu *et al.* proposed the PhysNet which is a combination of the three-dimensional CNN (3DCNN) based network and Recurrent Neural Network (RNN) based network [30]. 3DCNN based networks can simultaneously extract the semantic pulse features in spatial and temporal domain. Moreover, the RNN based networks directly estimate the pulse signals using calculated features.

In the medical domain, motion artifact removal techniques have been used not only for 2D medical images such as X-ray, computed tomography (CT), and magnetic resonance imaging (MRI), but also for 1D time series signals such as electroencephalogram (EEG), electromyography (EMG), ECG and PPG. In particular, 1D time series signals acquired from contact-type biosensors are essential to apply the motion artifact filtering process. Unlike the 2D medical image that is acquired in a restricted environment, 1D time series signals are acquired from the wearable medical devices or contact-type biosensors that may be damaged by sudden physical movements or unexpected variation in the external environment.

Digital filters, configured by a linear time invariant (LTI) system and Fast Fourier Transform (FFT) for searching motion artifact frequency ranges, have been used for removing motion [30], [33]. It is widely known that the frequency range ($<$0.2 Hz) of motion artifacts is lower than the frequency range ($>$1Hz) of pulse signals generated by cardiac activities. Additionally, for measuring accurate frequency ranges of motion artifacts, some research groups have attached accelerator sensors to their wearable system [35]. However, it is difficult to remove noise with a similar frequency range to that of the biosignal since the digital filter based on the LTI system. To solve this problem, adaptive filtering, such as least mean square (LMS), normalized least mean square (NLMS), and RLS, that changes the filter coefficients based on the input signal has been developed [14]. Since the measured biosignals are non-stationary signals, it is essential to construct a filtering system that changes with time. However, as adaptive filtering models are mainly designed for 1D time series signals such as ECG, PPG, EMG, and EEG, it is difficult to make the 2D face image as an input matrix of the adaptive filtering model.

To solve this problem, the dimensions of continuous face frames (width $\times$ height $\times$ time) are reduced by averaging each 2D image (1 $\times$ time) [36]–[38], but there are two problems with this approach. First, each point in the dimension-reduced face frame cannot guarantee the spatial information of the facial image. From the camera's perspective, motion artifacts are induced by face movements generated by changing each pixel value. Since the cause of the change in each pixel's value is different (e.g., facial shape), it is essential to apply a filtering process to each pixel. Second, unexpected spatial information such as the background of the facial image and subject's hair can be mixed while simply reducing the dimensions of continuous facial frame by averaging.

In case of deep learning processes such as CNN, LSTM, and RNN, the 2D facial images can be used as the input signal and the aforementioned problems can be solved [9], [30]–[32]. However, for deep learning models that use many continuous face frames as inputs, a large number of GPUs and considerable time are required during training. Furthermore, there are space-time constraints and complex computations for calculating rPPG information. To solve this problem, it is essential to construct a process that extracts rPPG information from each pixel value without heavy computation. In addition, to accurately estimate the rPPG signal, the filtering process is applied to only the facial component of the image as the remaining information prevents the approximation. In this paper, RLS-based ANC configured by matrix operations is applied to each pixel of 2D continuous facial frames, and the artifacts derived by face curvature are reduced by applying the modified HSI model as well. In other words, the proposed method (combination modified HSI and RLS based ANC) can help overcome the limitations of the traditional filtering process and solve the aforementioned problems such as omission of spatial information, inclusion of unnecessary pixel values and heavy computations.

III. METHOD

A. DATA COLLECTION

To track the blood flow changes using continuous face frames, an image acquisition system with a low-cost webcam (HD webcam C615, Logitech; RGB 640 $\times$ 480 pixels) was developed. Additionally, to measure the PPG signal from the index finger, a photoplethysmogram (RP320, Laxtha Corp) sensor was used, and this signal represented the reference signal. The PPG signals and continuous face video images were simultaneously acquired at 15 Hz sampling frequency.

For time synchronization between the PPG signal and continuous face frames, an external clock was configured. Whenever the PPG signal and continuous face frames were acquired, the external clock recorded the acquisition time points, and each time point was used for the synchronization between two signals. Finally, since the sampling rate of the rPPG signal is lower than the sampling rate of the PPG signal, a cubic spline method for interpolating missing data was applied. Fig. 2 illustrates the block diagram for the data acquisition and synchronization. To acquire continuous

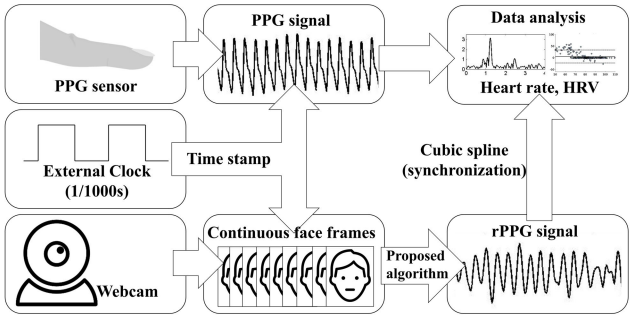

FIGURE 2. Block diagram depicting data acquisition of PPG and continuous face frames and synchronization procedure between PPG and estimated rPPG signal. A cubic spline method that interpolates the omitted data position is applied to the rPPG signal.

face frames, the Viola–Jones face detector, known for its face detection algorithm based on Haar-like features, was used [39]. Fifteen participants (males: 13, females: 2) were recruited. The average age of the participants was 27.1 years (standard deviation (SD): ±1.98). The subjects having cardiovascular diseases such as arrhythmia, ventricular tachycardia and bradycardia were excluded. The study was approved by the Public Institution Bioethics Committee designated by the Ministry of Health and Welfare of South Korea (IRB P01-201812-12-001). A written informed consent was obtained from each participant, all methods were performed in accordance with the relevant guidelines and regulations, and the whole procedure including the experimental protocol was conducted in agreement with the Public Institution Bioethics Committee designated by the Ministry of Health and Welfare of South Korea. Informed consent to publish identification images (face) was obtained from the concerned subject.

To evaluate the performance of the adaptive algorithms, two types of experiments were conducted. First, the participants did not move their faces in the image frame for three minutes. This dataset is called ''static images''. Second, the participants moved in two different manners: they shook their head from side to side (movement left and right side within a 60° angle on the transverse plane) and nodded their head up and down (movement top and bottom side within a 45° angle on the sagittal plane). A combination of these two movements were realized over three minutes; refer to the dataset of this second part as the ''motion artifact corrupted images (MA-corrupted images).'' The images are corrupted because the head motion can induce artifacts in the continuous image frame, which can disturb the accurate tracking of blood flow changes.

B. MODIFIED HSI MODEL

To reduce the motion artifacts, the color space HSI model, derived by transforming the RGB model, was modified. It stands for *hue* (H), *saturation* (S), and *intensity* (I) [17]. The I-plane in the HSI model represents the intensity of each pixel on a scale from 0 to 1. I-plane values of 0 and 1 indicate that the image comprising RGB has no and maximum intensity, respectively. To evenly distribute intensity across a skin image, the value of 1 is considered. In the HSI model, the value of 1 means that each pixel has the maximum intensity and minimizes the effect of noise generated by the face shape. Subsequently, a modified HSI model is inversely transformed into the RGB model. Finally, the irregular intensity of the light reflected from the face surface can be changed to a regular intensity. Fig. 3 illustrates the overall HSI modification (HSI(1)) method.

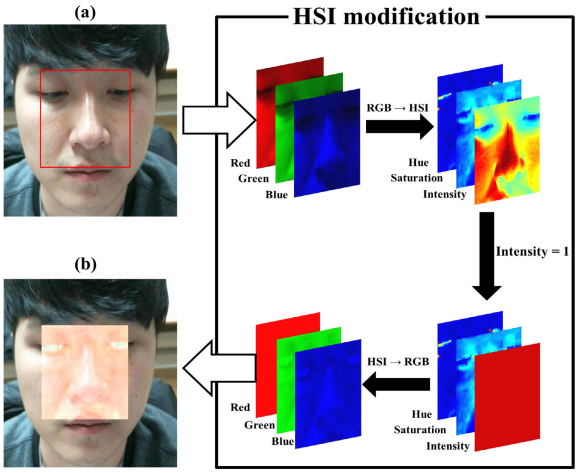

FIGURE 3. Basic scheme of the HSI modification. Three steps: (1) transformation from RGB to HSI (RGB → HSI); (2) change of I plane as 1 (Intensity = 1); (2) transformation of HSI to RGB (HSI → RGB). (a) Original face image. The red box represents the ROI. (b) Image after applying the HSI modification method.

In this study, since the wavelength bands of deoxyhemoglobin and hemoglobin are 520 and 590 nm, respectively, which are close to the wavelength of the G-plane (wavelength = 546.1 nm) of the RGB model, the G-plane is used as the input signal for the proposed method [40], [41].

C. RECURSIVE LEAST SQUARES (RLS)

Recursive least squares (RLS) is an adaptive filtering algorithm that recursively finds unknown filter coefficients by minimizing the weighted linear least squares error function. The convergence time of RLS is lesser than the least mean squares (LMS), where filter coefficients can be updated using the mean square error. In addition, the LMS has a greater step size compared to that of the RLS. Consequently, the LMS has a high computational load. In other words, the filtering performance of RLS is superior to other adaptive filtering methods (e.g., LMS and normalized LMS) when compared using the same computation load. Therefore, RLS is the best method for a non-stationary environment compared to other adaptive filtering methods [14], [15].

Let $\boldsymbol{i}_{x,y}(n) = [i_{x,y}(n), i_{x,y}(n-1), \ldots, i_{x,y}(n-p+1)]^T$ be the input signal vector, where x and y are coordinates, n is a time index, and p is the order in the two-dimensional input image. In addition, input signal value $i_{x,y}(n)$ is a sum of the desired signal $d_{x,y}(n)$ and noise signal $v_{x,y}(n)$. The estimated noise signal $\hat{v}_{x,y}(n)$ is calculated and the error function $e_{x,y}(n)$

as follows:

$$\hat{v}_{x,y}(n) = \hat{\boldsymbol{w}}_{\boldsymbol{x,y}}(n)^T \boldsymbol{i}_{x,y}(n), \tag{1}$$

$$e_{x,y}(n) = i_{x,y}(n) - \hat{v}_{x,y}(n) = \hat{d}_{x,y}(n), \tag{2}$$

where $\hat{d}_{x,y}(n)$ and $\hat{\boldsymbol{w}}_{\boldsymbol{x,y}}(n) = [\hat{\boldsymbol{w}}_{\boldsymbol{x,y}}{}^{\boldsymbol{0}}(n)\ \hat{\boldsymbol{w}}_{\boldsymbol{x,y}}{}^{\boldsymbol{1}}(n), \ldots, \hat{\boldsymbol{w}}_{\boldsymbol{x,y}}{}^{\boldsymbol{p-1}}(n)]^T$ are the estimated desired signal and estimated filter coefficients, respectively. The weighted linear least squares error is calculated as follows:

$$\Sigma_{i=0}^{n} \lambda^{n-i} |e_{x,y}(n)|^2 = 0, \tag{3}$$

where λ is the forgetting factor. The updated RLS equations are represented as follows:

$$\mathbf{g}_{x,y}(n) = \frac{\lambda^{-1}\mathbf{R}_{x,y}^{-1}(n-1)\,\boldsymbol{i}_{x,y}(n)}{1+\lambda^{-1}\boldsymbol{i}_{x,y}^{\mathrm{T}}(n)\mathbf{R}_{x,y}^{-1}(n-1)\,\boldsymbol{i}_{x,y}(n)}, \tag{4}$$

$$\alpha_{x,y}(n) = d_{x,y}(n) - \hat{\boldsymbol{w}}_{x,y}(n)^T \boldsymbol{i}_{x,y}(n), \tag{5}$$

$$\hat{\boldsymbol{w}}_{x,y}(n+1) = \hat{\boldsymbol{w}}_{x,y}(n) + \alpha_{x,y}(n)\,\mathbf{g}_{x,y}(n), \tag{6}$$

where $\boldsymbol{R}_{x,y}$ is the autocorrelation matrix of the input vector $\boldsymbol{i}_{x,y}(n)$, $\alpha_{x,y}(n)$ is a priori error, and $\mathbf{g}_{x,y}(n)$ is the gain vector. Fig. 4 illustrates the basic configuration of the RLS-based ANC method. Consequently, the estimated desired signal $\hat{d}_{x,y}(n)$ is equal to the error function $e_{x,y}(n)$ of the ANC in (2). Therefore, it is essential to determine the noise signal $v_{x,y}(n)$, which interferes with the estimation of an accurate desired signal $d_{x,y}(n)$.

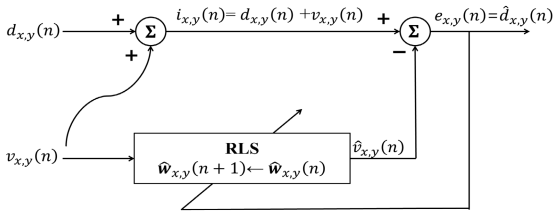

FIGURE 4. Basic configuration of the RLS-based ANC.

D. HUMAN SKIN MODELING FROM AN IMAGE

The human skin is a complex multilayer organic structure that consists of stratum corneum, epidermis, dermis, and subcutaneous fat tissues. When light shines on the human skin, a part of it is absorbed by melanin and hemoglobin distributed throughout the skin. Melanin, which is a natural chromophore on the epidermis, contributes to the skin color, and its location is fixed. However, hemoglobin is a part of red blood cells that transfer oxygen from the respiratory organs to the tissues of the entire body.

To describe this phenomenon mathematically, the Beer–Lambert law, which relates light attenuation to the material properties, was applied [21]. Representing the skin absorbance, wavelength, and time index as A, λ, and n, respectively, the modified Beer–Lambert law can be expressed as

$$A(\lambda, n) = v^m(\lambda)\, c^m + v^h(\lambda, n)\, c^h + A^0(\lambda), \tag{7}$$

where powers m and h represent melanin and hemoglobin, respectively. In addition, v is the product of the pigment extinction coefficient and the mean path length of photons in the skin layer, c is the pigment concentration, and A^0 is the baseline absorbance of skin and residual pigments [21]. The values of v^m, c^m, c^h and A^0 are time-invariant and only v^h changes with time index n. Thus, a change in skin absorbance A reflects a change in blood flow (7). Nevertheless, whenever the position of the human face is changed, the values of constants c^m, c^h, and A^0 may change, which introduces unexpected motion artifacts. Therefore, (7) is modified to reduce the motion artifact effect and is expressed as

$$A(\lambda, n) = v^m(\lambda, n)\, c^m + v^h(\lambda, n)\, c^h + A^0(\lambda, n), \tag{8}$$

where the values of v^m and A^0 which are time-invariant variables in (7), are changed to time-variant ones. In other words, the motion artifacts induced by the movement of the head can be represented by (8). In addition, absorbance can be represented as

$$A = -\log(L/E), \tag{9}$$

where L is the power of the transmitted light, and E is the power of the incident light. By combining (8) and (9),

$$L(\lambda, n) = E\exp\{-(v^m(\lambda, n)\, c^m + v^h(\lambda, n)\, c^h + A^0(\lambda, n))\} \tag{10}$$

is acquired. Additionally, the pixel intensity at coordinates x and y from (10) is represented as

$$i_{x,y}(n) = G\int L_{x,y}(\lambda, n)\, \mathrm{S}(\lambda) d\lambda, \tag{11}$$

where i is the pixel intensity, G is the camera gain, and S is the spectral response function. Assuming that the spectral response function S is a Kronecker delta function, (11) can be expressed as

$$i_{x,y}(n) = GE\exp\{-(v_{x,y}^m(n)\, c_{x,y}^m + v_{x,y}^h(n)\, c_{x,y}^h + A_{x,y}^0(n))\}. \tag{12}$$

Here, (12) implies that the change in intensity i of an image is related to the change in the absorbance of melanin, hemoglobin, and residual pigment. By applying natural logarithm to both sides of (12), it becomes

$$\ln i_{x,y}(n) = \ln GE - \left(v_{x,y}^m(n)\, c_{x,y}^m + v_{x,y}^h(n)\, c_{x,y}^h + A_{x,y}^0(n)\right). \tag{13}$$

Differentiating (13) with respect to n,

$$-\frac{i'_{x,y}(n)}{i_{x,y}(n)} = v_{x,y}^{m'}(n)\, c_{x,y}^m + v_{x,y}^{h'}(n)\, c_{x,y}^h + A_{x,y}^{0'}(n), \tag{14}$$

where apostrophe ($'$) indicates that the differentiation operation is acquired. As the camera gain G and incident light power E are constants, the differentiation of ln GE with respect to n equals zero. The changes in absorption of

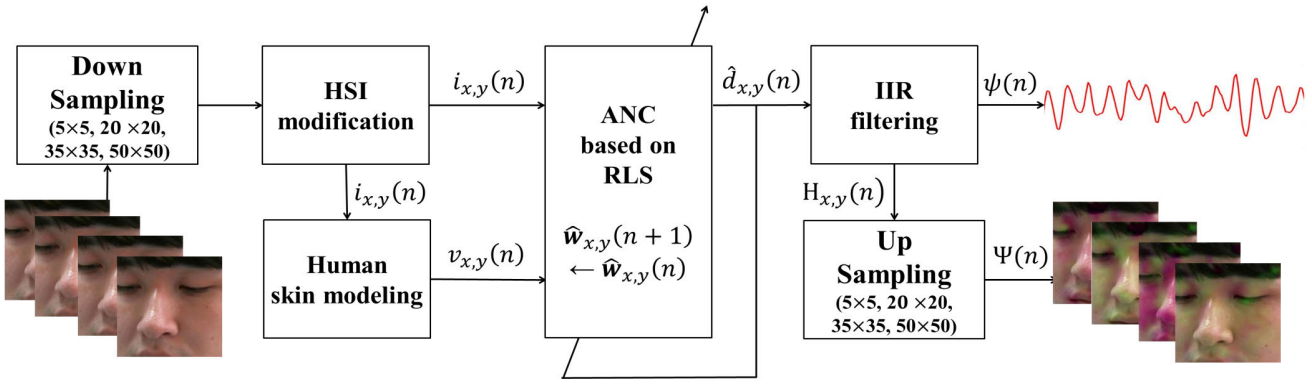

FIGURE 5. Block diagram for estimating the rPPG signal and rPPG image from the continuous face frames.

melanin and residual pigments are much higher than that of hemoglobin in the face image (because hemoglobin is present under the skin while melanin is present on the skin). As a result, (14) can be approximated as

$$-\frac{i'_{x,y}(n)}{i_{x,y}(n)} = v^{m'}_{x,y}(n)\, c^{m}_{x,y} + A^{0'}_{x,y}(n)\,. \tag{15}$$

Hence, (15) represents the absorption changes in melanin and residual pigments over time.

E. PROPOSED EXTRACTION METHOD: HSI(1)+RLS

Our study is based on two important assumptions. First, it is assumed that the irregular intensities induced by the uneven face shapes can trigger motion artifacts. Second, the intensities of hemoglobin, melanin, and residual pigments of the facial skin are time-varying values influenced by the heart activity and continuous movements. To remove the motion artifacts according to these two assumptions, an extraction method that combines the HSI modification (HSI(1)) and the RLS-based ANC was applied [14], [15]. The HSI(1) is the method that can reduce the effects of irregular intensity induced by uneven face shapes because the HSI(1) can change from an uneven intensity to an even intensity. ANC is the adaptive filtering method that removes the motion artifact $v_{x,y}(n)$ by updating the unknown filter coefficient $\hat{\boldsymbol{w}}_{\boldsymbol{x,y}}(n)$. By using (15) as the noise signal (e.g., $\boldsymbol{v_{x,y}}(n) = -i'_{x,y}(n)/i_{x,y}(n)$), the effects induced by the absorption change in melanin and residual pigments can be reduced. Additionally, zero-phase infinite impulse response filter (0.5 – 4 Hz bandpass) is used to enhance the rPPG information and reduce the unnecessary noise [14].

Assuming that the input image and rPPG information $\mathrm{I}_{x,y}(n)$ and $\mathrm{H}_{x,y}(n)$, respectively, the proposed algorithm (HSI(1) + RLS) is described in Algorithm 1.

For evaluating the performance of the proposed method, two different cases are constructed using $\mathrm{H}_{x,y}(n)$. Those are rPPG signal $\psi(n)$ and rPPG image $\boldsymbol{\Psi}(n)$, respectively. The rPPG signal $\psi(n)$ is similar to the PPG signal because $\psi(n)$ is a one-dimensional signal and varies according to a time series. The corresponding value is calculated as follows:

$$\psi(n) = \frac{1}{N}\frac{1}{N}\sum_x \sum_y \mathrm{H}_{x,y}(n). \tag{16}$$

The rPPG image $\boldsymbol{\Psi}(n)$ represents the variations in blood flow changes on the face as a 2D image. To visualize the changes, the original input image $\mathrm{I}_{x,y}(n)$ is overlapped with rPPG information $\mathrm{H}_{x,y}(n)$ multiplied by the amplification parameter γ, with a scale from 0 to 1. Assuming that the input image size is X and Y, the rPPG image $\boldsymbol{\Psi}(n)$ is described as follows:

$$\boldsymbol{\Psi}(n) = \mathrm{I}_{x,y}(n) + \mathrm{H}_{x,y}(n)\,\gamma \begin{cases} x = 1, \ldots, X \\ y = 1, \ldots, Y. \end{cases} \tag{17}$$

Before implementing the proposed algorithm, the original input facial image was downsampled to reduce the computational cost. Finally, the rPPG signal $\psi(n)$ is calculated using $\mathrm{H}_{x,y}(n)$, derived from the downsampled input facial image. Subsequently, the upsampling of $\mathrm{H}_{x,y}(n)$ is also performed to reconstruct the rPPG image $\boldsymbol{\Psi}(n)$. Overall procedure of the proposed method is schematically illustrated in Fig. 5.

Algorithm 1 Proposed Method: HSI(1) + RLS

Input: Input image $\mathrm{I}_{x,y}(n)$; length of time index N; order p
Output: rPPG information $\mathrm{H}_{x,y}(n)$
for $n = 1$: N **do**
Step 1: Calculate HSI(1): $i_{x,y}(n) \leftarrow \mathrm{I}_{x,y}(n)$
Step 2: Calculate the noise signal:

$$v_{x,y}(n) \leftarrow -i'_{x,y}(n) \Big/ i_{x,y}(n)$$

Step 3: ANC based on RLS:

$$\mathbf{g}_{x,y}(n) = \frac{\lambda^{-1}\mathbf{R}^{-1}_{x,y}(n-1)\boldsymbol{i}_{x,y}(n)}{1+\lambda^{-1}\boldsymbol{i}^{\mathrm{T}}_{x,y}(n)\mathbf{R}^{-1}_{x,y}(n-1)\boldsymbol{i}_{x,y}(n)}$$
$$\alpha_{x,y}(n) = d_{x,y}(n) - \hat{\boldsymbol{w}}_{x,y}(n)^T \boldsymbol{i}_{x,y}(n)$$
$$\hat{\boldsymbol{w}}_{x,y}(n+1) = \hat{\boldsymbol{w}}_{x,y}(n) + \alpha_{x,y}(n)\,\mathbf{g}_{x,y}(n)$$
$$\hat{d}_{x,y}(n) = i_{x,y}(n) - \hat{\boldsymbol{w}}_{x,y}(n)^T \boldsymbol{i}_{x,y}(n)$$

Step 4: Implement the IIR filter: $\mathrm{H}_{x,y}(n) \leftarrow \hat{d}_{x,y}(n)$
return $\mathrm{H}_{x,y}(n)$ $n = 1, \ldots, N$

TABLE 1. Summary of previous rPPG information estimation algorithm and proposed algorithm.

Algorithms	Input data type	Input data dimension	Data set, reference	Face detection	Comments
IIR [19]	G	1 × N	Male: 0, Female: 14, PPG	VJ	Butterworth bandpass filtering
REE [21]	R, G	2 × N	Ambiguous (face, palm, arm), PPG	VJ	Skin modeling based on Lambert–Beer law
ICA [13]	R, G, B	3 × N	Male: 10, Female: 2, PPG	VJ	Feature normalization, Fast ICA
EVM [22]	CFF	3 × 200 × 200 × N	-	VJ	Amplifying hidden information based on spatio–temporal filtering
POS [12]	R, G, B	3 × N	60 video sequences, ECG or PPG	VJ	Applying a projection plane orthogonal to the skin-tone
PhysNet [30]	CFF	3 × 128 × 128 × 64 (each epoch)	212 video sequences, ECG or PPG	VJ	Combine 3D-convolutional neural network and temporal encoder–decoder
rPPGNet [9]	CFF	3 ×128 × 128 × 64 (each epoch)	527 video sequences ECG	VJ, SD	Combine a spatio–temporal convolutional network and skin-based attention module
HSI(1) + RLS	CFF	3 × 200 × 200 × N	Male: 13, Female: 2, PPG	VJ	Skin modeling based on Lambert–Beer law, HSI modification, ANC process based on RLS

R, G, B: average pixels from 3 channels (red, green and blue) of each face frame, CFI: continuous face frames, VJ: Viola–Jones face detection algorithm [40], SD: skin detection algorithm [41], N: total video frame length for each subject.

F. VALIDATI ON PROCESS

To evaluate the performance of the proposed algorithm, the following baseline methods were implemented: infinite impulse response (IIR) filtering [19], robust efficient estimation (REE) [21], independent component analysis (ICA) [13], Eulerian video magnification (EVM) [22], Plane-Orthogonal-to-Skin (POS) [12], PhysNet [30] and rPPGNet [9]. Table 1 summarizes the previous studies of extracting rPPG information from continuous face frames and the proposed method. Additionally, Fig. 6 shows the reconstructed rPPG signal compared with the results of previous studies.

To validate the proposed methods, two analyses were performed: heart rate and heart rate variability (HRV). To estimate the accurate heart rate, the spectral analysis using the fast Fourier transform (FFT) algorithm was applied to the rPPG signal. Subsequently, the most dominant frequency value was represented as the heart rate. The number of beats per minute (bpm) was calculated by multiplying the most dominant frequency by 60 (i.e., $1.375 \times 60 = 82.5$ bpm). In this study, heart rate values were calculated from the rPPG signal (180 s). The first 10 seconds of the rPPG signal were excluded due to the initial transient time of the proposed method. Therefore, the 17 epochs per subject were obtained by dividing the rPPG signal into windows of 10 s length. The heart rate signal was calculated by applying FFT algorithm into each epoch. In other words, the total number of heart rate was 510 (15 subjects × 2 experiments ×17 epochs). The true heart rate was obtained from PPG signal. Similarly, the first 10 s of PPG signal were excluded and the 17 epochs were obtained by dividing PPG signal into window of 10 s length. Finally, the estimated heart rates (IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]) are in a one-one correspondence with the true heart rates (PPG).

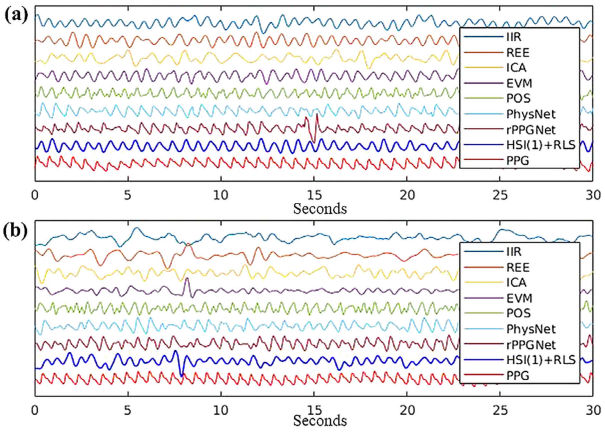

FIGURE 6. rPPG signal induced by various algorithms: IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9] and proposed algorithm (HSI(1) + RLS) in (a) static images and (b) MA-corrupted images. The reference signal is PPG.

HRV is measured by the variations in time between each heartbeat and the distance between heartbeats is called RR interval or NN interval. HRV can be represented using various parameters corresponding to time, frequency, and geometric domains [1]. In this study, five different time-domain parameters were estimated, i.e., Mean RR (average of the beat intervals), STD RR (standard deviation of the beat intervals), SDSD (standard deviation of differences between adjacent bit intervals), RMSSD (the square root of the average of the sum of the squares of differences between adjacent beat intervals), and pNN50 (percentage of the average of the pairs of adjacent beat intervals differing by more than 50 ms in the beat intervals in the entire recording). The HRV parameters were obtained via IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30], rPPGNet [9] and the PPG signal in the static and MA-corrupted images.

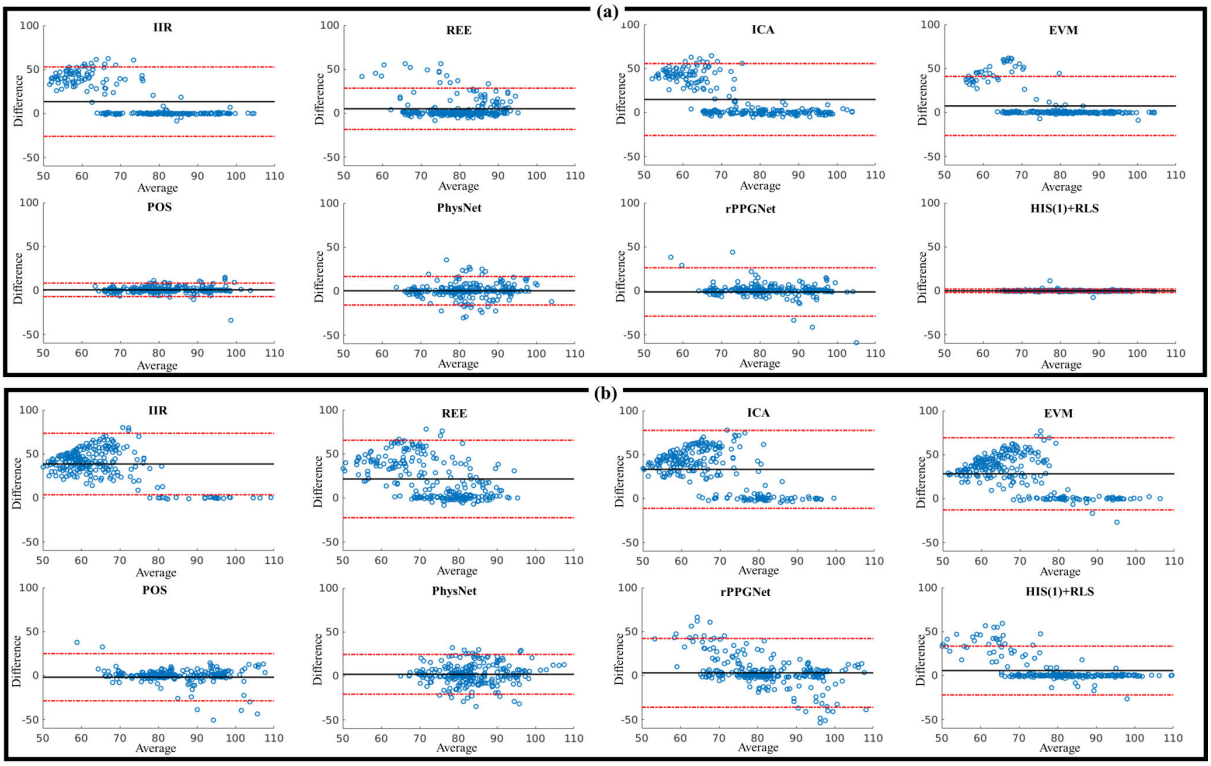

FIGURE 7. Bland–Altman plots for the proposed method (HSI(1) + RLS) and previous algorithms (IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]). The vertical axis represents the heart rate difference (true heart rates – estimated heart rates). The horizontal axis represents the average ([true heart rates + estimated heart rates]/2) (a) Static images. (b) MA-corrupted images.

In addition, the first 10 s of the rPPG signal were excluded due to the initial transient time.

IV. RESULT AND EVALUATION

A. HEART RATE ESTIMATION

To illustrate the performance of the proposed method, Bland–Altman method, which is a graphical approach to evaluate a statistical similarity, was used. In general, a Bland–Altman plot consists of the difference (vertical-axis) and average (horizontal-axis) between the estimated and true heart rates. Therefore, the Bland–Altman plot can describe mean $\tilde{m}$ and standard deviation $\tilde{d}$ for the heart rate difference as well as a 95% confidence interval ($\tilde{m} \pm 1.96\tilde{d}$) to show an agreement between the estimated and true heart rates. Fig. 7 shows the estimated heart rates from previous algorithms (IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]) and the proposed algorithm (HSI(1) + RLS), extracted from the static and MA-corrupted images. In particular, Fig. 7(a) shows that most of the data points stay within a 95% confidence interval. The proposed algorithm has the smallest $\tilde{d}$ for static images. The results of previous algorithms (IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]) show that the bias of $\tilde{m}$ and $\tilde{d}$ are greater than those for the proposed algorithm. Fig. 7(b) shows that $\tilde{d}$ and bias of $\tilde{m}$ in the proposed algorithm are greater than the results of Fig. 7(a) for the MA-corrupted case. However, the changes of bias in the proposed algorithm are less than those in other algorithms. Tables 2 and 3 represent the quantitative values of the Bland–Altman plot. Correlation coefficients between the estimated heart rates were calculated from the rPPG signal obtained by various algorithms and true heart rates calculated from the PPG signal. In addition, when the p-value is less than 0.001, the correlation coefficient value is considered to be significant.

Table 2 shows the Bland–Altman results for static images. Here, $\tilde{m}$ is 0.04 bpm, and the widths of 95% confidence interval ($\tilde{m} \pm 1.96\tilde{d}$) are 3.89 bpm in the proposed algorithm. In addition, the correlation coefficient ($\approx$0.99) and p-value ($<$0.001) indicate a good heart-rate agreement between the PPG signal and the estimated rPPG signals in all subjects. IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9] obtain worse values of $\tilde{m}$ and 95% confidence interval ($\tilde{m} \pm 1.96\tilde{d}$) compared to the proposed algorithm. In addition, although p-value ($<$0.001) indicates highly significant results, the correlation coefficient ($<\pm$0.5) is lower than the results of the proposed algorithm.

Table 3 shows the Bland–Altman results for MA-corrupted images. The result of $\tilde{m}$ and the widths of 95% confidence interval ($\tilde{m} \pm 1.96\tilde{d}$) obtained by the proposed algorithm arehigher than the results for static images. However, the correlation analysis of the proposed algorithm shows that the heart rate calculated by the proposed algorithm is somewhat similar to the true heart rate induced by the PPG signal (although the estimated rPPG signal was corrupted by

TABLE 2. Statistical analysis of various algorithms for static images. * indicates the significance at $p < 0.001$.

Algorithms	$\tilde{m}$ (bpm)	$\tilde{m}$+1.96$\tilde{d}$ (bpm)	$\tilde{m}$–1.96$\tilde{d}$ (bpm)	Correlation coefficient
IIR [19]	13.49	52.85	−25.88	0.4087*
REE [21]	4.99	28.41	−18.44	0.2778*
ICA [13]	7.46	40.97	−26.06	0.3069*
EVM [22]	14.8	55.66	−26.05	0.2884*
POS [12]	−1.18	24.62	−26.98	0.4905*
PhysNet [30]	−0.37	17.02	−17.76	0.3995*
rPPGNet [9]	−3.07	32.81	−38.94	0.2902*
HSI(1) + RLS	0.04	1.98	−1.91	0.9942*

TABLE 3. Statistical analysis of various algorithms for MA-corrupted images. * indicates the significance at $p < 0.001$.

Algorithms	$\tilde{m}$ (bpm)	$\tilde{m}$+1.96$\tilde{d}$ (bpm)	$\tilde{m}$–1.96$\tilde{d}$ (bpm)	Correlation coefficient
IIR [19]	38.71	73.85	−3.58	0.2845
REE [21]	21.55	64.58	−22.48	−0.1241
ICA [13]	28.10	69.18	−12.98	0.1310
EVM [22]	33.36	77.82	−11.11	−0.0230
POS [12]	−6.64	42.38	−55.65	0.1902
PhysNet [30]	1.71	20.76	−17.34	0.3738*
rPPGNet [9]	5.14	48.40	−38.11	0.2112*
HSI(1) + RLS	5.77	33.46	−21.92	0.6272*

motion artifacts). However, the correlation analysis of other algorithms indicates that the estimated heart rate calculated by other algorithms is influenced by motion artifacts.

B. HRV ESTIMATION

Table 4 summarizes the HRV parameters obtained via the various algorithms for the static image. Generally, the difference between the HRV parameters calculated from the PPG signal and the proposed algorithm is less than that calculated from the PPG signal and the algorithms such as IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]. In general, the smaller the difference between the HRV parameter of the PPG signal and the HRV parameter of the estimated signal, the better the performance is demonstrated. In particular, the Mean RR values obtained via the proposed and other algorithms are similar to that obtained via the PPG signal. However, HRV parameters such as STD RR, SDSD, RMSSD, and pNN50 calculated via the other algorithms are greater than those calculated via the proposed algorithm.

Table 5 lists the HRV parameters obtained via various algorithms for the MA-corrupted image, which are larger than those obtained for the static image. In particular, the Mean RR values calculated via the other algorithms for the static image are similar to those of the PPG signal; however, the Mean RR values calculated via the other algorithms (IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]) for the MA-corrupted image exhibit a difference of more than 100 ms than those of the PPG signal. Other HRV parameters such as STD RR, SDSD, RMSSD, and pNN50 are also higher for the MA-corrupted image than for the static image. However, the increase in the HRV parameter values obtained via the proposed algorithm is less than those obtained via the other algorithms for the MA-corrupted image.

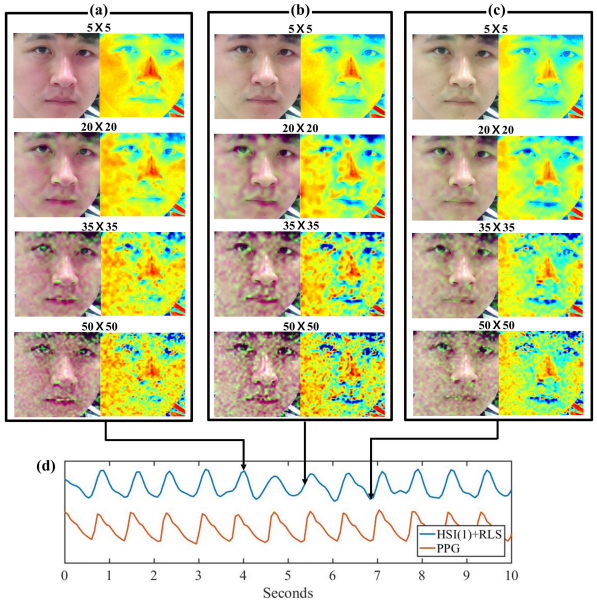

FIGURE 8. Face images reconstructed by the proposed algorithm according to the downsampling sizes of 5 × 5, 20 × 20, 35 × 35, and 50 × 50 pixels. Each left-side face image and right-side face image is the rPPG image and the rPPG information, respectively. (a), (b), and (c) are the reconstructed face images in high, intermediate, and low peak positions of the rPPG signal, respectively. (d) shows the rPPG signal (35 × 35 downsampling size) and the PPG signal.

C. IMAGE RECONSTRUCTION AND REMOTE PPG MEASUREMENT

An ROI size of 200 × 200 pixels was used for applying the proposed algorithm to face images of 640 × 480 pixels. In addition, to reduce the computational complexity, the sizes of downsampled images such as 5 × 5, 20 × 20, 35 × 35, and 50 × 50 pixels were established. Each left-side face image in Fig. 8 shows the rPPG image calculated by the proposed algorithm according to the downsampled sizes of input images. Each right-side face image is the G-plane of the RGB model in the rPPG image. These were calculated from the high, intermediate, and low peak positions of the rPPG signal in Fig. 8 (a), (b), and (c), respectively. The rPPG images in Fig. 8 (a) show high rPPG information at the positions of cheek and nose, compared with that of Fig. 8 (c). Therefore, the blood flow changes mainly occur in the cheeks and nose because capillaries are widely distributed. Furthermore, the downsampling size affects the resolution of rPPG images. For example, the downsampled image of 50 × 50 pixels can show obvious spatial patterns on the rPPG image. Meanwhile, the downsampled image of 5 × 5 pixels has only vague spatial patterns. Fig. 8 (d) shows the rPPG signal calculated from the original face images and the PPG signal measured at the finger. It indicates that the shape of the rPPG signal is similar to the PPG signal, and the proposed algorithm can accurately extract the pure blood flow changes.

TABLE 4. HRV parameters using various algorithms in the static image.

Algorithms	Mean RR (ms)	STD RR (ms)	SDSD (ms)	RMSSD (ms)	pNN50 (ms)
IIR [19]	761.25 ± 90.2	91.62 ± 37.9	103.69 ± 42.1	103.63 ± 42.2	74.91 ± 10.3
REE [21]	764.43 ± 60.5	67.19 ± 19.2	67.68 ± 19.5	67.57 ± 19.5	63.1 ± 8.6
ICA [13]	768.43 ± 80	100.95 ± 40	106.66 ± 41.3	106.46 ± 41.3	74.14 ± 7.4
EVM [22]	749.51 ± 73.2	84.97 ± 28	88.21 ± 30.6	88.05 ± 30.3	73.88 ± 9.6
POS [12]	676.61 ± 80.6	81.94 ± 25.5	82.12 ± 26.9	81.95 ± 26.7	68.51 ± 9.4
PhysNet [30]	624.83 ± 75.3	90.3 ± 20.2	88.12 ± 25.1	88.07 ± 25.1	70.48 ± 9.6
rPPGNet [9]	651.44 ± 51.8	93.08 ± 24.7	97.59 ± 26.2	97.64 ± 26.7	75.76 ± 5.7
HSI(1)+RLS	735.01 ± 77.1	64.93 ± 13.5	66.33 ± 18	66.15 ± 17.9	63.73 ± 8.3
PPG signal	738.85 ± 79.8	47.67 ± 11.1	59.8 ± 15.4	59.64 ± 15.3	56.39 ± 9.6

TABLE 5. HRV parameters using various algorithms in the MA-corrupted image.

Algorithms	Mean RR (ms)	STD RR (ms)	SDSD (ms)	RMSSD (ms)	pNN50 (ms)
IIR [19]	940.58 ± 158.4	193.1 ± 54.2	203.31 ± 61.8	202.84 ± 61.2	86.81 ± 6.0
REE [21]	813.27 ± 55.9	103.42 ± 34	110.61 ± 44.5	110.37 ± 44.3	75.86 ± 8
ICA [13]	875.8 ± 48.3	142.82 ± 48.9	151.84 ± 54.1	151.33 ± 53.6	81.24 ± 7.9
EVM [22]	833.72 ± 105.4	152.31 ± 42.8	159.13 ± 44.9	158.66 ± 44.8	82.48 ± 7
POS [12]	597.52 ± 63.6	90.7 ± 18.8	93.32 ± 23.1	93.3 ± 23.1	74.02 ± 8.5
PhysNet [30]	601.56 ± 39.6	98.88 ± 11.3	95.63 ± 16.9	95.63 ± 16.8	72.92 ± 4.9
rPPGNet [9]	637.6 ± 29.4	102.76 ± 25.1	105.02 ± 27	104.97 ± 26.9	78.05 ± 6.6
HSI(1)+RLS	723.74 ± 74.8	80.52 ± 24.6	81.46 ± 28.7	81.22 ± 28.6	70.05 ± 10.5
PPG signal	718.74 ± 82.3	48.7 ± 9.8	60.05 ± 12.4	59.89 ± 12.3	55.38 ± 7.8

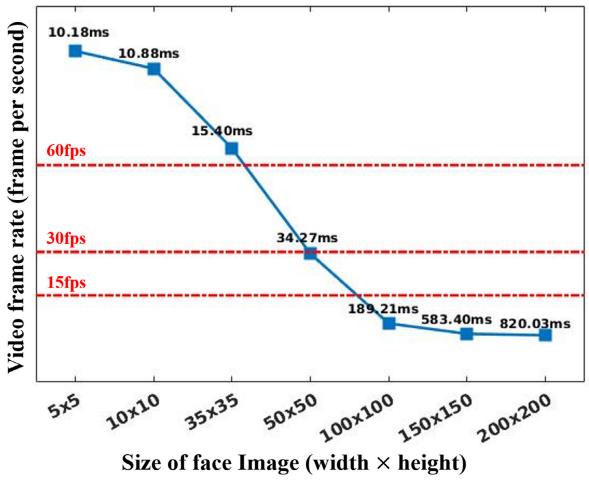

FIGURE 9. Video frame rates according to the downsampled sizes such as 5 × 5, 10 × 10, 35 × 35, 50 × 50, 100 × 100, 150 × 150, and 200 × 200 pixels. The numbers on the blue squares represents the computation time for processing one frame. The red dot lines represent the limit of camera specification for configuring the real-time system.

The proposed algorithm is intended for real-time systems, and the relationship between the computation time and face image sizes is plotted in Fig. 9. This figure shows that the ROI size is inversely proportional to the video frame rate. Therefore, a suitable downsample is essential to configure a real-time system based on the camera specification.

V. DISCUSSION

In this study, it is demonstrated that the proposed algorithm (HSI(1) + RLS) can estimate the correct rPPG signal and the rPPG image from continuous face frames. In general, estimating the rPPG signal from a regular webcam has a critical limitation because the color changes induced by blood flow are much smaller than the color changes of human movements. In this study, it is assumed that the motion artifacts can be triggered by two problems: a) unevenly distributed intensity induced by face curves and b) the color changes caused by the movement of melanin and residual pigments.

The first problem implies that the rPPG signal is influenced by an irregular face surface. Although the light source (e.g., fluorescent light) is uniformly distributed, the light reflected from the human face has irregular intensity, which is measured by a webcam. Therefore, head movements may cause motion artifacts, which are generated by the amplification of unevenly distributed intensity of face curves. Motion artifacts may considerably overlap with the rPPG signal. To solve this problem, the HSI is modified and shown in Fig. 3.

To solve the second problem, the Lambert–Beer law, which can represent the absorbance of melanin, hemoglobin, and residual pigments by (7), was applied. A previous study that estimated rPPG signals from continuous face frames assumed that the melanin and residual pigments were constant, and the change of each pixel was induced only by the change

of hemoglobin [13]. However, as motion artifacts can be triggered by the movements of melanin and residual pigments, it becomes difficult to estimate pure rPPG signals. Therefore, in the ordinary Lambert–Beer law, melanin and residual pigments defined as constants, were replaced with variables which change with various head motions in (8). Accordingly, the ANC method based on the RLS algorithm was used to reduce the effects of variation of melanin and residual pigments, as shown in Fig. 4.

Accordingly, the proposed algorithm is constructed by combining the HSI modification and ANC based on RLS (HSI(1) + RLS) to reduce motion artifacts induced by head movements and irregular reflected intensity, as shown in Fig. 5. To evaluate the performance of the proposed algorithm, the rPPG signal was calculated using the methods of previous studies, such as IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]. In addition, to model different environments, two different experiments were conducted with static and MA-corrupted images. First, the static images show the human face without movements. Therefore, melanin and residual pigments are fixed at each pixel of the face image, and the irregular reflected intensity from the face does not trigger motion artifacts. Second, the MA-corrupted images show the human face with movements and the variety of motion artifacts occurred.

A. EVALUATION OF HEART RATE

To quantitatively evaluate the proposed algorithm, the heart rate by previous studies was estimated and compared with the heart rate of the PPG signal. Table 2 shows that the heart rates estimated by the proposed and previous algorithms are not significantly different from that of the PPG signal in static images. Table 3 shows that the heart rates obtained by the PhysNet [30], rPPGNet [9] and proposed algorithm are not significantly different from the true heart rate of PPG signal in the MA-corrupted images. However, correlation coefficients of PhysNet (0.3738) and rPPGNet (0.2112) are smaller than that of proposed algorithm (0.6272). In other words, in comparison with previous methods, the proposed algorithm can estimate the accurate heart rate robustly in the MA-corrupted environment. Additionally, the value of HSI(1) + RLS listed in Tables 2 and 3 are 0.4 bpm and 5.77 bpm mean($\tilde{m}$), respectively. However, a bias of 5.77 bpm was observed. The proposed algorithm considers two different assumptions. First, melanin and residual pigments are time-varying with respect to the recording camera. To reduce the motion artifacts triggered by melanin and residual pigments, the RLS algorithm is applied. In RLS, the desired signal entails the sum of the variance of melanin and residual pigments, as given in (15). Therefore, whenever motion artifacts occur, the desired signal increases. Since the value of variance of hemoglobin is considerably smaller than that of the desired signal in MA-corrupted environments, RLS can remove not only the noise signal derived by motion artifacts, but also a part of the rPPG signal. Therefore, a bias is observed in the MA-corrupted environment.

B. EVALUATION OF HEART RATE VARIABILITY

Based on the HRV parameter values listed in Tables 4 and 5, compared with the previous methods, the proposed algorithm can effectively estimate the rPPG signal. Since the HRV parameters are calculated with respect to the beat interval of heart activity, the results are very sensitive, particularly to the motion artifacts. Therefore, the HRV parameters calculated via the proposed and other algorithms are similar to those calculated for the PPG signal in the static image as only a small motion artifact effect is observed. However, in the MA-corrupted image, since correct beat intervals are difficult to obtain, the HRV parameter exhibits a larger bias than those obtained for the static image. Therefore, the Mean RR of the HRV parameters calculated via the other algorithms exhibit a difference of more than 100 ms and results of other HRV parameters such as STD RR, SDSD, RMSSD and pNN50 are more than twice that of reference, as summarized in Table 5. The table also indicates that the proposed algorithm effectively removes motion artifacts and estimates HRV parameters more precisely.

C. EVALUATION OF REMOTE PPG INFORMATION

The shapes of rPPG signals calculated by previous studies and proposed algorithm are similar to that of the PPG signal in Fig. 7 (a) because changes in the position of melanin and residual pigments barely occur. In addition, the reflected intensity change of the face image is also small. Meanwhile, the shapes of rPPG signals calculated by previous studies are different from that of the PPG signal because the changes in melanin and residual pigments as well as irregular intensity caused by face curves can trigger various motion artifacts in Fig. 7 (b). In particular, the rPPG signal of IIR has a low frequency component. Because the IIR filter is a linear time-invariant system, it is difficult to eliminate nonlinear signals, such as motion artifacts [14], [15]. The motion artifacts also affect the results of REE because the REE does not assume the changes in melanin and residual pigments in (7). Therefore, noise signals triggered by motion artifacts were unstably included in the results [22]. ICA assumes that the rPPG signal is an IC. In short, the interactions among ICs are minimal, and the characteristics of ICs themselves are maximal. However, the results of ICA show that the signals triggered by melanin and residual pigments and the signals estimated from the movement of hemoglobin are dependent, and the method eventually fails to separate them, as shown in Fig. 7 and Table 3, respectively [13]. The rPPG signal calculated by EVM is also vulnerable to motion artifacts. The concept of EVM is not to extract rPPG signals but to amplify slight movements. Therefore, it is impossible to separate the noise signals (melanin and residual pigments) from the rPPG signals. In other words, EVM barely considers motion artifacts compared with other previously developed algorithms [22]. POS can remove the noise signal by assuming that the noise signal is orthogonal only to skin-tone and by projecting averaged signals of face image onto the plane [12]. However, POS does not consider the unevenly distributed face curve,

and unevenly distributed face curve eventually caused motion artifacts in the MA-corrupted image. PhysNet and rPPGNet are based on the convolutional neural network (CNN) and have a form of supervisor learning. Therefore, learning model requires a high-purity raining sets of face image [9], [30]. Although, results of PhysNet and rPPGNet are better than the previous studies, performances of PhysNet and rPPGNet are lower than that of proposed algorithm in the static and MA-corrupted image. Therefore, we see that the two algorithms (PhysNet and rPPGNet) does not consider the optical properties of facial skin, resulting in a poor result.

Finally, the proposed algorithm also reconstructed the rPPG image in real-time. However, the proposed algorithm may have a high computational cost because the proposed algorithm is applied to each pixel of face images in (17). Consequently, it is necessary to reduce the size of images for extracting rPPG signals and optimizing a real-time system. Fig. 8 shows the results of rPPG image reconstruction depending on the downsampling size. Especially, a rPPG image of 50 × 50 pixels has a high resolution. However, it is difficult to show the changes in clear rPPG patterns because the proposed algorithm is applied to each pixel, and the filtered signals in each pixel are independent. Meanwhile, the rPPG images of 5 × 5 and 20 × 20 pixels have a suitable sparse resolution. The downsampling process for reducing the computational cost can prevent displaying the excessive sparse information produced by independent filtering processes in each pixel and provide a blur effect. In addition, Fig. 9 shows that downsampling size selection depends on the camera specification used for configuring real-time system. To implement the real-time system, it is necessary to complete rPPG signal estimation process before acquiring the next face image. For instance, a camera at 60 fps should be configured with a downsampling process of at least 35 × 35 pixels, whereas, a camera at 15 fps is capable of a downsampling process up to 50 × 50 pixels. Therefore, the proper downsampling process not only produces a more efficient real-time system but also displays distinct rPPG pattern images.

D. LIMITATIONS

The proposed method, complying with the Lambert–Beer law, can be applied to human skin modeling to remove the artifacts derived by the reflective irregular intensity as well as various head movements. The results for heart rate estimation and HRV obtained in this study are better than those obtained in previous studies that employed IIR [19], REE [21], ICA [13], EVM [22], POS [12], PhysNet [30] and rPPGNet [9]. However, the proposed algorithm exhibits two limitations. First, the rPPG signal is influenced by the intensity of the light source and variation in the location of the light source. The focus is on the reflective irregular intensity caused by face shape, not the irregular intensity caused by the variation in the light source. Therefore, extracting the rPPG signal in an outdoor environment is difficult since the light source intensity and location can vary unexpectedly. Second, in MA-corrupted environments, heart rate value calculated with the rPPG signal is biased compared with that calculated with the reference PPG signal. In RLS, the desired signal is the sum of melanin and pigment variation and is larger in MA-corrupted environments. The large value of desired signal can be removed from original signal and the result of proposed algorithm is unstable. Therefore, an adaptive attenuation value should be studied to reduce the effect of the desired signal in MA-corrupted environments.

VI. CONCLUSION AND FUTURE WORK

In this study, we proposed a novel method that combined the HSI modification with RLS-based ANC to estimate the rPPG signal and rPPG image. To evaluate the performance of the proposed algorithm, the estimation performances of heart rate and HRV were investigated, and compared the results with previous algorithms. Consequently, the proposed algorithm shows a better heart rate agreement between the estimated rPPG signal and the PPG. However, the cross-correlation in MA-corrupted images remained low. It means that the motion artifacts were not removed perfectly. In addition, it is necessary to evaluate the performance of the proposed algorithm according to different skin colors as well because the distribution of melanin under the face is different according to the race, and can affect the extraction of an accurate rPPG signal. Finally, there is a plan to recruit the participants of various races for analyzing the noise cancellation performance.

REFERENCES

[1] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, and F. Lombardi, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.

[2] G. Lu, F. Yang, J. Taylor, and J. Stein, "A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects," *J. Med. Eng. Technol.*, vol. 33, no. 8, pp. 634–641, 2009.

[3] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," *J. Med. Eng. Technol.*, vol. 32, no. 6, pp. 479–484, 2008.

[4] M. Bolanos, H. Nazeran, and E. Haltiwanger, "Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 6, Aug. 2006, pp. 4289–4294.

[5] I. Nishidate, N. Tanaka, T. Kawase, T. Maeda, T. Yuasa, Y. Aizu, T. Yuasa, and K. Niizeki, "Noninvasive imaging of human skin hemodynamics using a digital red-green-blue camera," *J. Biomed. Opt.*, vol. 16, no. 8, 2011, Art. no. 086012.

[6] Q. He and R. Wang, "Hyperspectral imaging enabled by an unmodified smartphone for analyzing skin morphological features and monitoring hemodynamics," *Biomed. Opt. Exp.*, vol. 11, no. 2, pp. 895–910, 2020.

[7] D. Cho and B. Lee, "Non-contact robust heart rate estimation using HSV color model and matrix-based IIR filter in the face video imaging," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, Art. no. 16379859.

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

[9] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 151–160.

[10] T. Luguev, D. Seus, and J.-U. Garbas, "Deep learning based affective sensing with remote photoplethysmography," in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2020, Art. no. 19668761.

[11] Q. Zhan, W. Wang, and G. de Haan, "Analysis of CNN-based remote-PPG to understand limitations and sensitivities," *Biomed. Opt. Exp.*, vol. 11, no. 3, pp. 1268–1283, 2020.
[12] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2016.
[13] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Exp.*, vol. 18, no. 10, pp. 10762–10774, 2010.
[14] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Hoboken, NJ, USA:Wiley, 1996.
[15] S. Ardalan, S. Moghadami, and S. Jaafari, "Motion noise cancelation in heartbeat sensing using accelerometer and adaptive filter," *IEEE Embedded Syst. Lett.*, vol. 7, no. 4, pp. 101–104, Dec. 2015.
[16] G. Lu, J. S. Brittain, P. Holland, J. Yianni, A. L. Green, J. F. Stein, T. Z. Aziz, and S. Wang, "Removing ECG noise from surface EMG signals using adaptive filtering," *Neurosci. Lett.*, vol. 462, no. 1, pp. 14–19, Oct. 2009.
[17] O. Ikeda, "Segmentation of faces in video footage using HSV color for face detection and image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 3, 2003, pp. 913–916.
[18] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, Dec. 2008.
[19] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, 2007.
[20] A. Roggan, M. Friebel, K. Dorschel, A. Hahn, and G. Müller, "Optical properties of circulating human blood in the wavelength range 400-2500 nm," *J. Biomed. Opt.*, vol. 4, no. 1, pp. 36–46, 1999.
[21] S. Xu, L. Sun, and G. K. Rohde, "Robust efficient estimation of heart rate pulse from video," *Biomed. Opt. Exp.*, vol. 5, no. 4, pp. 1124–1135, 2014.
[22] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, Aug. 2012.
[23] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Feb. 2018.
[24] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.
[25] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018.
[26] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, Aug. 2018.
[27] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805.
[28] X. Han, *Research on Face Recognition Based on Deep Learning*. Beijing, China: Shenyang Normal Univ., 2018.
[29] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *Proc. Int. Conf. Modern Electr. Energy Syst. (MEES)*, Nov. 2017, pp. 376–379.
[30] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *Proc. BMVC*, 2019, pp. 1–12.
[31] E. Lee, E. Chen, and C.-Y. Lee, "Meta-rPPG: Remote heart rate estimation using a transductive meta-learner," in *Proc. Comput. Vis.-Workshops (ECCV)*, 2020, pp. 392–409.
[32] W. Chen and D. Mcduff, "DeepPhys: Video-based physiological measurement using convolutional attention networks," in *Proc. Comput. Vis.-Workshops (ECCV)*, 2018, pp. 349–365.
[33] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Exp.*, vol. 6, no. 5, pp. 1565–1588, May 2015.
[34] H. Fukushima, H. Kawanaka, M. S. Bhuiyan, and K. Oguri, "Estimating heart rate using wrist-type photoplethysmography and acceleration sensor while running," in *Proc. IEEE-EMBC*, Aug./Sep. 2012, pp. 2901–2904.
[35] G. B. Papini, P. Fonseca, M. M. van Gilst, J. W. M. Bergmans, R. Vullings, and S. Overeem, "Wearable monitoring of sleep-disordered breathing: Estimation of the apnea–hypopnea index using wrist-worn reflective photoplethysmography," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, Dec. 2020.
[36] R.-Y. Huang and L.-R. Dung, "A motion-robust contactless photoplethysmography using chrominance and adaptive filtering," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Oct. 2015, pp. 1–4.
[37] K. Lee, J. Lee, C. Ha, M. Han, and H. Ko, "Video-based contactless heart-rate detection and counting via joint blind source separation with adaptive noise canceller," *Appl. Sci.*, vol. 9, no. 20, p. 4349, Oct. 2019.
[38] P.-W. Huang, B.-J. Wu, and B.-F. Wu, "A heart rate monitoring framework for real-world drivers using remote photoplethysmography," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1397–1408, May 2021.
[39] D.-S. Chen and Z.-K. Liu, "Generalized Haar-like features for fast face detection," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Aug. 2007, pp. 2131–2135.
[40] W. G. Zijlstra and A. Buursma, "Spectrophotometry of hemoglobin: Absorption spectra of bovine oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin," *Comparative Biochemistry Physiol. B, Biochemistry Mol. Biol.*, vol. 118, no. 4, pp. 743–749, Dec. 1997.
[41] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human skin detection using RGB, HSV and YCbCr color models," in *Proc. Int. Conf. Commun. Signal Process. (ICCASP)*, vol. 137, 2017, pp. 324–332.

DONGRAE CHO (Member, IEEE) received the B.S. degree in biomedical engineering from Inje University, Gimhae, South Korea, and the M.S. degree from the School of Mechatronics, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. In 2017, he started a business at Deepmedi Inc., as a Co-Founder. He has been developing algorithms to measure bio-signals using camera embedded in smartphone. He is also interested in non-contact measurement technology. He is currently working with Deepmedi Inc., as a Research Institute Director. His current research interests include machine learning, biomedical signal processing, and non-contact healthcare systems.

JONGIN KIM (Member, IEEE) was born in Seoul, Republic of Korea, in 1987. He received the B.S. degree in biomedical engineering from Yonsei University, in 2012, and the M.S. and Ph.D. degrees in biomedical engineering from Gwangju Institute of Science and Technology (GIST), in 2014 and 2019, respectively. Since 2017, he has been working with Deepmedi Inc., as a Co-Founder and a CTO. His research interests include biomedical signal processing, non-contact healthcare technology, and artificial neural networks.

KWANG JIN LEE (Member, IEEE) was born in Seoul, Republic of Korea, in 1985. He received the B.S. degree in mechanical engineering from Kyung Hee University, in 2010, and the M.S. degree in bio-medical engineering from Gwangju Institute of Science and Technology (GIST), in 2014. Since 2017, he has been running the company at Deepmedi Inc., Republic of Korea, where he is currently a CEO. His research interests include bio-signal processing and diagnosis systems using artificial intelligence.

SAYUP KIM (Member, IEEE) received the M.S. and Ph.D. degrees in biomedical engineering from Yonsei University, South Korea. He is currently a Principal Researcher with the Digital Transformation Research and Development Department, Korea Institute of Industrial Technology. He is interested in developing a healthcare service system by measuring bio-signals using non-contact measurement technology. His research interests include sleep quality evaluation, classification of sleep stage, and smart bedding systems for healthcare applications.