

Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin

Litong Feng, *Student Member, IEEE*, Lai-Man Po, *Senior Member, IEEE*,
Xuyuan Xu, *Student Member, IEEE*, Yuming Li, and Ruiyi Ma

Abstract—Remote imaging photoplethysmography (RIPPG) can achieve contactless monitoring of human vital signs. However, the robustness to a subject's motion is a challenging problem for RIPPG, especially in facial video-based RIPPG. The RIPPG signal originates from the radiant intensity variation of human skin with pulses of blood and motions can modulate the radiant intensity of the skin. Based on the optical properties of human skin, we build an optical RIPPG signal model in which the origins of the RIPPG signal and motion artifacts can be clearly described. The region of interest (ROI) of the skin is regarded as a Lambertian radiator and the effect of ROI tracking is analyzed from the perspective of radiometry. By considering a digital color camera as a simple spectrometer, we propose an adaptive color difference operation between the green and red channels to reduce motion artifacts. Based on the spectral characteristics of photoplethysmography signals, we propose an adaptive bandpass filter to remove residual motion artifacts of RIPPG. We also combine ROI selection on the subject's cheeks with speeded-up robust features points tracking to improve the RIPPG signal quality. Experimental results show that the proposed RIPPG can obtain greatly improved performance in accessing heart rates in moving subjects, compared with the state-of-the-art facial video-based RIPPG methods.

Index Terms—Blood volume pulse (BVP), motion artifact, photoplethysmography (PPG), radiometry, remote imaging, skin optics.

I. INTRODUCTION

PHOTOPILETHYSMOGRAPHY (PPG) is an electro-optic technique for noninvasively measuring the tissue blood volume pulses (BVPs) in the microvascular tissue bed underneath the skin [1]. Vital physical signs, such as the heart rate (HR), respiratory rate, and arterial oxygen saturation, can be accessed by PPG. With the rapid advancement of portable imaging devices, especially for smartphones and laptops, there is a trend for transforming conventional contact PPG (CPPG) to remote imaging PPG (RIPPG) [2], [3]. The emerging RIPPG technique has good potential as an innovative way

Manuscript received February 27, 2014; revised June 15, 2014 and August 13, 2014; accepted October 14, 2014. Date of publication October 22, 2014; date of current version May 1, 2015. This work was supported by City University of Hong Kong, Hong Kong, through the CityU Strategic Research Project under Grant 7004058. This paper was recommended by Associate Editor A. Signoroni.

The authors are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: litonfeng2-c@my.cityu.edu.hk; eelmpo@cityu.edu.hk; xuyuanxu2-c@my.cityu.edu.hk; yumingli4-c@my.cityu.edu.hk; ruiyi.ma@my.cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2014.2364415

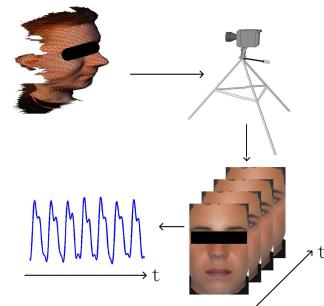


Fig. 1. Operating principle of RIPPG.

to access cardiac pulsations, because only a low-cost digital camera is needed and contact probes or dedicated light sources are not required [4], [5]. As a noncontact technique, RIPPG can be applied in a flexible way, such as RIPPG applications in smartphone applications and a functional mattress [6], [7].

The operating principle of facial video-based RIPPG is shown in Fig. 1. Ambient light is used as the light source. A digital camera is focused on a region of interest (ROI) on the human face. A sequence of facial images is recorded in a video format. Since hemoglobin in blood can absorb light, BVPs beneath the skin surface modulate light absorption by the skin during cardiac cycles, appearing as slight color variations in the skin. Even though the slight color variations of the skin of the face due to BVPs are invisible to human eyes, they can be detected by digital cameras. From the sequence of frames of the facial video, the PPG signal can be extracted by video/signal processing techniques.

Current RIPPG can provide relatively accurate physiological assessment for still subjects. However, it is difficult for current RIPPG to accurately measure vital signs in moving subjects owing to motion artifacts. Typically, environmental light is not distributed with a spatially uniform radiant intensity. Hence, spatial motion will change the radiant flux on the ROI. In addition, motion can also cause ROI fluctuation or even ROI loss in RIPPG. All of these influences will cause motion artifacts in the RIPPG signals. Thus, the measurement accuracy of RIPPG is susceptible to the subject's movements, especially when ambient light and low-cost digital cameras are used [9]. In practice, it is difficult to keep subjects absolutely still. Therefore, a high tolerance for motion artifacts is necessary for RIPPG. Recently, some research works with a focus on reducing motion artifacts for RIPPG have been

reported [5], [9]–[12]. In general, these methods for motion artifacts reduction can be classified into two categories: ROI tracking and the color difference method.

The averaging of image pixel intensities in the ROI produces the raw RIPPG signal. ROI tracking aims to locate and adjust the ROI with the subject's motions, so that the pixels in the ROI used for calculating RIPPG signals belong to a skin region invariant to the subject's motions. ROI tracking is an indispensable first step for obtaining a reliable RIPPG signal for moving subjects. The representative methods of ROI tracking include face tracking based on the Viola–Jones (VJ) algorithm [5] and face tracking based on the Kanade–Lucas–Tomasi (KLT) algorithm [11]. The VJ face detector implements multiscale face detection by training a boosted cascade of classifiers [13]. However, the accuracy of the face location is not well considered, so the ROI fluctuates significantly during tracking. Most of the time, only frontal face images are utilized for training classifiers in the VJ face detector, so ROI tracking will fail when the face position deviates from the frontal position. Meanwhile, the KLT algorithm can provide better face tracking than the VJ algorithm, because the KLT algorithm implements face tracking by tracking feature points [14], [15]. The motion vector representing face motion can be estimated by the KLT algorithm, so the ROIs location, shape, and size can be well-adjusted with the face's motion.

In general, a digital camera sensor has red, green, and blue (RGB) color channels. The color difference method aims to extract the RIPPG signal through a linear combination of raw RIPPG signals from the RGB channels. Because RIPPG signals in different color channels undergo a similar motion modulation caused by a subject's motions, the weighted subtraction of one RIPPG signal from other RIPPG signals in different color channels can effectively reduce motion artifacts. Two color difference methods are independent component analysis (ICA) [5] and the skin tone-based method [12]. The theoretical foundation of ICA is non-Gaussianity is independence, based on the central limit theorem. By maximizing the non-Gaussianity of ICA output components, the linear combination coefficients can be estimated and motion artifacts can be separated from the RIPPG signal. Meanwhile, the skin tone-based color difference method utilizes a standardized skin tone to reduce the effects of specular reflection and nonwhite illumination. The linear combination coefficients are fixed first, and then fine-tuned according to the standard deviations of two color difference signals.

All of the methods mentioned above are derived from digital signal/image processing techniques. The RIPPG signal originates from the radiant intensity variation of human skin due to BVPs, which is an optical signal. The digital camera sensor acts as a photoelectric converter. In this paper, we study the origins of motion artifacts of RIPPG based on the optical properties of human skin. An optical RIPPG signal model is built so that we can solve the problem of motion artifacts more directly and effectively compared with previous works. The ROI of skin can be treated as an extended Lambertian source, which obeys Lambert's law of intensity. Hence, the effect of ROI tracking on reducing motion artifacts can be

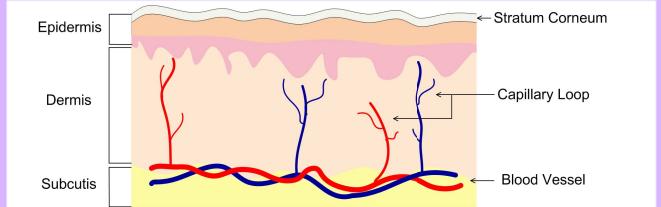


Fig. 2. Multilayered skin model.

analyzed from the perspective of radiometry. In addition, we found that, in theory, ROI tracking cannot remove all motion artifacts. If the digital color camera is regarded as a simple spectrometer, then the combined spectrum of skin tone and ambient illumination can be determined by checking the amplitude of each raw RIPPG signal in the RGB channels. Hence, nonwhite illumination and skin tone variation can be compensated. Based on the wavelength dependency of reflection PPG and skin optics, an adaptive color difference method between the green and red channels is proposed. In addition, an adaptive bandpass filter (ABF) is constructed according to the spectral characteristics of the PPG signal. Based on these findings, a motion-resistant RIPPG method is proposed in this paper. Experiments on HR estimation for moving subjects demonstrated the effectiveness of the proposed RIPPG method in comparison with the state-of-the-art methods.

The rest of this paper is organized as follows. In Section II, the motion artifacts of RIPPG are analyzed from the perspective of radiometry, and an optical RIPPG signal model is proposed. Based on the analysis results in Section II, we propose a new RIPPG method in Section III. To verify our proposals, experimental assessment details are illustrated in Section IV. Experimental results are shown in Section V. Discussion is given in Section VI. Finally, the conclusion is drawn in Section VII.

II. MOTION ARTIFACTS ANALYSIS

The raw RIPPG signal is obtained by averaging all the image pixel values within the ROI in one color channel

$$\text{RIPPG}_{\text{raw}}(t) = \frac{\sum_{x,y \in \text{ROI}} P(x, y, t)}{|\text{ROI}|} \quad (1)$$

where $|\text{ROI}|$ is the size of the ROI and $P(x, y, t)$ is the pixel value at location (x, y) , at time t . The RIPPG signal originates from the light absorption variation of the skin with BVPs. A model of light propagation in skin is used here for describing RIPPG signals in detail. The multilayered skin model has been used to estimate realistic reflectance spectra of skin [16], as shown in Fig. 2.

For RIPPG, we consider only the regular reflection by the stratum corneum, scattering and absorption by the epidermis, and scattering and absorption by the dermis. There is a change in refractive index between the air and the stratum corneum, and thus a small fraction of incident light (4%–7%) will be reflected by the stratum corneum. This reflection is diffuse, because the skin surface is not optically smooth and skin reflection does not maintain an image [17]. Various skin

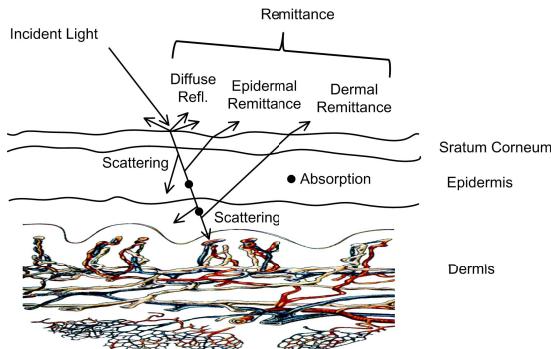


Fig. 3. Light propagation in a simplified skin model.

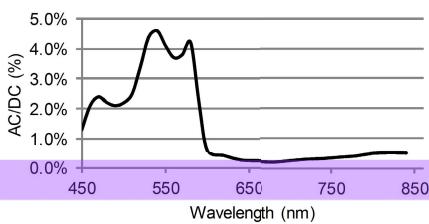


Fig. 4. Wavelength dependence of the ac/dc ratio of a reflection CPPG signal.

chromophores in the epidermis and the dermis absorb incident light. In the visible spectrum, the major chromophores of human skin are melanin and hemoglobin. Melanin is located in the epidermis. In contrast, hemoglobin is located in the microvascular network in the dermis. Scattering also occurs in both the epidermis and the dermis owing to refractive index fluctuations on a microscopic level [18]. In this paper, we simplify the multilayered skin model to only three layers. Light propagation in human skin is shown in Fig. 3.

The remittance from skin includes diffuse reflection by the stratum corneum, and scattered light by the epidermis and the dermis. Since the scattered light can also be regarded as diffuse light (both obey Lambert's law of intensity), we can treat the ROI of the skin as an extended Lambertian source. Absorption by the skin also inversely determines the remittance. The amount of hemoglobin in the dermis will change quasi-periodically with BVPs, and this will increase/decrease the light absorbed by the skin. The RIPPNG signal is generated by recording this radiant intensity variation.

The amplitude of light absorption variation with BVP is very small compared with that of the average remittance, appearing as a small ac component added to a large dc component. For the reflection mode CPPG, the ac/dc ratio of the PPG signal is wavelength dependent, as shown in Fig. 4 [19]. For the RIPPNG signal, the peak ac/dc ratio is even smaller, only about 2%. Based on the above analysis, the optical signal of the RIPPNG can be represented as

$$I_i(t) = \alpha_i \beta_i (S_0 + \gamma_i S_0 \text{Pulse}(t) + R_0) M(t) \quad (2)$$

where $\text{Pulse}(t)$ is the normalized ideal RIPPNG signal, S_0 is the average scattered light intensity from the ROI of the skin with white light illumination, R_0 is the diffuse reflection light intensity from the surface of the ROI of the skin with white

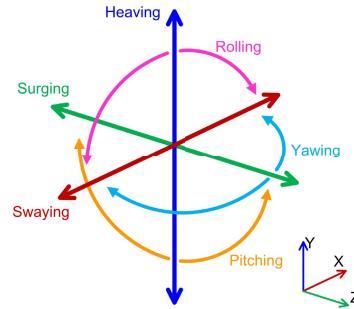


Fig. 5. 6 DOF.

light illumination, the subscript i denotes one RGB, α_i is the power of the i th color light in the normalized practical illumination spectrum, β_i is the power of the i th color light in the normalized diffuse reflection spectrum of the skin, and γ_i is the ac/dc ratio of a RIPPNG signal in the i th color channel. The spectral characteristics of the environmental illumination and the skin remittance are both considered in this optical signal model of RIPPNG. In (2), we assume that the scattering spectrum of the skin and the diffuse reflection spectrum of the stratum corneum have the same spectral characteristics. Thus, the same β_i acts on both S_0 and R_0 . Since capillaries are located in the dermis, not in the epidermis, γ_i can affect S_0 , but not R_0 . This assumption may not be accurate enough for simulating light propagation in the skin. However, the aim of this paper is to reduce motion artifacts for RIPPNG, so this simplified optical signal model for RIPPNG is reasonable. Equation (2) can describe RIPPNG operation in an ideal case. When the subject is moving, the motion will modulate all three optical RIPPNG signals in the RGB channels in the same way, as shown in

$$I_i(t) = \alpha_i \beta_i (S_0 + \gamma_i S_0 \text{Pulse}(t) + R_0) M(t) \quad (3)$$

where $M(t)$ is the motion modulation on the ideal RIPPNG signal.

A. Effect of ROI Tracking

ROI tracking is used to locate and adjust the ROI along with the subject's motions. Thus, the pixels in (1) belong to an invariant skin region, even though the subject moves. Then, motion artifacts due to ROI fluctuation, ROI drift, and ROI loss can be significantly reduced. Thus far, there is no detailed description of ROI tracking for RIPPNG. Since the ROI of the skin can be regarded as an extended Lambertian source, whose radiance obeys Lambert's law of intensity [20], the effectiveness of ROI tracking can be analyzed from the perspective of radiometry. In this analysis, we assume that there is an ideal ROI tracking with no errors and that the environmental illumination is uniform (such as in an integrating sphere). If we treat the face as a rigid body, then the subject's motions can be classified into six different types in 3-D space: swaying, heaving, surging, yawing, pitching, and rolling, as shown in Fig. 5.

The focal length of the digital camera is f , the size of its entrance pupil is A , the distance from the camera to the face

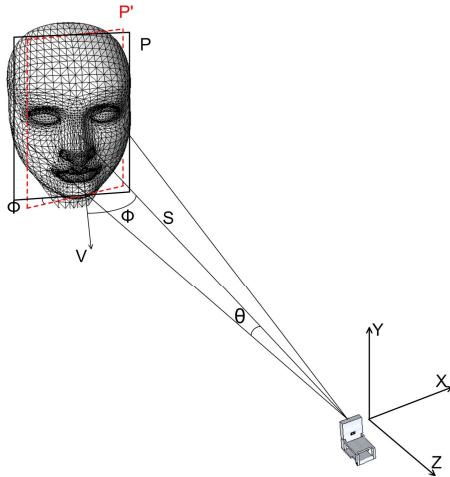


Fig. 6. Radiance analysis of RIPPNG for a yawing motion.

is S , the face width is $2W$, and the camera subtends an angle of 2θ to the width of the face on the frontal plane P' (dotted rectangle), as shown in Fig. 6.

The ROI is assumed to be a square with an area of $4W^2$. Based on the imaging principle of the digital camera, the size of the image ROI on the imaging plane is

$$S_{\text{image_ROI}} = \frac{4W^2 f^2}{S^2}. \quad (4)$$

With the assumption of a Lambertian source of finite size for the face, when the face is located on the optical axis of the digital camera with a frontal position, the radiant intensity H received by the camera from the radiation of the face is

$$H = \pi N \sin^2 \theta \quad (5)$$

where N is the radiance of the face with units of $\text{Watt} \cdot \text{Ster}^{-1} \cdot \text{cm}^{-2}$, and $\sin \theta = \sqrt{2W/2S}$. Since the entrance pupil size is A , the total power received by the camera is

$$P_w = AH = \frac{A\pi NW^2}{2S^2}. \quad (6)$$

The raw RIPPNG signal is proportional to the irradiance on the image ROI on the imaging plane, which is

$$I_{\text{image_ROI}} = \frac{P_w}{S_{\text{image_ROI}}} = \frac{A\pi N}{8f^2}. \quad (7)$$

When the face yaws with an angle Φ , as shown in Fig. 6, the face yaws from the frontal plane P' (red dashed rectangle) to the current plane P (solid rectangle). The angle from the normal of face plane V to the z -axis is also Φ . Even with a yaw angle Φ , it can be proved that the irradiance on the image ROI is still equal to (7). Thus, ideal ROI tracking can cancel the motion artifacts caused by a yawing motion under uniform illumination. Under the same model, it can be proved that the irradiance on the image ROI is also invariant with pitching, rolling, and surging motions. Therefore, ideal ROI tracking can also cancel the motion artifacts caused by these motions under uniform illumination.

However, we found that the results for swaying and heaving motions are not the same, as shown in (7). A swaying motion

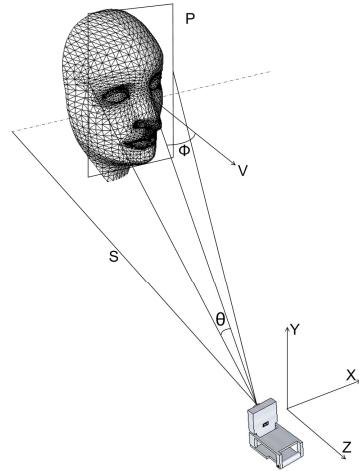


Fig. 7. Radiance analysis of RIPPNG for a swaying motion.

occurs in the form of a left/right translation, as shown in Fig. 7. The face is still on the frontal plane. The line from the camera to the face center makes an angle Φ to the face normal V owing to the translation. The irradiance on the image ROI sourced from the radiation of the face becomes

$$I_{\text{image_ROI}} = \frac{A\pi N \cos^4 \Phi}{8f^2}. \quad (8)$$

In addition, a heaving motion also obtains the same result as shown in (8). The detailed proofs are given in the Appendix. Since there is a factor of $\cos^4 \Phi$ in (8), ROI tracking cannot compensate for motion artifacts caused by swaying and heaving. It can be concluded that ideal ROI tracking can compensate for motion artifacts caused by yawing, pitching, rolling, and surging motions under uniform illumination, but does not work for swaying and heaving motions.

The practical situation is much more complicated than the above analysis. First, it is difficult to obtain a uniform illumination for RIPPNG in everyday environments. Typically, the illumination intensity is spatially inhomogeneous. Second, the actual motion of a subject simultaneously comprises several kinds of motions in the 6 DOF. It is difficult to estimate the irradiance variance of the image ROI due to the motion even under uniform illumination, because it is difficult to determine Φ for a combination of motions. It is an ill-posed problem of estimating 3-D motion based on 2-D images. In conclusion, good-quality ROI tracking plays an important role in reducing motion artifacts. However, it cannot completely eliminate motion artifacts.

B. Effect of the Color Difference Method

For convenience, we rewrite the optical RIPPNG signal model of (3) as

$$I_i(t) = \alpha_i \beta_i (S_0 + \gamma_i S_0 \text{Pulse}(t) + R_0) M(t). \quad (9)$$

There is an identical motion modulation in all three color channels. The aim of the color difference method is to reduce motion artifacts by the weighted subtraction of one RIPPNG

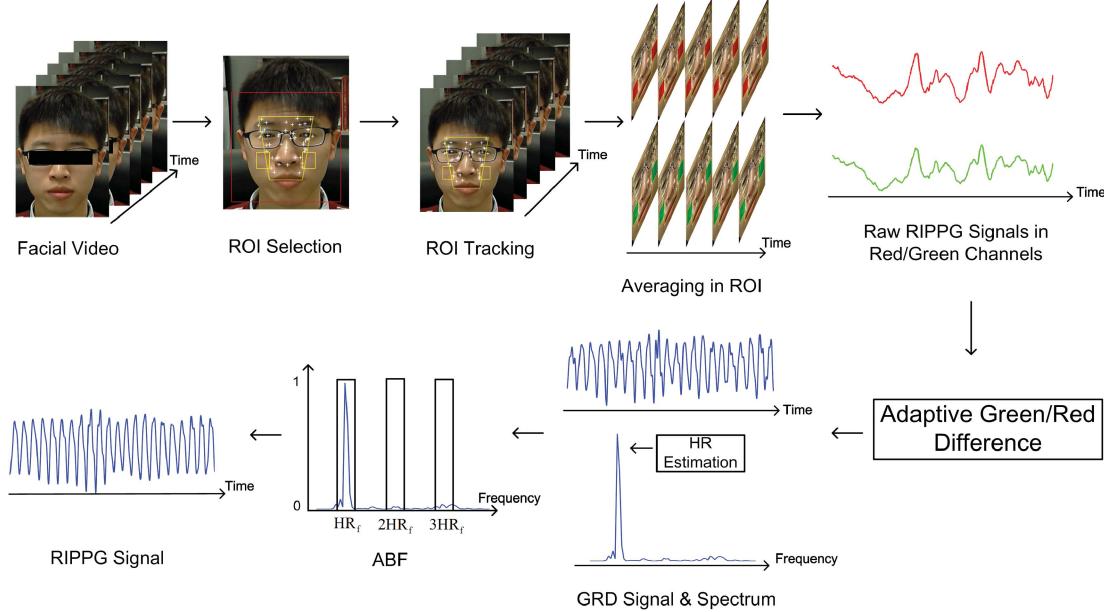


Fig. 8. Flowchart of the proposed RIPPNG method.

signal from another RIPPNG signal in different color channels. The color difference signal can be described as

$$D(t) = \frac{I_i(t)}{\alpha_i \beta_i} - \frac{I_j(t)}{\alpha_j \beta_j} = (\gamma_i - \gamma_j) S_0 \text{Pulse}(t) M(t)$$

for $i, j \in \{R, G, B\}, i \neq j$. (10)

Here, the motion-modulated scattered light $S_0 M(t)$ and the motion-modulated diffuse reflection light $R_0 M(t)$ are removed by the color difference operation. The remainder is a motion-modulated BVP signal $\text{Pulse}(t) M(t)$. Since there is still a motion modulation left in the color difference signal $D(t)$, it is still necessary to use ROI tracking before performing the color difference operation. After ROI tracking, the motion modulation $M(t)$ can be significantly reduced. Hence, the motion modulation left in the color difference signal $D(t)$ is residual.

III. METHODOLOGY

Based on the analysis of motion artifacts for facial video-based RIPPNG, we proposed a new motion-resistant RIPPNG method, as shown in Fig. 8. First, the subject's face is detected by a VJ face detector. After face detection, the ROI is selected on the cheeks region and speeded-up robust features (SURFs) points are detected on the face. Second, an ROI tracking module based on the KLT algorithm is utilized to adjust the ROI with the subject's motions. During ROI tracking in consecutive frames, raw RIPPNG signals in the green and red channels are obtained by averaging pixels in the ROI in the green and red channels, respectively. Next, an adaptive color difference operation is performed between raw RIPPNG signals in the green and red channels. This operation is called the adaptive green/red difference (GRD). Afterward, HR frequency can be estimated using the spectrum of the GRD signal. With the estimated HR frequency, the ABF is

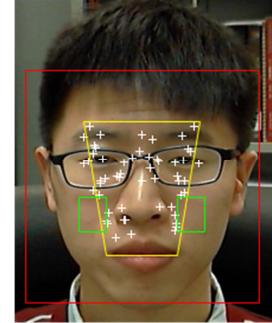


Fig. 9. ROI selection and tracking.

created to further remove motion artifacts and noises with frequencies that differ significantly to the HR frequency. Finally, a clear RIPPNG signal can be obtained. A video clip illustrating the proposed RIPPNG method is provided as supplementary material (<http://youtu.be/pS9R7Ed-uhI>). Detailed descriptions for each module are given in the following sections.

A. ROI Selection and Tracking

Since the face is assumed to be a planar extended Lambertian source, we treat the face as a rigid object and estimate its motion by tracking feature points. SURFs are utilized to detect points of interest that are scale and rotation invariant [21].

The VJ face detector is used to detect a rectangular facial region ($\text{width} \times \text{height}$) in the initial frame of a video (red rectangle in Fig. 9), and afterward a trapezoidal region (yellow trapezoid in Fig. 9) is automatically selected by ratios ($0.5 \times \text{width}$, $0.4 \times \text{width}$, and $0.58 \times \text{height}$). SURF points are detected within this trapezoid (white points in Fig. 9) and updated consecutively, per frame. The purpose of using such a trapezoid is to improve the robustness of ROI tracking, because

the SURF point concentration is high in this area and it can avoid background interference around the face boundary. The ROI for RIPPNG is selected on the cheeks region, indicated by the two green rectangles on the cheeks in Fig. 9. The size of each ROI rectangle is $0.12 \times \text{width}$ multiplied by $0.15 \times \text{height}$. The reason for selecting the cheeks region as the ROI for facial video-based RIPPNG is that the skin of the cheek can provide an RIPPNG signal with a higher signal-to-noise ratio (SNR), compared with other skin regions on the face [22].

Feature points can be reliably tracked by the KLT algorithm, and transformation matrices describing motions between subsequent frames are obtained during the tracking process. The transformation matrix describing the face motion from the $(k-1)$ th frame to the k th frame is H_k , which is expressed as an affine transformation matrix

$$H_k = \begin{pmatrix} a_k & b_k & 0 \\ c_k & d_k & 0 \\ e_k & f_k & 1 \end{pmatrix}. \quad (11)$$

The ROI in the $(k-1)$ th frame can be transformed into a coordinate of the k th frame using H_k

$$\begin{pmatrix} x_k \\ y_k \\ 1 \end{pmatrix} = H_k \begin{pmatrix} x_{k-1} \\ y_{k-1} \\ 1 \end{pmatrix}. \quad (12)$$

Hence, the location and shape of the ROI can be adjusted along with the subject's motions.

B. Adaptive GRD

According to (10), the amplitude of the color difference signal $D(t)$ is proportional to $(\gamma_i - \gamma_j)$. To keep the amplitude of $D(t)$ as large as possible, i and j should be selected by maintaining a large value of $(\gamma_i - \gamma_j)$. The wavelengths of blue light are from 450 to 495 nm, the wavelengths of green light are from 495 to 570 nm, and the wavelengths of red light are from 620 to 750 nm. It can be observed that $\gamma_G > \gamma_B > \gamma_R$ in Fig. 4. Compared to γ_G , γ_R is very small. Thus, we choose raw RIPPNG signals in the green and red channels to perform the color difference operation, and the GRD signal can be expressed as

$$\text{GRD}(t) = \frac{I_G(t)}{\alpha_G \beta_G} - \frac{I_R(t)}{\alpha_R \beta_R} = (\gamma_G - \gamma_R) S_0 \text{Pulse}(t) M(t). \quad (13)$$

However, $\alpha_G \beta_G$ and $\alpha_R \beta_R$ are still unknown. The digital color camera sensor can be treated as a simple spectrometer, as the camera sensor can split the light into the three different colors of RGB. Even though we cannot determine the illumination spectrum and the diffuse reflection spectrum separately, we can estimate the product of them using the received light intensity in RGB colors by a camera, because the radiant light of the skin is influenced by the illumination spectrum and the reflection spectrum in the form of consecutive multiplications. In the raw RIPPNG signal in (9), the amplitude of component $\gamma_i S_0 \text{Pulse}(t)$ varies with color, and the amplitude of $\gamma_i S_0 \text{Pulse}(t)$ is much smaller than that of S_0 and R_0 . To estimate $\alpha_G \beta_G$ and $\alpha_R \beta_R$ using raw RIPPNG

signals, we make an assumption that the raw RIPPNG signal becomes

$$I_i(t) = \alpha_i \beta_i (S_0 + R_0) M(t) \quad (14)$$

where the component $\gamma_i S_0 \text{Pulse}(t)$ is removed. Hence, the difference between the raw RIPPNG signal in green $I_G(t)$ and the raw RIPPNG signal in red $I_R(t)$ is only caused by $\alpha_G \beta_G$ and $\alpha_R \beta_R$. For the frame at time t , the raw RIPPNG signals $I_G(t)$ and $I_R(t)$ are known. The product of the normalized illumination spectrum and the normalized diffuse reflection spectrum in green and red can be approximately estimated by

$$\tilde{\alpha}_G(t) \tilde{\beta}_G(t) = \frac{I_G(t)}{\sqrt{I_R^2(t) + I_G^2(t) + I_B^2(t)}} \quad (15)$$

and

$$\tilde{\alpha}_R(t) \tilde{\beta}_R(t) = \frac{I_R(t)}{\sqrt{I_R^2(t) + I_G^2(t) + I_B^2(t)}} \quad (16)$$

respectively.

It should be noted that $\tilde{\alpha}_G(t) \tilde{\beta}_G(t)$ and $\tilde{\alpha}_R(t) \tilde{\beta}_R(t)$ cannot be directly utilized with raw RIPPNG signals for GRD, as shown in (13), because they are estimated based on the assumption that $\gamma_i S_0 \text{Pulse}(t)$ can be neglected compared with S_0 and R_0 . If we substitute $\tilde{\alpha}_G(t) \tilde{\beta}_G(t)$ and $\tilde{\alpha}_R(t) \tilde{\beta}_R(t)$ into (13), then $(\gamma_G - \gamma_R) S_0 \text{Pulse}(t)$ will be canceled using this assumption. Therefore, we first remove the power of S_0 and R_0 using a bandpass filter. A fixed bandpass filter (0.7–4 Hz) will be operated on $I_G(t)$ and $I_R(t)$ before the GRD operation, and the passband corresponds to an HR of 42–240 beats/min. Then, $\tilde{\alpha}_G(t) \tilde{\beta}_G(t)$ and $\tilde{\alpha}_R(t) \tilde{\beta}_R(t)$ are utilized with the bandpass-filtered RIPPNG signals for GRD

$$\begin{aligned} \text{GRD}(t) &= \frac{I_{Gf}(t)}{\tilde{\alpha}_G(t) \tilde{\beta}_G(t)} - \frac{I_{Rf}(t)}{\tilde{\alpha}_R(t) \tilde{\beta}_R(t)} \\ &= (\gamma_G - \gamma_R) S_0 \text{Pulse}(t) M_f(t) \end{aligned} \quad (17)$$

where $I_{Gf}(t)$ and $I_{Rf}(t)$ are the bandpass-filtered $I_G(t)$ and $I_R(t)$, respectively, and $M_f(t)$ is the motion modulation at frequencies between 0.7 and 4 Hz. In the signal $\text{GRD}(t)$, the motion-modulated scattered light $S_0 M(t)$ and the diffuse reflection light $R_0 M(t)$ are removed and the BVP signal is maintained with the highest amplitude. Since $M_f(t)$ has already been significantly reduced by ROI tracking, the HR frequency can be easily determined by finding the peak in the frequency domain via the fast Fourier transform (FFT) of $\text{GRD}(t)$. Since $\alpha_G \beta_G$ and $\alpha_R \beta_R$ are not fixed and can be changed with environmental variation, the proposed color difference method is more adaptive compared with the skin tone-based method.

C. Adaptive Bandpass Filter

After ROI tracking and adaptive GRD, there is still a motion modulation $M_f(t)$ left in the RIPPNG signal, as shown in (17). The ABF can reduce the motion modulation in one more step. A typical PPG signal is shown in Fig. 10, and its normalized power spectral density is shown in Fig. 11.

The spectrum of a typical PPG signal has a sharp predominant peak and another two clear harmonics. The signal

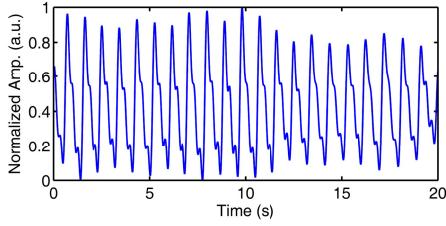


Fig. 10. Typical PPG signal.

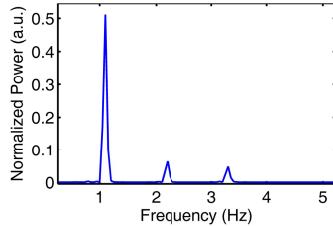


Fig. 11. Normalized power spectral density of a typical PPG signal.

power is highly centralized around the frequency of the HR, $2 \times \text{HR}$ and $3 \times \text{HR}$. Since the HR frequency can be well estimated using the spectrum of the GRD signal, we can design a bandpass filter adaptively with the estimated HR according to spectral characteristics of PPG signals. The frequency response of the ABF is shown in Fig. 12.

The ABF can maintain the power of the RIPP signal around frequencies of HR, $2 \times \text{HR}$ and $3 \times \text{HR}$, and remove the artifacts and noises at other frequencies. The frequency response of the ABF is

$$\text{ABF}(f) = \begin{cases} 1, & \text{HR}_f - 0.15 \leq f \leq \text{HR}_f + 0.15 \text{ Hz} \\ 1, & 2\text{HR}_f - 0.15 \leq f \leq 2\text{HR}_f + 0.15 \text{ Hz} \\ 1, & 3\text{HR}_f - 0.15 \leq f \leq 3\text{HR}_f + 0.15 \text{ Hz} \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where HR_f represents the HR frequency.

In this paper, the window size of the ABF operation for RIPP is 20 s. To avoid phase distortion, the ABF is implemented using an FFT. The frequency resolution of HR_f is 1/20 Hz, which is determined by the window size of the ABF. In (18), the first sub-band of the ABF is $\text{HR}_f \pm 0.15$ Hz. Given the frequency resolution of 1/20-Hz per bin for the ABF, the sub-band covers seven bins around the HR frequency, corresponding to ± 0.175 Hz (± 10.5 beats/min). The second and third sub-bands have the same sub-bandwidth. This sub-bandwidth is an experimental value that allows for HR variation within 20 s. The same sub-band has also been adopted for calculating the SNR of the RIPP signal [11]. There are 10 subjects in our database, and seven sets of continuous HR measurements lasting 20 s were performed for each subject. Of a total of 70 sets of HR data, the maximum 1.96 standard deviation and the maximum difference of HR values within 20 s are 8.67 and 15.15 beats/min, respectively. Hence, a sub-bandwidth of ± 10.5 beats/min can cover the normal HR variation within 20 s. For a long-term RIPP analysis, the ABF operation can be shifted along the GRD signal.

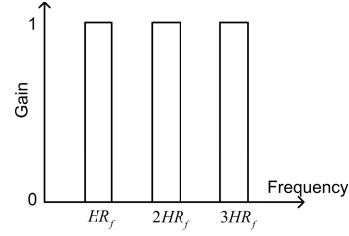


Fig. 12. Frequency response of the ABF for RIPP.

IV. ASSESSMENT DETAILS

In our experiments, we wish to assess the performances of the proposed method in three ways: 1) the effect of ROI tracking on reducing motion artifacts; 2) the HR measurement accuracy of the proposed RIPP for moving subjects; and 3) the robustness of the proposed RIPP with relation to different motions within the 6 DOF. An instantaneous HR measurement is used to test the RIPP performance. After a pulse wave is extracted from a facial video by the RIPP, a simple HR estimation algorithm is used to calculate the instantaneous HR in the time domain. A running rectangular window with a size of 4 s will locate a 4-s long pulse wave each second. The time locations of systolic peaks within the rectangular window are detected using a simple peak detection method [*findpeaks()* in MATLAB]. The instantaneous HR is calculated as

$$\text{HR}_{\text{ins}} = \frac{60 \times (\text{Num} - 1)}{\text{Time}(j) - \text{Time}(i)} \quad (19)$$

where Num is the number of peaks within the rectangular window, $\text{Time}(j)$ is the location of the last peak within the rectangular window on the time axis, and $\text{Time}(i)$ is the location of the first peak within the rectangular window on the time axis. To eliminate the interference of diastolic peaks, the minimum peak separation is set to $0.5 \times (60/\text{HR}_{\text{ave}})$, where HR_{ave} is the average HR during a long period (e.g., 20 s), which can be estimated by finding a peak in the frequency domain.

A. Material

In total, 10 healthy volunteers (seven males and three females) ranging in age from 20 to 33 participated in the study. None of them had any known cardiovascular diseases. Informed consent was obtained from each subject, and the study was approved by the Research Committee of City University of Hong Kong. A low-cost webcam (Logitech C270) was used as the imaging device. The webcam worked at a frame rate of 30 frames/s. All videos were recorded in color space (24-bit RGB) with a resolution of 640×480 , and saved in uncompressed AVI format. During the video recording process, there were no dedicated light sources utilized except for normal indoor fluorescent light. Seven videos were recorded by the webcam for each subject. The duration of each video was 20 s. The subject sat in front of the webcam at a distance of approximately 75 cm.

In the first six videos, the subject was asked to make yawing, pitching, rolling, surging, swaying, and heaving motions, respectively. Visible markers were used to help the subjects standardize and limit their motions. Yawing was kept in the range of $\pm 35^\circ$. Pitching was kept in the range of $\pm 30^\circ$. Rolling was kept in the range of $\pm 45^\circ$. Surging was kept in the range of ± 30 cm. Swaying was kept in the range of ± 40 cm. Heaving was kept in the range of ± 20 cm. All the above ranges were determined from feasible motions limited by the sitting posture or sustainable ROI tracking. In the seventh video, each subject was asked to move their body and head freely with large-scale motions, as long as their head still appeared in the field of view of the webcam. The first six videos were used to test the robustness of RIPPNG to the six separate motions. The last video was used to check the HR measurement accuracy for moving subjects. During the video recording, an U.S. Food and Drug Administration approved commercial vital signs monitor (EDAN M3 Vital Signs Monitor) was contacted to the subject's index finger to record the BVP wave for [23]. The index finger was prevented from moving to obtain an accurate reference.

B. Benchmark Methods

To benchmark the proposed RIPPNG method, we compare the experimental results with two state-of-the-art RIPPNG methods [5], [12]. In the first, ICA is combined with ROI tracking using the VJ algorithm [5]. We call this method the ICA-based method. We use the robust ICA algorithm with maximized kurtosis contrast for the implementation of the ICA-based method [24], and the ICA sorting problem is solved by selecting the ICA output component with the highest peak in the power spectral density [25]. Other implementation details are the same, as described in [5]. In the second method [12], specular reflection is assumed to exist and a standardized skin tone is used to compensate for nonwhite illumination. A color difference operation is performed on RIPPNG signals in RGB channels with a set of fixed linear combination coefficients. Then, the combination coefficients are refined based on the standard deviations of color difference signals. For convenience, we call this method the skin tone-based method. Both the overlap-add Hanning window and running average window for normalization had a size of 2 s. All the algorithms were implemented using custom software written in MATLAB.

C. Statistics

We employ Bland–Altman analysis as the main method to verify agreement between RIPPNG and CPPG [26]. We examine an agreement between methods of measurement with multiple observations per subject [27]. In addition, Pearson's correlation coefficients (CCs) and the corresponding p -values are calculated to measure the linear dependency between RIPPNG and CPPG [28]. When plotting Bland–Altman plots, the HR measurement results of CPPG are regarded as the gold standard. The HR measurement differences between RIPPNG and CPPG will be plotted against the results of CPPG. The measurement bias and 95% limits of agreement between the two will be

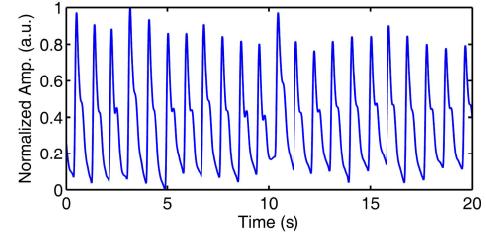


Fig. 13. Reference CPPG waveform for the sample video.

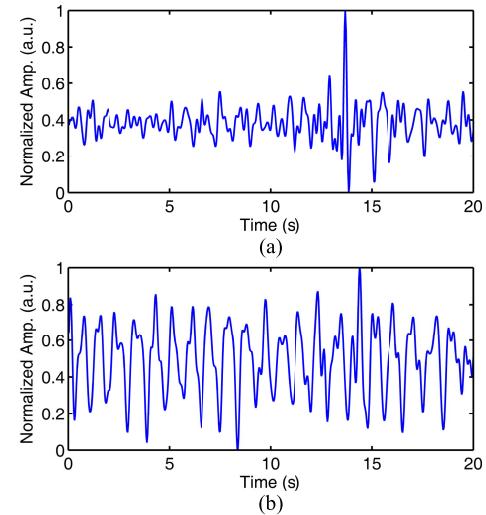


Fig. 14. RIPPNG waveforms for the sample video measured using (a) original ICA-based RIPPNG and (b) modified ICA-based RIPPNG with ROI tracking.

calculated to check the accuracy of RIPPNG. The 95% limits of agreement are defined as the bias ± 1.96 times the standard deviation of the difference, which shows how far apart the HR measurements made by RIPPNG are likely to be for most individuals.

V. EXPERIMENTAL RESULTS

A. Effect of the Proposed ROI Tracking on RIPPNG

To evaluate the effect of the ROI tracking module of the proposed method, we replace ROI tracking based on the VJ algorithm in the ICA-based method and the skin tone-based method with the proposed ROI tracking module. Ten videos of combinational motions were analyzed by the original benchmark methods and the revised benchmark methods. Fig. 13 shows a reference CPPG waveform for one sample facial video of the 10 videos. For the same video, the RIPPNG waveforms measured by the original ICA-based RIPPNG and the modified ICA-based RIPPNG with ROI tracking are shown in Fig. 14. The RIPPNG waveforms measured by the original skin tone-based RIPPNG and the modified skin tone-based RIPPNG with ROI tracking are shown in Fig. 15. The Bland–Altman plots with multiple instantaneous HR measurements per subject for the original ICA-based RIPPNG and the modified ICA-based RIPPNG with ROI tracking are shown in Fig. 17. The Bland–Altman plots with multiple instantaneous HR measurements per subject for the original skin tone-based

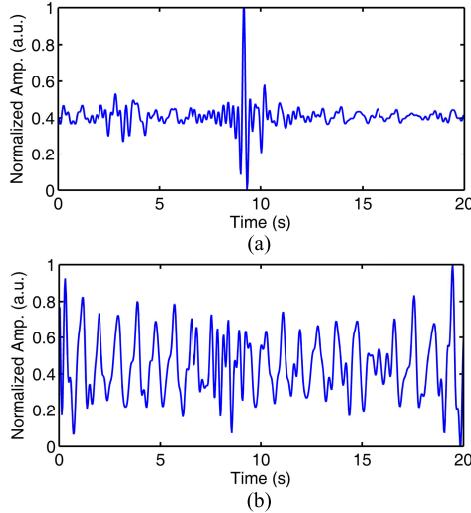


Fig. 15. RIPP waveform for the sample video measured using the (a) original skin tone-based RIPP and (b) modified skin tone-based RIPP with ROI tracking.

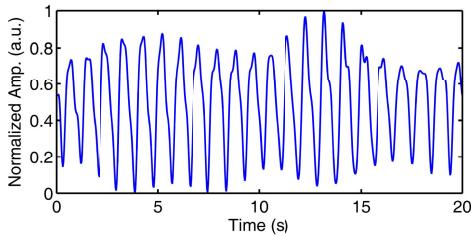


Fig. 16. RIPP waveform for the sample video measured using the proposed RIPP.

RIPP and the modified skin tone-based RIPP with ROI tracking are shown in Fig. 18. The gold standard for the Bland–Altman plots is the HR measurement by CPPG, shown on the horizontal axis in Figs. 17 and 18.

In Figs. 14 and 15, it can be observed that the proposed ROI tracking module can significantly improve the signal quality of ICA-based RIPP and skin tone-based RIPP. There are spikes in the RIPP waveform due to ROI fluctuation or loss, as shown in Figs. 14(a) and 15(a). These spikes can be removed by ROI tracking, as shown in Figs. 14(b) and 15(b). The statistics in Figs. 17 and 18 also show the effectiveness of ROI tracking. For ICA-based RIPP, the HR measurement bias between CPPG and the original ICA-based RIPP is -16.5 beats/min, with 95% limits of agreement -121.5 to 88.4 beats/min. With the help of the proposed ROI tracking module, the HR measurement bias was 0.7 beats/min, with 95% limits of agreement -10.8 to 12.2 beats/min. For skin tone-based RIPP, the HR measurement bias between CPPG and the original skin tone-based RIPP is -32.2 beats/min, with 95% limits of agreement -161.3 to 96.8 beats/min. With the help of the proposed ROI tracking module, the HR measurement bias becomes 5.6 beats/min, with 95% limits of agreement -15.5 to 26.8 beats/min. The CC between the instantaneous HR measured by CPPG and ICA-based RIPP increased from 0.1427 (p -value = 0.11) to 0.8394 (p -value < 0.001) with the help of the proposed ROI tracking module.

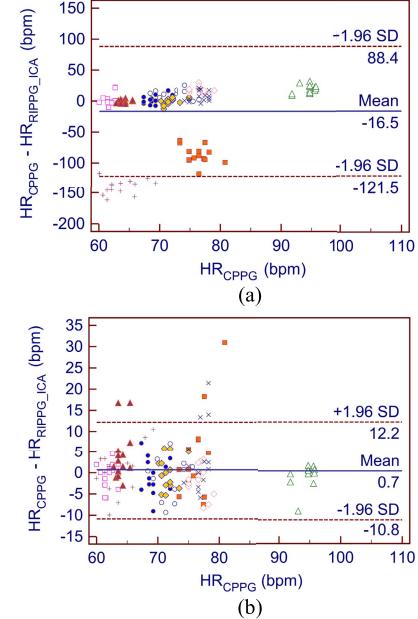


Fig. 17. Bland–Altman plots with multiple instantaneous HR measurements per subject. (a) Comparison of CPPG and the original ICA-based RIPP. (b) Comparison of CPPG and the modified ICA-based RIPP with ROI tracking.

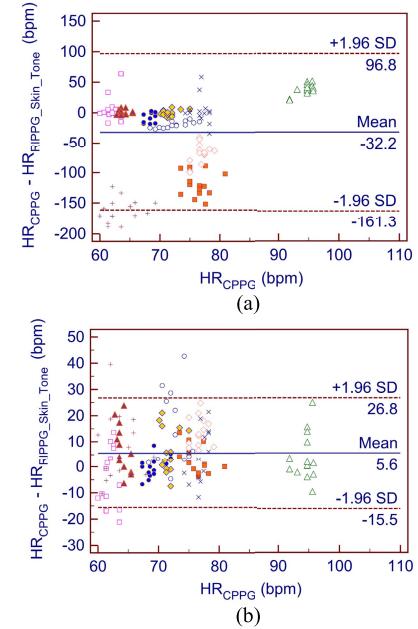


Fig. 18. Bland–Altman plots with multiple instantaneous HR measurements per subject. (a) Comparison of CPPG and the original skin tone-based RIPP. (b) Comparison of CPPG and the modified skin tone-based RIPP with ROI tracking.

module. The CC between the instantaneous HR measured by CPPG and skin tone-based RIPP increased from 0.1366 (p -value = 0.14) to 0.6191 (p -value < 0.001) with the help of the proposed ROI tracking module.

B. Performance of the Proposed RIPP Method

We used the videos of combined motions to evaluate the complete proposed RIPP method (Fig. 8). For the same

TABLE I
ROBUSTNESS OF THE PROPOSED RIPPNG TO SEPARATE MOTIONS IN THE 6 DOF

Statistic	Yawing	Pitching	Rolling	Surging	Swaying	Heaving
Bias (bpm)	-0.4	-0.1	-0.1	-0.1	0.1	0.5
Limits of agreement	Upper limit (bpm)	4.8	6.3	6.4	5.5	6.7
	Lower limit (bpm)	-5.7	-6.4	-6.5	-5.7	-6.4
CC	0.97	0.96	0.97	0.97	0.96	0.93
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

TABLE II
ROBUSTNESS OF THE MODIFIED ICA-BASED RIPPNG WITH ROI TRACKING TO SEPARATE MOTIONS IN THE 6 DOF

Statistic	Yawing	Pitching	Rolling	Surging	Swaying	Heaving
Bias (bpm)	2.3	1.0	5.4	0.6	1.1	1.3
Limits of agreement	Upper limit (bpm)	22.1	14.3	28.1	9.8	11.3
	Lower limit (bpm)	-17.4	-12.3	-17.3	-8.5	-9.0
CC	0.59	0.82	0.49	0.92	0.92	0.89
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

TABLE III
ROBUSTNESS OF THE MODIFIED SKIN TONE-BASED RIPPNG WITH ROI TRACKING TO SEPARATE MOTIONS IN THE 6 DOF

Statistic	Yawing	Pitching	Rolling	Surging	Swaying	Heaving
Bias (bpm)	2.2	6.3	8.1	1.6	3.0	8.5
Limits of agreement	Upper limit (bpm)	46.2	44.4	46.2	50.0	21.0
	Lower limit (bpm)	-41.8	-31.9	-30.1	-46.7	-14.9
CC	0.09	0.33	0.1	0.1	0.75	0.27
p-value	0.34	<0.001	0.26	0.47	<0.001	<0.001

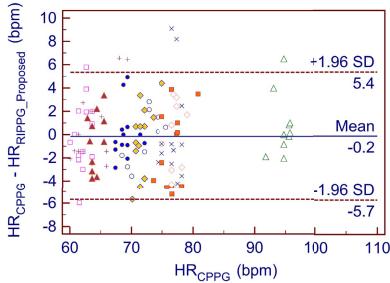


Fig. 19. Bland-Altman plot with multiple instantaneous HR measurements per subject for the comparison of CPPG and the proposed RIPPNG.

sample video used in Figs. 14 and 15, a good-quality RIPPNG waveform can be measured using the proposed method, as shown in Fig. 16. The Bland-Altman plot with multiple instantaneous HR measurements per subject for the comparison of CPPG and the proposed RIPPNG is shown in Fig. 19. Even though subjects made large-scale motions, the HR measurement bias between CPPG and the proposed RIPPNG is only -0.2 beats/min. Moreover, the 95% limits of agreement are between -5.7 and 5.4 beats/min. The CC between the instantaneous HR measured by CPPG and the proposed RIPPNG is 0.9542 (p -value < 0.001). All of these performance parameters were greatly improved compared to the results of the state-of-the-art RIPPNG methods, as shown in Section V-A.

C. Robustness of RIPPNG to Separate Motions

To evaluate the robustness of the proposed RIPPNG to separate motions in the 6 DOF (Fig. 5), the video sets with these motions were analyzed using the proposed RIPPNG.

These separate motions consisted of yawing, pitching, rolling, surging, swaying, and heaving. HR measurements were performed using the proposed RIPPNG on each video set corresponding to a separate motion. Bland-Altman analysis of the HR measurements was performed for each separate motion. We analyzed the agreement between the methods of measurement with multiple observations per subject. The Pearson's correlation was also calculated for the HR measurements in each video set. The statistical results are shown in Table I. The same analyses were performed using ICA-based RIPPNG and skin tone-based RIPPNG. The ROI tracking modules in those two state-of-the-art methods were replaced with the proposed ROI tracking module; this is because the original methods cannot handle the large-scale motions in the experiments. In addition, the statistical results of the two methods are shown in Tables II and III. By comparing Table I with Tables II and III, the proposed RIPPNG is found to have the best performance, and there is no obvious sensitivity to a certain type of motion.

VI. DISCUSSION

Good-quality ROI tracking plays an important role in reducing motion artifacts for facial video-based RIPPNG, as the results show in Section V-A. After ROI tracking based on the VJ algorithm was replaced by the proposed ROI tracking module, the accuracy of HR measurement was significantly improved. The advantage of our ROI tracking module is that the KLT algorithm can handle more complicated motions than the VJ algorithm, and ROI selection on the cheeks can also improve the SNR of the RIPPNG signal. In the analysis of the robustness of the proposed RIPPNG to separate motions, as shown in Section V-C, we did not find any obvious sensitivity to a certain separate motion. The proposed RIPPNG exhibited

similar performance for all kinds of separate motions. This result did not match the results of the analysis of the effect of ideal ROI tracking in Section II-A, in that ROI tracking cannot compensate well for swaying and heaving motions. This is because the assumption of a planar rigid object for the human face is not sufficiently accurate. The face is a curved 3-D surface, rather than a planar surface, so performing only an affine transformation on the whole face cannot implement ideal ROI tracking. In practice, motion compensations for all separate motions are not ideal. In addition, the assumption of uniform illumination in Section II-A is not possible in everyday environments. Hence, the performances of the proposed RIPPg for all kinds of separate motions are balanced.

By comparing Fig. 19 with Figs. 17(b) and 18(b), we can observe that the proposed RIPPg method has a better performance than the state-of-the-art RIPPg methods. In this comparison, the proposed ROI tracking module was used by both ICA-based RIPPg and skin tone-based RIPPg, and hence the comparison shows the effectiveness of the proposed adaptive GRD method and the ABF. There is a basic assumption of independence (non-Gaussianity) between the BVP signal and motion artifacts for RIPPg in the ICA-based method. Referring to (3), we can treat $S_0M(t)$ and $R_0M(t)$ as motion artifacts and $\text{Pulse}(t)M(t)$ as the BVP signal. The BVP signal is not absolutely independent with motion artifacts, because all of them are modulated by an identical motion modulation $M(t)$. When $M(t)$ is large, ICA will not work well as shown in Fig. 17 and Table II. For skin tone-based RIPPg, there is an assumption of a standardized skin tone. Referring to (3), the standardized skin tone is actually a fixed approximation for β_i (reflection spectrum), so β_i is eliminated with this approximation. Afterward, skin tone-based RIPPg uses normalization of $I_i(t)$ to further eliminate α_i (illumination spectrum). The proposed GRD method can adaptively eliminate $\alpha_i(t)$ and $\beta_i(t)$ simultaneously without using a fixed approximation. Furthermore, the assumption of specular reflection on the skin will also affect the accuracy of skin tone-based RIPPg. The surface of human skin is not optically smooth enough for specular reflection.

The improvement in RIPPg performance exhibited by our proposals demonstrates that the proposed optical RIPPg signal model works better in describing the origins of the RIPPg signal and motion artifacts, compared with the state-of-the-art methods.

VII. CONCLUSION

RIPPg revolutionizes the operation mode of conventional CPPG. The required reduction of motion artifacts is an unavoidable problem in the development of RIPPg for clinical field. Previous works have focused on reducing motion artifacts from the perspective of signal/image processing. In this paper, it was shown that the RIPPg signal arises from radiant intensity variation of the skin due to BVP. We built an optical signal model to analyze the motion artifacts of RIPPg, based on the optical properties of human skin. ROI selection on the cheeks was combined with ROI tracking. An adaptive color

difference operation between the green and red channels was proposed based on optical RIPPg signal analysis. An ABF was proposed based on the spectral characteristics of PPG signals, which can further improve the SNR of the RIPPg signal after the determination of the HR frequency. In an experimental comparison with state-of-the-art RIPPg methods, the improvement in HR measurement accuracy verified the effectiveness of our proposals. We believe that the motion-resistant RIPPg has good potential for applications in special situations, such as lie detection, infant monitoring, and telemedicine.

APPENDIX

The irradiance on the image ROI on the imaging plane of a camera is analyzed within the 6 DOF. Uniform environmental illumination is assumed. The different variables are defined in Section II.

A. Yawing

When the face tilts with a yawing angle Φ , the entrance pupil size projected on the face normal V is $A \cos \Phi$, so the radiant power received by the camera is

$$P_w = \frac{A\pi NW^2 \cos \Phi}{2S^2}. \quad (20)$$

With a yawing angle Φ , the size of the image ROI becomes

$$S_{\text{image_ROI}} = \frac{4W^2 f^2 \cos \Phi}{S^2}. \quad (21)$$

Hence, the irradiance on the image ROI is

$$I_{\text{image_ROI}} = \frac{A\pi N}{8f^2}. \quad (22)$$

In (22), there is no variation for the irradiance of the image ROI with a yawing motion.

B. Pitching

The case of a pitching motion is very similar to that of a yawing motion, except that the pitching angle is located at the vertical axis y , not on the horizontal axis x . Thus, the same conclusion as that for yawing motion can be obtained. With ideal ROI tracking, a pitching motion has no effect on the irradiance of the image ROI.

C. Rolling

When there is a rolling motion with an angle of Φ , the face is still on the frontal plane with the same distance S away from the camera. The radiant power received by the camera is still

$$P_w = \frac{A\pi NW^2}{2S^2}. \quad (23)$$

The size of the image ROI also does not change. Therefore, the irradiance of the image ROI is still

$$I_{\text{image_ROI}} = \frac{A\pi N}{8f^2}. \quad (24)$$

Thus, a rolling motion does not affect the irradiance of the image ROI.

D. Surging

A surging motion is a translation on the z -axis. The distance S will change along with the forward/backward translation. The face is still on the frontal plane, so the radiant power received by the camera is

$$P_w = \frac{A\pi NW^2}{2S^2}. \quad (25)$$

The size of the image ROI is

$$S_{\text{image_ROI}} = \frac{4W^2 f^2}{S^2}. \quad (26)$$

Even though S changes, the irradiance of the image ROI does not change, and it is still

$$I_{\text{image_ROI}} = \frac{A\pi N}{8f^2}. \quad (27)$$

Hence, a surging motion does not affect the irradiance of the image ROI.

E. Swaying

A swaying motion is a translation on the x -axis. The line from the camera to the face center makes an angle Φ to the face normal V , resulting from the left/right translation. The radiant intensity received by the camera from the radiation of the face is

$$H = \pi N \sin^2 \theta \cos^4 \Phi. \quad (28)$$

The radiant power received by the camera is

$$P_w = \frac{A\pi NW^2 \cos^4 \Phi}{2S^2}. \quad (29)$$

The size of the image ROI does not change, owing to the invariant quantity S . Therefore, the irradiance of the image ROI is

$$I_{\text{image_ROI}} = \frac{A\pi N \cos^4 \Phi}{8f^2}. \quad (30)$$

A swaying motion has an impact on the irradiance of the image ROI with a factor of $\cos^4 \Phi$. Even if ROI tracking works well, a swaying motion can still cause motion artifacts.

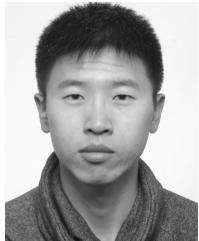
F. Heaving

A heaving motion corresponds to an up/down translation on the y -axis. The effect of a heaving motion on the irradiance of the image ROI is the same as that of a swaying motion. A heaving motion will cause a $\cos^4 \Phi$ factor on the irradiance of the image ROI, as shown in (30).

REFERENCES

- [1] A. B. Hertzman and C. R. Speelman, "Observations on the finger volume pulse recorded photoelectrically," *Amer. J. Physiol. Meas.*, vol. 119, pp. 334–335, 1937.
- [2] 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.
- [3] E. Jonathan and M. Leahy, "Investigating a smartphone imaging unit for photoplethysmography," *Physiol. Meas.*, vol. 31, no. 11, pp. N79–N83, 2010.
- [4] W. Verkruyse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [5] 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.
- [6] 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 EMBC*, Aug./Sep. 2012, pp. 2174–2177.
- [7] M. Y. Wong, E. Pickwell-MacPherson, and Y. T. Zhang, "Contactless and continuous monitoring of heart rate based on photoplethysmography on a mattress," *Physiological Meas.*, vol. 31, no. 7, pp. 1065–1074, 2010.
- [8] T. Aoyagi, N. Kobayashi, and T. Sasaki, "Apparatus for determining the concentration of a light-absorbing material in blood," U.S. Patent 4 832484, May 23, 1989.
- [9] G. Cennini, J. Arguel, K. Akşit, and A. van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Exp.*, vol. 18, no. 5, pp. 4867–4875, 2010.
- [10] 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, pp. 077010–1–077010–9, 2011.
- [11] N. Kawasaki, "Improving motion robustness of contact-less monitoring of heart rate using video analysis," M.S. thesis, Dept. Math. Comput. Sci., Eindhoven Univ. Technology, Eindhoven, The Netherlands, 2011.
- [12] 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.
- [13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. CVPR*, vol. 1, Dec. 2001, pp. I-511–I-518.
- [14] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th IJCAI*, 1981, pp. 674–679.
- [15] C. Tomasi and T. Kanade, "Detection and tracking of point features," Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, Apr. 1991.
- [16] S. Takano, W. Fujita, and E. Okada, "Multi-layered models for prediction of diffuse reflectance spectra of skin and lip," in *Biomed. Opt. 3-D Imag., OSA Tech. Dig. (CD)*, 2010, paper BME4.
- [17] R. R. Anderson and J. A. Parrish, "The optics of human skin," *J. Invest. Dermatol.*, vol. 77, no. 1, pp. 13–19, 1981.
- [18] G. Zonios, J. Bykowski, and N. Kollias, "Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed *in vivo* using diffuse reflectance spectroscopy," *J. Invest. Dermatol.*, vol. 117, no. 6, pp. 1452–1457, 2001.
- [19] J. A. Crowe and D. Damianou, "The wavelength dependence of the photoplethysmogram and its implication to pulse oximetry," in *Proc. 14th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 6, Oct./Nov. 1992, pp. 2423–2424.
- [20] W. J. Smith, "Principles of radiometry and photometry," in *Modern Optical Engineering: The Design of Optical Systems*, 4th ed. New York, NY, USA: McGraw-Hill, 2008.
- [21] H. Bay, A. Ess, T.uytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [22] G. Lempe, S. Zaunseder, T. Wirthgen, S. Zipser, and H. Malberg, "ROI selection for remote photoplethysmography," in *Bildverarbeitung für die Medizin*. Berlin, Germany: Springer-Verlag, pp. 99–103, 2013.
- [23] O. P. Faris. (Aug. 30, 2013). *RE: K131818 Vital Signs Monitor M3 & M3A*. [Online]. Available: <http://owen.faris@fda.hhs.gov>
- [24] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 248–261, Feb. 2010.
- [25] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [26] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

- [27] J. M. Bland and D. G. Altman, "Agreement between methods of measurement with multiple observations per individual," *J. Biopharmaceutical Statist.*, vol. 17, no. 4, pp. 571–582, 2007.
- [28] B. Jacob, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009, pp. 1–4.



Litong Feng (S'12) received the B.E. degree in electronic science and technology from Harbin Institute of Technology, Harbin, China, in 2008 and the M.E. degree in optical engineering from Tianjin Jinhang Institute of Technical Physics, Tianjin, China, in 2011. He is currently working toward the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong.

His research interests include video processing for vital signs and optical system design.



Lai-Man Po (M'92–SM'09) received the B.S. and Ph.D. degrees in electronic engineering from City University of Hong Kong, Hong Kong, in 1988 and 1991, respectively.

He has been with the Department of Electronic Engineering, City University of Hong Kong, since 1991, where he is currently an Associate Professor and the Laboratory Director of Texas Instruments Educational Training Centre. He has authored over 140 technical journal and conference papers. His research interests include image and video coding with an emphasis on fast encoding algorithms, new motion compensated prediction techniques, and 3-D video processing.

Dr. Po is a member of the Technical Committee on Multimedia Systems and Applications and the IEEE Circuits and Systems Society. He was the Chairman of the IEEE Signal Processing Hong Kong Chapter in 2012 and 2013. He was an Associate Editor of *HKIE Transactions* in 2011 to 2013. He also served on the organizing committees of the IEEE International Conference on Acoustics, Speech, and Signal Processing in 2003, the IEEE International Conference on Image Processing in 2010, and other conferences.



Video.

Xuyuan Xu (S'11) received the B.E. degree in information engineering from City University of Hong Kong, Hong Kong, in 2010, where he is currently working toward the Ph.D. degree with the Department of Electronic Engineering.

His research interests include 3-D video coding and 3-D view synthesis.

Mr. Xu received the Best Tertiary Student Project of the Asia-Pacific International and Communication Award in 2010 for his final year project entitled Stereoscopic Video Generation from Monoscopic



Yuming Li received the B.Eng. and M.Phil. degrees from Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with City University of Hong Kong, Hong Kong.

His research interests include image and video processing, multiscale analysis, and machine learning.



Ruiyi Ma received the B.Eng. degree from Guangdong University of Technology, Guangzhou, China, in 2013. He is currently working toward the M.Sc. degree with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong.

His research interests include image processing, motion detection, and computer vision.