Using blood volume pulse vector to extract rPPG signal in infrared spectrum

Lin, X.

Award date:
2014

Technische Universiteit
Eindhoven
University of Technology

DEPARTMENT OF ELECTRICAL ENGINEER
VIDEO & IMAGE PROCESSING GROUP AT PHILIPS RESEARCH

Using blood volume pulse vector to extract rPPG signal in infrared spectrum

Master Thesis

Author: Xueming Lin
Email: < x.lin@student.tue.nl >
ID: 0868944

Supervisors:
Prof. Dr. Gerard de Haan (TU/e & Philips)
Dr. Sander Stuijk (TU/e)

Eindhoven, August 2014

Table of Contents

Abstract

Pulse rate, the number of arterial pulses per minute, is caused by the contractions of the heart and thereby considered closely related to heart rate. Abnormal heart rate may be an indication of illness. Therefore, it is crucial to measure heart rates in a non-invasive and continuous manner. Remote photoplethysmography (rPPG) techniques enable contactless heart rate monitoring by detecting blood volume variations during each heartbeat using regular RGB cameras. Recent research has developed successful rPPG methods in ambient light conditions. State-of-the-art rPPG methods are aiming to improve motion robustness. However, for subjects like premature babies, who need 24/7 continuous vital sign monitoring, it is preferable to have a method that can work in darkness. In this paper, we present the (existing) PBV method and study its application in the infrared spectrum. From a population of 41 stationary subjects, we obtain the PBV signatures using 5 different methods for both monochrome cameras and an RGB camera. Using the most reliable PBV signatures with least variations, we extract the pulse signals and heart rates from stationary subjects. With the simple peak detection in frequency domain, the results show the SNRs for monochrome cameras range from 2.5 dB to 9.3 dB and the average SNR over all 41 stationary subjects is around 6.0 dB. For the RGB camera, the SNRs range from 0.6 dB to 8.6 dB and the average SNR is 4.7 dB. The accuracy of the PBV method in the infrared spectrum is over 96% compared to the reference sensor. After that, we test the motion robustness on 13 subjects with three types of head motions: rotation, scaling and translation. The average SNRa for all motions is found to be around 2.0 dB and the correlation between our method and the reference is greater than 0.98. The accuracy is found to be higher than 91%. In a further attempt to improve the motion robustness, we aim to obtain the optimal PBV vector by changing the optical filters in the monochrome cameras. Considering the limited optical filters available, we find out the three optimal filters shall have center wavelengths at 661 nm, 720 nm and 842 nm, resulting in the PBV signature with maximum angle with the motion vector [1 1 1]. Using these new filters in the monochrome cameras, we carry out another experiment on 11 stationary subjects. The results show an accuracy of over 98% and the average SNR is increased to 6.4 dB. Overall, the performance of the PBV method in infrared spectrum for stationary subjects is nearly identical to a commercial patient monitor. This method shows a great potential for 24/7 vital signs monitoring, particularly for sleeping subjects without significant motions.

Key words: Photoplethysmography (PPG), remote PPG (rPPG), pulse, heart rate, vital sign, biomedical monitoring, blood volume pulse (PBV), infrared (IR).

Acknowledgements

I would like to express my sincere gratitude to my supervisor from both Philips and university, Prof. Dr. Gerard de Haan, for his expertise, guidance, patience and support throughout my work. Besides, I would like to thank Dr. Sander Stuijk for his guidance and assistance in writing reports and following the correct graduation procedures. Furthermore, I am very grateful to Mark van Gastel, who has given me strong supports for my work. I really enjoyed working with them and have learned a lot from them, which will be a valuable asset to me.

I would also like to thank all of the Philips colleagues who have assisted me in the experimental setup, especially Willem Verkruijsse, Ihor Kirenko, Patriek Bruins, and Marek Bartula. Also, I am grateful to all of the volunteers that participated in my recordings experiment.

I. Introduction

Heart rate (pulse rate), typically expressed as beats per minute (bpm), is a critical vital sign for medical diagnosis. The pulse rate can vary during physical exercise, sleep, anxiety and illness, when the body needs to absorb oxygen (O_2) and excrete carbon dioxide (CO_2). The normal resting heart rate for healthy adults ranges from 60 to 100 bpm. Children and women tend to have faster heart rates than men. Athletes, who do a lot of cardiovascular conditioning, may have heart rates around 40 bpm and experience no problems [1].

Heart rate variability (HRV), measured by the variation in the beat-to-beat interval, indicates the degree to which the sympathetic and parasympathetic nervous systems work together to modulate cardiac activity. Some studies have shown that HRV can serve as a predictor of sudden cardiac death [2], hemorrhagic shock [3] and impending septic shock [4]. Abnormal (usually reduced) HRV may also be associated with congestive heart failure, depression and poor survival in premature babies [5]. Therefore, there is a strong need to measure heart rate in a non-invasive and continuous manner.

In the 1930s, photoplethysmography (PPG) was first proposed as an optical technique to detect pulse rate [6]. During each cardiac cycle, the heart pumps blood against the resistance of the vascular bed. This pressure pulse will then distend the arteries and arterioles in the subcutaneous tissue, thereby changing the blood volume and causing the consequential variations of the light absorption. Based on this observation, PPG measures pulse rate by illuminating the skin tissue with a light-emitting diode (LED) and then detecting the variations in light intensity by a photo-detector, either transmitted or reflected from the skin. Because the skin is so richly perfused, it is relatively easy to detect the pulsatile component of the cardiac cycle. The PPG signal has two components, conventionally referred to as AC and DC, shown in Figure 1. The AC component is directly attributable to variation in blood volume in the skin, which is synchronous with the heart rate. The AC component is superimposed onto a large DC component which is attributable to the bulk absorption of the skin tissue. Since PPG is entirely non-invasive, it has been predominantly applied in hospitals in the form of contact device clipped to finger/toe/ear, or patched to skin.

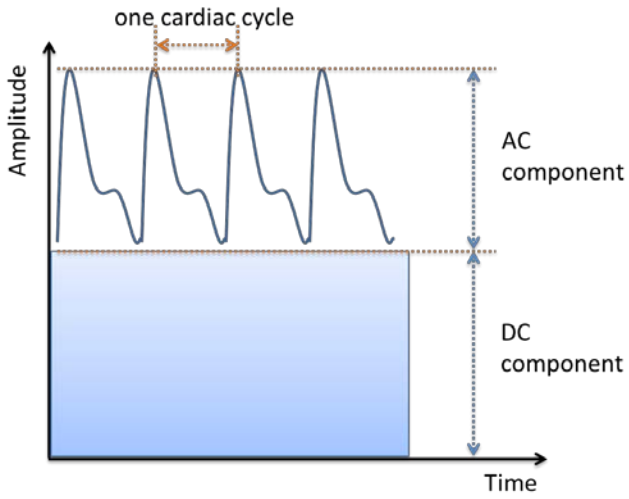

Figure 1. A typical PPG signal that consists of AC and DC component.

This contact-based PPG method inevitably causes some burdens or discomfort to subjects. There is a growing interest in contactless HR measurement, especially for subjects whose skin is too sensitive to wear devices, such as premature babies and severely burned patients. This need triggered the new technology – remote-PPG (rPPG) [7], [8]. Recent studies have shown some successful rPPG methods using a regular RGB camera in ambient light conditions [9]-[14].

Verkruysse *et al.* observed that the PPG signal has different relative strengths in the three color channels of a camera. The green channel contains the strongest PPG signal, consistent with the fact that (oxy-)hemoglobin absorbs green light better than red and penetrates sufficiently deeper into the skin as compared to blue light to probe the vasculature [9]. In 2010, Poh *et al.* used a blind source separation (BSS) technique to remove noises from physiological signals. By performing Independent Component Analysis (ICA) along with the reference signal, the independent component in close agreement with the cardiovascular pulse can be obtained. The feasibility to simultaneously measure multiple people in front of a camera was also presented in their study [10]. As an alternative, Lewandowska *et al.* chose to apply Principal Component Analysis (PCA) to extract the changeable component containing the heart rate. According to their work, PCA and ICA based methods have comparable results, while PCA reduces computational complexity. They further investigated the possibility to use only two, instead of three, color channels to extract heart rates so as to improve the computational efficiency [11]. In addition to these color-based methods, Balakrishnan *et al.* successfully extracted pulse rates by detecting subtle head motions caused by the Newtonian reaction to the pumping of blood at each heartbeat [12].

With these earlier studies, recent research has changed the focus on motion robustness improvement. BSS techniques – ICA and PCA – are based on the assumption that the pulse signal carries the strongest periodicity. This assumption is valid when no or slight subject motion is present, but may be invalid in a fitness scenario where subjects perform periodic exercises, thereby hindering the selection criteria. Moreover, a fairly long observation interval is required to perform an accurate selection, which prohibits adaption to fast motions. In 2013, de Haan *et al.* proposed the chrominance-based rPPG method (CHROM method) to reconstruct the pulse signal with a linear combination of three color channels under a standardized skin-tone assumption. This method eliminates the weak component selection heuristic used in BSS and performs very well with short interval. Consequently, the motion robustness was greatly improved [13]. More recently, de Haan *et al.* further improved their work by proposing a new rPPG method (PBV method) [14]. They derived the unique signature of the blood volume pulse (PBV-signature) from the color channels of an RGB camera and used it to reconstruct the pulse signal. As a further extension, the weighting vectors derived from the CHROM and PBV methods can be used to guide the component selection of the BSS methods. These results have shown superior accuracy and motion robustness compared to existing rPPG methods.

However, all of these methods are based on ambient light conditions. For subjects like premature babies, who need 24/7 continuous vital sign monitoring, it is preferable to have a method that can work even in darkness. This stimulated a new research area on infra-red (IR) based rPPG method, yet few breakthroughs have been made so far. As demonstrated in [15] and [16], the amplitude of the PPG signal in light reflected from the skin varies as a function of the wavelength. The peak in the visible light spectrum was found to be around $\lambda = 580$ nm. This is why the green channel carries the strongest pulsatility. But in the (near) IR spectrum ($\lambda = 700 - 1000$ nm), the PPG amplitude is very low

[15] and consequently contains a lot of noise [16]. Therefore, the existing rPPG methods may not be simply applied in IR spectrum.

Since the PBV method [14] has the highest accuracy among all of the existing rPPG methods, the goal of this paper is to study this method in detail and to investigate its application in the IR spectrum. BSS techniques, both PCA and ICA, are used to select a reliable PBV-signature. Based on the assumption that motion artifacts have equal influence over all color channels, a series of simulations are done to find the optimal combination of available filters, in an attempt to derive a PBV signature that has the maximum angle with the vector [1 1 1] (i.e., the motion vector).

Therefore, the main contributions of our work are: (1) measuring the PBV signatures for both the RGB camera and the monochrome cameras with different optical filters, (2) evaluating the accuracy of the PBV method compared to the reference sensor under stationary and motion induced scenarios, (3) performing simulations to find out the theoretically optimal optical filters for the monochrome cameras that lead to a higher degree of motion robustness, and (4) discussing the possible reasons for the discrepancy between simulated and measured results of the PBV signatures.

In Section II of this paper, we shall introduce the blood volume pulse signature (PBV signature) and five methods to obtain the signature from raw videos of stationary subjects. After that, we describe the PBV method and then show the experimental results of stationary subjects. In Section III we use the PBV signatures to evaluate the motion robustness in three scenarios. In section IV we show the theoretical optimal selection of IR filters from simulations, verify it by evaluations on stationary subjects, and discuss the discrepancy between simulation and measurement. Finally, we draw the conclusions in Section V.

II. Methods

In this section, we first give an introduction of the PBV signature proposed by de Haan *et al.*[14], followed by techniques to optimize the data, methods to compute the PBV signatures and the complete PBV method to extract the pulse signals. The evaluation metrics and the results shall be presented at the end of this section.

A. PBV-signature

Hülsbusch described that the *relative* PPG-amplitude ($RPPG(\lambda)$) is determined by the contrast between the blood and the blood-free tissue, as a function of the wavelength λ [15]. Assuming a melanin density of 3%, and given the absorption spectrum of the oxygenated arterial blood, dermis and epidermis, Hülsbusch's simulations led to the relative PPG-amplitude, the normalized version of which is shown in Figure 2a. While two peaks (λ = 540 nm and λ = 580 nm) can be clearly recognized in the visible light spectrum, the absorption level of oxygenated blood is very low in the IR spectrum (< 0.1 for λ > 600 nm).

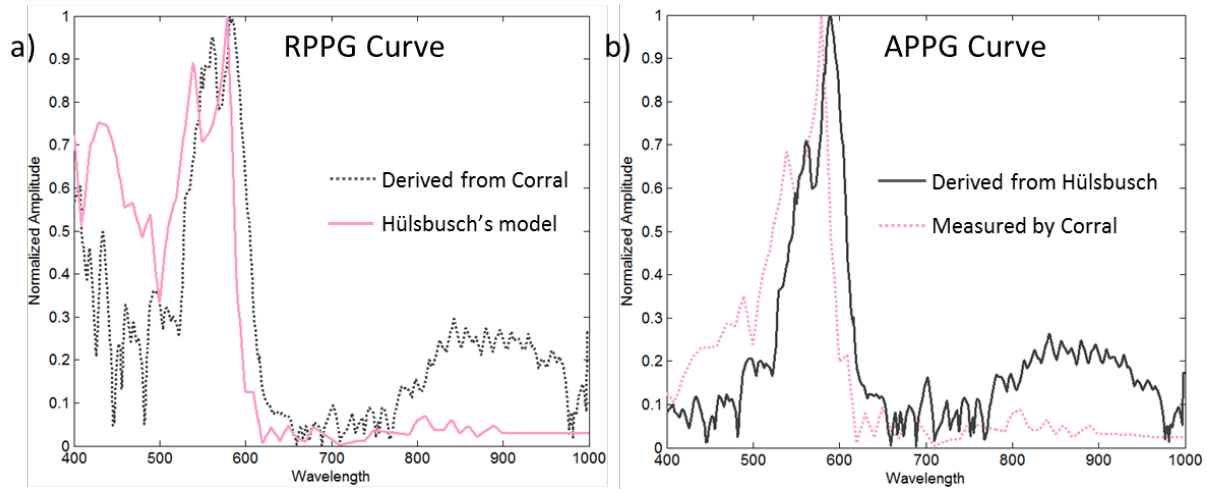

Figure 2. a) Relative PPG-amplitudes, simulated from Hülsbusch's model and derived from Corral's measured absolute PPG-amplitude. b) Absolute PPG-amplitudes, measured by Corral and derived from Hülsbusch's simulation. It shows a good match in the spectrum from 500 nm to 700 nm for both RPPG and APPG curves.

As an alternative, Corral *et al.* measured the PPG-amplitude by analyzing the light reflected from forehead illuminated by a quartz-halogen lamp which emits radiation in the visible and near infrared spectrum [16]. Since their measurement is not independent of illumination and skin types, de Haan *et al.* [14] referred to it as the *absolute* PPG-amplitude ($APPG(\lambda)$) to distinguish it from the *relative* PPG-amplitude from Hülsbusch's simulations. As can be seen from Figure 2b, Corral's normalized RPPG spectrum has a strong peak at around λ = 590 nm. De Haan *et al.* then proposed the equation expressing the relationship between *absolute* and *relative* PPG:

$$APPG(\lambda) = \rho_s(\lambda) I_h(\lambda) RPPG(\lambda), \quad (1)$$

where $\rho_s(\lambda)$ is the skin reflectance spectrum and $I_h(\lambda)$ is the halogen lamp spectrum. The derived APPG from Hülsbusch's model using this equation shows a good agreement with Corral's measured

APPG in the spectrum from λ = 500 nm to λ = 700 nm, as is shown in Figure 2b. The derived RPPG from Corral's measurement also matches Hülsbusch's model, as can be seen in Figure 2a.

De Haan *et al.* [14] defined the normalized blood volume pulse vector $\vec{P}_{bv}$ (PBV-vector) as the relative PPG-amplitude in the normalized color channels of the video camera registering a stationary skin region. They concluded that given the skin reflectance spectrum $\rho_s(\lambda)$, the illumination spectrum $I(\lambda)$ and, the responses of three color channels $H_{ch1}(\lambda)$, $H_{ch2}(\lambda)$ and $H_{ch3}(\lambda)$, the PBV-vector can be computed as follows:

$$\vec{P}_{bv} = \left[\frac{AC_{ch1}}{DC_{ch1}}, \frac{AC_{ch2}}{DC_{ch2}}, \frac{AC_{ch3}}{DC_{ch3}}\right] = \left[\frac{\int_{600}^{1000} H_{ch1}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}APPG(\lambda)\,d\lambda}{\int_{600}^{1000} H_{ch1}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}\rho_s(\lambda)\,d\lambda}, \frac{\int_{600}^{1000} H_{ch2}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}APPG(\lambda)\,d\lambda}{\int_{600}^{1000} H_{ch2}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}\rho_s(\lambda)\,d\lambda}, \frac{\int_{600}^{1000} H_{ch3}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}APPG(\lambda)\,d\lambda}{\int_{600}^{1000} H_{ch3}(\lambda)\frac{I(\lambda)}{I_h(\lambda)}\rho_s(\lambda)\,d\lambda}\right] \quad (2)$$

assuming a spectrum ranging from λ = 600 nm to λ = 1000 nm and where the noise-free RPPG curve from Hülsbusch's model is used to compute $APPG(\lambda)$ as shown in Eq. (1). Although the equation depends on the skin reflectance $\rho_s(\lambda)$, which varies with the melanin density, $\vec{\mathrm{P}}_{\mathrm{bv}}$ was found to be stable. This is because the $APPG(\lambda)$ is also related to skin reflectance, according to Eq. (1). As a consequence, the skin reflectance $\rho_s(\lambda)$ of the AC and DC component will likely compensate each other.

Since the stable $\vec{\mathrm{P}}_{\mathrm{bv}}$ represents the relative pulsatilities of the color channels, it can be used as the 'signature' of the variations of blood volume to distinguish the pulse signal from noises. This eliminates the assumption used by Blind Source Separation (BSS) algorithms that the pulse signal is the only significant periodic component in the video.

To confirm their assumption on the stable PBV signature, de Haan *et al.* measured $\vec{\mathrm{P}}_{\mathrm{bv}}$ on a large data set of 105 subjects with different skin types, which was built by Thomas in 2011 [17]. The experimental results showed a stable $\vec{\mathrm{P}}_{\mathrm{bv}}$ under white illumination [14]. The differences between the measured and predicted $\vec{\mathrm{P}}_{\mathrm{bv}}$ using Hülsbusch's simulations and Corral's spectrum were 4^o and 7^o, respectively.

However, this conclusion may not be applicable in the IR spectrum, because Hülsbusch's RPPG amplitude is too low and Corral's APPG spectrum contains a lot of noise, as can be seen in Figure 2. To find out whether it is indeed impossible to use the PBV method in the IR spectrum, we decided to build a new dataset of recordings in the IR spectrum and then verify the PBV method accordingly. Before presenting the methods used to extract the PBV signature, we first demonstrate how to construct the optimized signals from raw videos in the next section.

B. Construct the optimized signals from raw videos

Figure 3 demonstrates the overview of the approach to reconstruct the optimized color signals from raw videos, which will be used to compute the normalized blood volume pulse vector $\vec{P}_{bv}$.

a) Track ROI

We first select the face of the subject as the region of interest (ROI) – the small rectangle with red borders shown in Figure 3a – because forehead is found to be one of the skin regions that contains

the strongest pulsatilities and it was widely used by previous rPPG methods [9]-[14]. If we assume the subject is stationary, we can simply set the same ROI for every frame. To minimize the motion artifacts, we used OpenCV 2.4 [18] to track the ROI for each frame. Given the initial ROI, we can track it accurately using the available global and local motion compensation techniques: the Circulant Structure of Tracking-by-detection with Kernels (CSK) developed by Henriques *et al.* [19] and the Farnebäck dense optical flow algorithm [20], respectively. The skin pixel classification [23] enables us to distinguish skin pixels from non-skin pixels. As can be seen from Figure 3a, the non-skin pixels are set to black and removed for further processing. The complete tracking algorithm had been implemented by Wenjin *et al.* [23] earlier this year and we simply used it for the experiments.

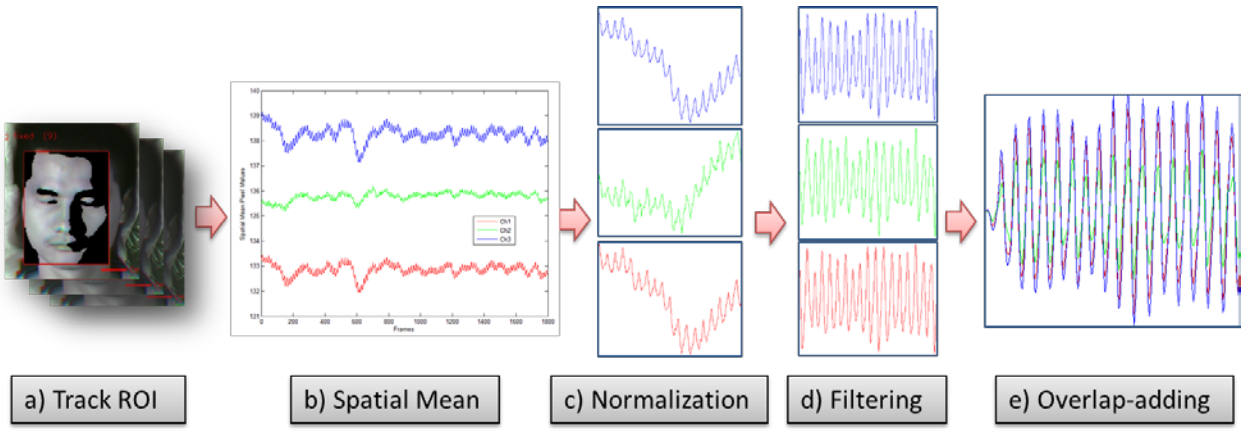

Figure 3. Overview of the method to reconstruct the optimized color signals for later $\vec{P}_{bv}$ extraction. a) ROI tracking by frames using global and local motion compensation techniques from OpenCV. b) Take spatial mean within the ROI for each of the three channels. c) Normalize the signals to remove DC component. d) Filter out the distortions not in the range of [40, 240] bpm. e) Using overlap-adding method to further optimize the normalized color signals.

b) Spatial Mean

For each video frame, we calculate the mean skin-pixel value within the ROI for each of the three color channels and store the results to a 3 × N matrix C_{raw}, where each row $\vec{C}_{raw1}$, $\vec{C}_{raw2}$ and $\vec{C}_{raw3}$ contains N frames of spatial mean pixel values.

c) Normalization

Since the DC component contains no pulsatilities, we normalize the raw signals C_{raw} to remove it and consequently obtain the 3 × N DC-free color matrix C_n, each row of which contains N frames of the normalized color signals $\vec{C}_{n1}$, $\vec{C}_{n2}$ and $\vec{C}_{n3}$, respectively:

$$\vec{C}_{ni} = \frac{1}{\mu(\vec{C}_{rawi})}\vec{C}_{rawi} - 1, \quad with\ i = 1,2,3. \tag{3}$$

where the operator μ corresponds to temporal mean over N frames.

d) Filtering

Given the fact that the heart rate of a normal healthy subject ranges from approximately 40 to 240 beats per minute (bpm), corresponding to a frequency band from 0.6 to 4 Hz, we can simply apply a

bandpass filter to remove the distortions out of this frequency range. Consequently, the filtered signals clearly show the AC component of the raw signals, as can be seen from Figure 3d.

e) Overlap-adding

To further improve the resulting signals for later processing, we use the overlap-adding method, as the one used by de Haan *et al.* in [13]. By normalizing and filtering the signals over a short interval instead of the complete length of the video, the method is more adaptive to quick changes, and thereby more motion robust. The overview of the overlap-add approach is shown in Figure 4, assuming the overlapping size is 50% of the interval in order to better illustrate the method. In practice, we start processing the 3 × N matrix C_{raw} within the first interval (Frame 1 to 64), e.g., normalization and filtering. After that, the filtered signal is multiplied with a Hanning Window to obtain the optimized signal. This process is repeated with the interval sliding along the time axis in a step of 1 frame until it reaches the end of the signal, which means 63 out of 64 frames of any two adjacent intervals that are overlapping each other. Finally, we add all of these separate, partially overlapping, signals together to construct the final optimized 3 × N matrix C_o, as can be seen from Figure 3e.

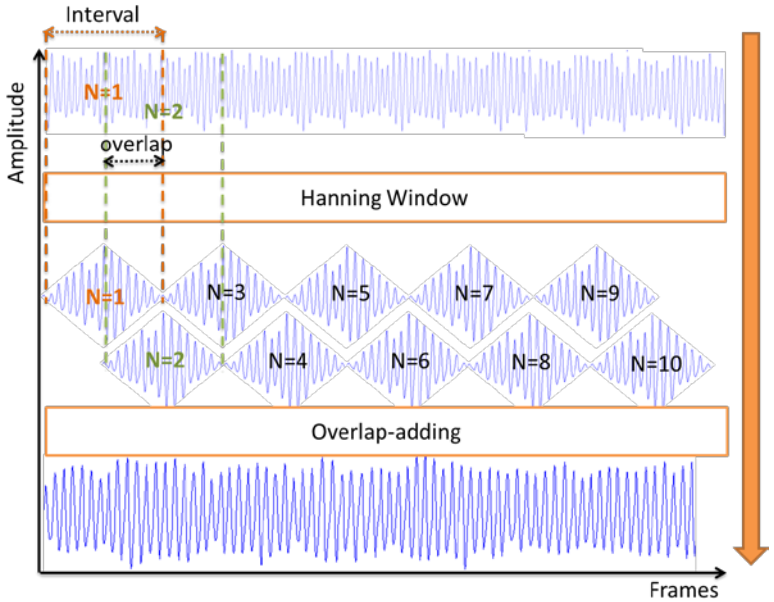

Figure 4. Demonstration of the overlap-adding method. During each step, the segmented signal is normalized and filtered and then multiplied with a Hanning window. Then the interval window moves along to the next segment. The bottom shows the final optimized output signal as the sum of all of these overlapping segments.

C. Methods to compute the PBV signature

Given the optimized color signals, we can extract the PBV signature using five different methods proposed by Mark van Gastel and Gerard de Haan: the simple correlation method, two different BSS-PCA based methods and two different BSS-ICA based methods. Some modifications have been made to improve the component selection of the BSS-ICA based methods, and we will review all these methods as following.

a) Simple correlation method

In Figure 3e, we intentionally put three optimized color signals $\vec{C}_{o1}$, $\vec{C}_{o2}$ and $\vec{C}_{o3}$ in one figure so as to have an overview of their amplitudes. Since a larger amplitude indicates a stronger pulsatility, we can simply choose the color channel with the maximum amplitude as the pulse signal $\vec{S}_{pulse}$:

$$\vec{S}_{pulse} = \max(\vec{C}_{oi}), with\ i = 1,2,3. \tag{4}$$

To improve motion robustness, we can redefine the pulse signal as:

$$\vec{S}_{pulse} = \max(\vec{C}_{oi}) - \min(\vec{C}_{oi}), with\ i = 1,2,3. \tag{5}$$

This is because we assume that motions have equal influence over all color channels. By subtracting the channel with minimum amplitude from the one with maximum amplitude, the motion artifacts can be greatly removed. Since the PBV vector $\vec{\mathrm{P}}_{\mathrm{bv}}$ contains the relative pulsatilities of the video, the correlation between the pulse signal $\vec{S}_{pulse}$ and the N × 3 optimized color signals matrix C_o^T shall be equal to $\vec{\mathrm{P}}_{\mathrm{bv}}$:

$$\vec{S}_{pulse} C_o^T = k\vec{\mathrm{P}}_{\mathrm{bv}}, \tag{6}$$

with scalar factor k chosen such that the $\vec{\mathrm{P}}_{\mathrm{bv}}$ is normalized, and we define it as $\mathrm{Corr\_}\vec{\mathrm{P}}_{\mathrm{bv}}$. The limitation of this method is that we assume that the pulse signal $\vec{S}_{pulse}$ to be constructed using Eq. (4) or (5) is the true pulse signal. To relax this assumption, we proposed other methods using BSS algorithms for the selection of the pulse signal.

b) Blind Source Separation (BSS)

Blind source separation (BSS) is the separation of a set of source signals from a set of mixed signals, without any information concerning the source signals or the mixing process. Therefore, the BSS-based rPPG methods require no assumptions regarding the main distortions, skin color, or blood volume pulse vectors. The source signals are separated by BSS assuming that the pulse is the only periodic signal. This assumption limits its usage in scenarios that contain periodic motions. But since the PBV signature is independent of motion, our aim is to compute the most reliable $\vec{\mathrm{P}}_{\mathrm{bv}}$ from our data set of stationary subjects and then use the signature for any kinds of scenarios. In such a scenario, the BSS algorithms shall yield the best results.

Principle Component Analysis (PCA) is a BSS technique that uses an orthogonal transformation to convert a set of mixed signals into a set of linearly uncorrelated signals called principal components. This transformation ensures that the first principal component accounts for as much of the variability in the mixed signals as possible, and each of the subsequent components in turn has the largest possible variance under the constraint that it is orthogonal to the preceding components. We thereby can select the true pulse signal and calculate the $\vec{\mathrm{P}}_{\mathrm{bv}}$ by performing the function ***pca***() in MATLAB:

$$[coeff, score, \sim, \sim, explained, \sim] = pca(C_o^T), \tag{7}$$

where C_o^T is the normalized signal optimized using the overlap-adding method (with an interval window size of 256 frames and a step size of 1 frame). Regarding the output variables of the PCA function, each column of the 3 × 3 matrix ***coeff*** contains coefficients for one principal component in descending order of component variance, columns of the ***score*** correspond to the principal components, and ***explained*** corresponds to the percentage of the total variance explained by each principal component.

When the signal C_o^T is clean enough, which should be expected after normalization and optimization, the first column of ***coeff*** equals the PBV-vector of each sliding overlapping window and the corresponding column of ***score*** contains the pulse signal. We can thereby reconstruct the complete pulse signal $PCA\_\vec{\mathrm{S}}_{\mathrm{pulse}}$ by overlap-adding the ***score*** values. Since we do not need to construct a pulse signal as Eq. (4) or (5), the results are expected to be more accurate and reliable, especially for stationary subjects. Given the pulse signal $PCA\_\vec{\mathrm{S}}_{\mathrm{pulse}}$, we can obtain the PBV vector $PCA\_\vec{\mathrm{P}}_{\mathrm{bv}}$ using Eq. (6) by calculating the correlation between $PCA\_\vec{\mathrm{S}}_{\mathrm{pulse}}$ and the normalized color signals C_o^T. We refer to this method as *Correlation with PCA*.

The problem with the correlation methods is that the correlation is calculated along the complete length of the videos, which implies that even a small distortion like eye blinking or facial motion will be inevitably reflected in the results. We need to find a way to minimize such distortions. Since the ***explained*** values serve as an indicator of the accuracy of component selection, as an alternative, we sort the first column of ***coeff*** (the PBV vectors obtained from each sliding overlapping window) accordingly. Given the sorted results, we simply take the median value within the top 5% as the PBV vector $PCA\_sorted\_\vec{\mathrm{P}}_{\mathrm{bv}}$ and the corresponding ***explained*** value indicates how accurate the component selection is. We refer to this method as *Median of sorted PCA*.

Another widely used BSS technique is Independent Component Analysis (ICA), which attempts to decompose a set of mixed signals into a set of independent non-Gaussian signals by maximizing the statistical independence of the estimated components. Here we choose the joint approximate Diagonalization of Eigen-matrices (JADE) algorithm developed by Cardoso to perform ICA [22]:

$$[A, S] = jade(C_o), \tag{8}$$

where C_o is the 3 × N normalized signal optimized using the overlap-adding method as shown in section B. Regarding the output variables, ***A*** is a 3 × 3 matrix containing the mixture coefficients, and ***S*** is a 3 × N matrix corresponding to the estimated source signals in such a way that:

$$AS = C_o\ . \tag{9}$$

Unlike PCA, there is no fixed ordering of the ICA components. Hence, during each iteration of the overlap-adding process, we need to identify the component corresponding to the pulse signal. To this end, we construct a guidance signal using Eq. (4). Then we select one row of ***S*** as the pulse signal in such a way that the correlation between this row and the guidance signal is maximal. Consequently, the corresponding column of ***A*** (take the absolute values) is equal to the PBV-vector of the current sliding overlapping window. Similar to the PCA methods, we can reconstruct the complete pulse signal $ICA\_\vec{\mathrm{S}}_{\mathrm{pulse}}$ by overlap-adding the selected rows of ***S***, and we then obtain the

PBV-vector $ICA\_\vec{\mathrm{P}}_{\mathrm{bv}}$ using Eq. (6) by calculating the correlation between $ICA\_\vec{\mathrm{S}}_{\mathrm{pulse}}$ and the normalized color signals C_o^T. We refer to this method as *Correlation with ICA*.

To minimize the distortions along the complete length of the videos, we sort the selected columns of ***A*** (the PBV-vector obtained from each sliding overlapping window) by the corresponding correlation values. Given the sorted results, we take the median value within the top 5% as the PBV-vector $ICA\_sorted\_\vec{\mathrm{P}}_{\mathrm{bv}}$. We refer to this method as *Median of sorted ICA*.

D. Experimental results of PBV signatures:

For the experiments, we recorded the videos in a lab using two light sources with incandescent light bulbs as illumination. This setup is shown in Figure 5a. We obtained the infrared videos by filtering out the visible light wavelengths with optical filters. To this end, we put a high-pass filter with a cut-off wavelength of 650 nm inside an RGB color camera (AVT Marlin F-046C). As an alternative, we used three identical monochrome cameras (AVT Stingray F-046B), each of which contains a different IR band-pass optical filter. By changing the filters inside the monochrome cameras, we expect to obtain a different PBV-vector, which gives us more flexibility in the future to improve the motion robustness. Since both Hülsbusch's RPPG-amplitude and Corral's APPG-amplitude curves are not quite reliable in the IR spectrum, as a starting point, we chose one filter centering at 675 nm and two filters centering further into the infrared spectrum (800 nm and 842 nm). Therefore, we used four cameras simultaneously to record two raw videos (640 × 480 pixels, 15 frames per second), each of which contains three channels. The placement of the four cameras is shown in Figure 5b.

Figure 5. a) Overview of the experiment environment. Two incandescent lamps are used as illumination. b) The four cameras: Cam1-3 are three identical monochrome cameras AVT Stingray F-046B and Cam4 is the RGB color camera AVT Marlin F-046C. c) C+K Multi Display Device (MDD4) and the probe for melanin measurement. d) Philips IntelliVue X2 patient monitor for reference heart rates measurement.

In total, 41 healthy volunteers with different skin types, male and female, age between 20 and 32, participated in our research study. Informed consent was obtained from each subject, and the study was approved by the Internal Committee Biomedical Experiments of Philips Research. We quantified the skin types by measuring the melanin concentration on the anterior surface of the subjects' forearms using the Multi Display Device (MDD4) from Courage and Khazaka. Before starting the recordings, subjects were asked to sit and relax for around 3 minutes to ensure a stable pulse rate. The duration of all recordings was set to 2 minutes, during which time subjects were sitting still on the dentist chair shown in Figure 5a. To obtain the synchronized reference pulse signals during recordings, subjects were asked to attach a pulse-oximeter finger clip from the Philips IntelliVue X2 patient monitor.

We extracted the PBV vectors from these raw videos using the above-mentioned methods, the results of which are shown in Figure 6. In general, the color camera gives a more consistent PBV-vector than the monochrome cameras do. This is probably because the relative PPG-amplitude curve has a dip at around λ = 660 nm, where the noise is even stronger than the signal [16]. When we use a band-pass filter near this dip, like the one with center wavelength of 675 nm, the results will be harmed by noise. Since the PBV-vector $\vec{P}_{bv}$ is normalized, the inaccurate component of this channel will affect the other two, thereby causing the distortions of the $\vec{P}_{bv}$ computed from the monochrome cameras. For the color camera, the distortions are relatively small, because the long-pass filter transmits wavelength from 650 nm to 1000 nm, along which spectrum the PPG signal is generally much stronger than the noise. The extracted $\vec{P}_{bv}$ is thereby more consistent.

	675 nm	800 nm	842 nm	Sum
Mono_Pbv_Corr	0.032	0.048	0.031	0.111
Mono_Pbv_PCA	0.010	0.031	0.015	0.055
Mono_Pbv_PCA_sorted_median	0.009	0.034	0.016	0.058
Mono_Pbv_ICA	0.033	0.040	0.025	0.098
Mono_Pbv_ICA_sorted_median	0.030	0.035	0.025	0.090

(a)

	Red	Green	Blue	Sum
RGB_Pbv_Corr	0.036	0.034	0.028	0.098
RGB_Pbv_PCA	0.013	0.013	0.011	0.037
RGB_Pbv_PCA_sorted_median	0.014	0.012	0.010	0.036
RGB_Pbv_ICA	0.015	0.015	0.013	0.042
RGB_Pbv_ICA_sorted_median	0.016	0.013	0.011	0.040

(b)

Table 1. Standard deviation of PBV-vectors measured with different methods for monochrome cameras (a) and the color camera (b). The sum of standard deviation of all three channels is shown in the last column.

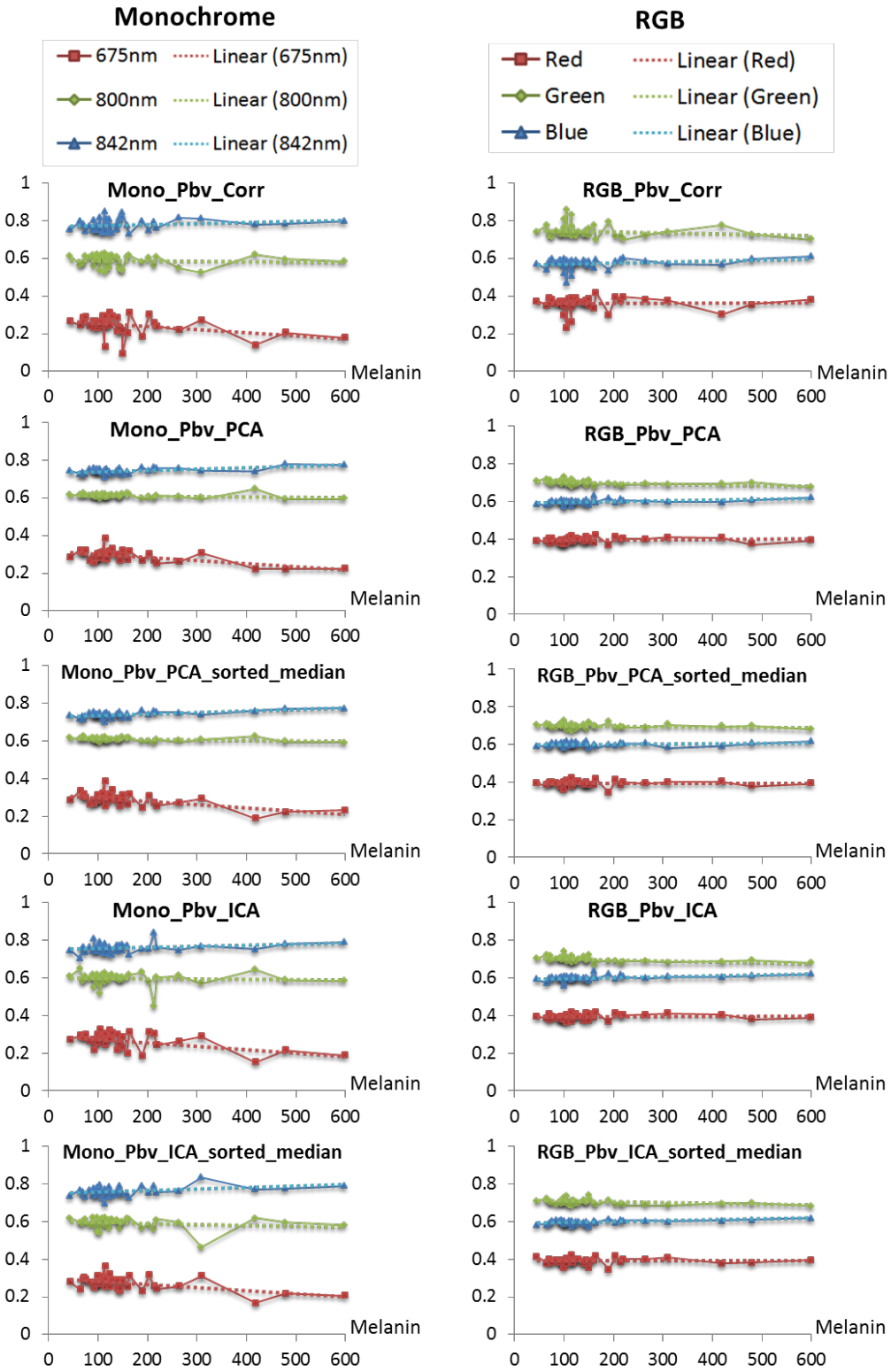

Figure 6. The normalized blood volume pulse vectors extracted from monochrome cameras AVT Stingray F-046B (left column) and the RGB color camera AVT Marlin F-046C (right column), plotted by measured melanin values. The bandpass optical filters used in the monochrome cameras have center wavelengths of 675 nm, 800 nm and 842 nm, respectively, while the long-pass filter used in color camera has a cut-off wavelength of 650 nm. The methods used, from top to bottom, are Simple Correlation, Correlation with PCA, Median of sorted PCA, Correlation with ICA, and Median of sorted ICA.

If we rank these five methods by the sum of standard deviation of all three channels, as shown in Table 1, we conclude that the PCA based methods are the best while the simple correlation method is the worst. This conclusion is consistent with our expectations: 1) BSS algorithms perform better for stationary subjects. 2) Correlation related methods suffer from distortions along the complete videos. 3) PCA based methods yield the cleanest results because we do not need to artificially construct a pulse signal (simple correlation method) or a guidance signal (ICA).

The overview of the normalized mean PBV vectors obtained from our data set of 41 subjects using five different methods are shown in Table 2a for monochrome cameras and Table 2b shows the results for the color camera. In conclusion, the measured results have confirmed our assumption that the PBV signature is stable both in the visible light spectrum and in the infrared spectrum. Among all five methods to compute PBV vectors, the *Median of sorted PCA* method is found to be the most reliable one. The corresponding PBV signatures for the monochrome cameras and the RGB camera are [0.29 0.61 0.74] and [0.39 0.70 0.60], respectively.

	675 nm	800 nm	842 nm	Angle
Mono_Pbv_Corr	0.24	0.58	0.77	22.40
Mono_Pbv_PCA	0.29	0.61	0.74	19.17
Mono_Pbv_PCA_sorted_median	0.29	0.61	0.74	19.34
Mono_Pbv_ICA	0.27	0.60	0.76	20.83
Mono_Pbv_ICA_sorted_median	0.27	0.59	0.76	20.47

(a)

	Red	Green	Blue	Angle
RGB_Pbv_Corr	0.36	0.74	0.57	15.60
RGB_Pbv_PCA	0.39	0.70	0.60	12.70
RGB_Pbv_PCA_sorted_median	0.39	0.70	0.60	12.76
RGB_Pbv_ICA	0.39	0.70	0.60	12.70
RGB_Pbv_ICA_sorted_median	0.39	0.70	0.60	13.01

(b)

Table 2. Normalized mean PBV-vectors measured with different methods for monochrome cameras (a) and the color camera (b). Angle corresponds to the angle with the motion vector [1 1 1] and Angle_off corresponds to the angle with the simulated results.

E. PBV method to extract pulse signals

Just as many other rPPG methods, we assume that the pulse signal can be constructed by a linear combination of normalized color channels:

$$\vec{S}_{pulse} = \vec{W}_{pbv} C_o \,, \tag{10}$$

where the weighting $\vec{W}_{pbv}$ has unit length such that $\vec{W}_{pbv}\vec{W}_{pbv}^{T} = 1$ and the 3 × N matrix C_o are the segmented optimized signals, acquired in the same way as described in section B (the segment size is set to 64 to be more adaptive to motion). Using Eq. (6) to rewrite Eq. (10), we will get:

$$\overrightarrow{W}_{pbv} C_o C_o^T = k\overrightarrow{\mathrm{P}}_{\mathrm{sig}} , \tag{11}$$

where $\overrightarrow{\mathrm{P}}_{\mathrm{sig}}$ corresponds to the known PBV signature. The weighting $\overrightarrow{W}_{pbv}$ can thus be calculated as:

$$\overrightarrow{W}_{pbv} = k\overrightarrow{\mathrm{P}}_{\mathrm{sig}} Q^{-1} \quad with \quad Q = C_o C_o^T . \tag{12}$$

Thereby, the segmented pulse signal can be reconstructed using Eq. (10):

$$\vec{S}_{pulse} = \overrightarrow{W}_{pbv} C_o = k\overrightarrow{\mathrm{P}}_{\mathrm{sig}} (C_o C_o^T)^{-1} C_o = k\overrightarrow{\mathrm{P}}_{\mathrm{sig}} (C_o^T)^{-1}. \tag{13}$$

At the end, all segmented pulse signals are multiplied with a Hanning Window and overlap-added to construct the complete pulse signal. According to these equations, given the known PBV signatures, we shall be able to extract the pulse signals from any video. This is the PBV method proposed by de Haan *et al.* in [14].

F. Evaluation metrics:

For evaluation, we use the signal-to-noise-ratio (SNR) as one of the metrics. Essentially, SNR is computed as the ratio of the energy around the fundamental frequency plus its first harmonic of the pulse signal and the remaining energy in the spectrum:

$$SNR = 10 \log_{10} \left(\frac{\sum_{f=36}^{210} \left(U_t(f) \hat{S}_{pulse}(f) \right)^2}{\sum_{f=36}^{210} \left(\left(1 - U_t(f) \right) \hat{S}_{pulse}(f) \right)^2} \right), \tag{14}$$

where $\hat{S}_{pulse}(f)$ is the frequency spectrum of the extracted pulse signal $\vec{S}_{pulse}$, f is the frequency in beats per minute, and $U_t(f)$ is a binary template window shown in Figure 7. It is worth noting that the fundamental frequency is measured every frame by detecting the peak in the frequency domain of the reference signal using a sliding Fourier window of 256 frames. The mean value *SNRa* over all of the partially overlapping segments will then be used as the evaluation metric.

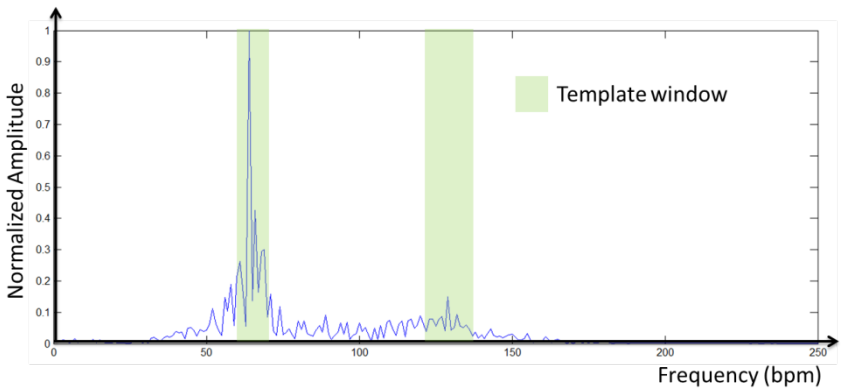

Figure 7. In the frequency spectrum of the extracted pulse signal, the binary template window is defined as ± 3 bpm away from the reference pulse rates measured by Philips IntelliVue X2 patient monitor. Another template window is set as ± 6 bpm from the first harmonic, e.g., twice of the reference pulse rates. The SNR is then calculated as the energy ratio of the spectral segments inside and outside the template windows.

As an alternative, we first calculate the mean absolute percentage error (MAPE) of the PBV method:

$$MAPE = \mu\left(\left|\frac{\vec{S}_{ref}-\vec{S}_{pulse}}{\vec{S}_{ref}}\right|\right) \times 100\% \ , \tag{15}$$

where μ corresponds to the mean value over all of the partially overlapping segments, $\vec{S}_{pulse}$ is the pulse signal extracted using PBV method and $\vec{S}_{ref}$ is the reference pulse signal measured by Philips IntelliVue X2 patient monitor. To be more straightforward, we thus define the second evaluation metric ACCU (accuracy) as:

$$ACCU = 100\% - MAPE \ , \tag{16}$$

The last metric we shall use for evaluation is the Bland–Altman plot, a widely used method in medical statistics to analyze the agreement between two different approaches. For each subject i, we have two sets of pulse rates, one calculated using the PBV method and one measured using the reference device. The temporal mean values of these two datasets are then defined as HB_Cal_i and HB_Ref_i, respectively. Hence, to compare the PBV method with the reference contact sensor, we can plot the data in the Cartesian plane with coordinates as:

$$S(x_i,\ y_i) = \left(\frac{HB\_Cal_i + HB\_Ref_i}{2},\ HB\_Cal_i - HB\_Ref_i\right) \ . \tag{17}$$

It is common to compute 95% limits of agreements for each comparison, which is the average difference ± 1.96σ of the difference, where σ corresponds to the standard deviation. If the differences are not significant and few outliers exist, we can conclude our method is in a good agreement with the reference.

Based on the temporal mean values HB_Cal_i and HB_Ref_i, we can also plot the Pearson Correlation graph to visually demonstrate the correlation between the PBV method and the reference.

G. Evaluation results

Assuming that the normalized mean PBV-vectors measured with the Median of sorted PCA method are reliable, we used the PBV signatures ([0.29 0.61 0.74] for the monochrome cameras and [0.39 0.70 0.60] for the color camera) to extract the pulse signals from raw videos. To minimize the motion artifacts and the influence of non-skin pixels, we implemented the PBV method in OpenCV with global and local tracking algorithms [23]. Another advantage of using OpenCV is that it provides a graphical interface to compare the PBV method with the reference frame by frame, as shown in Figure 8a. Overlap-adding and Hanning Window techniques (window size is set to 64 frames) are used to construct both pulse signals, and the heart rates (bpm) are then obtained by simple peak detection over the respective frequency spectrum. Figure 8b shows an example of the results.

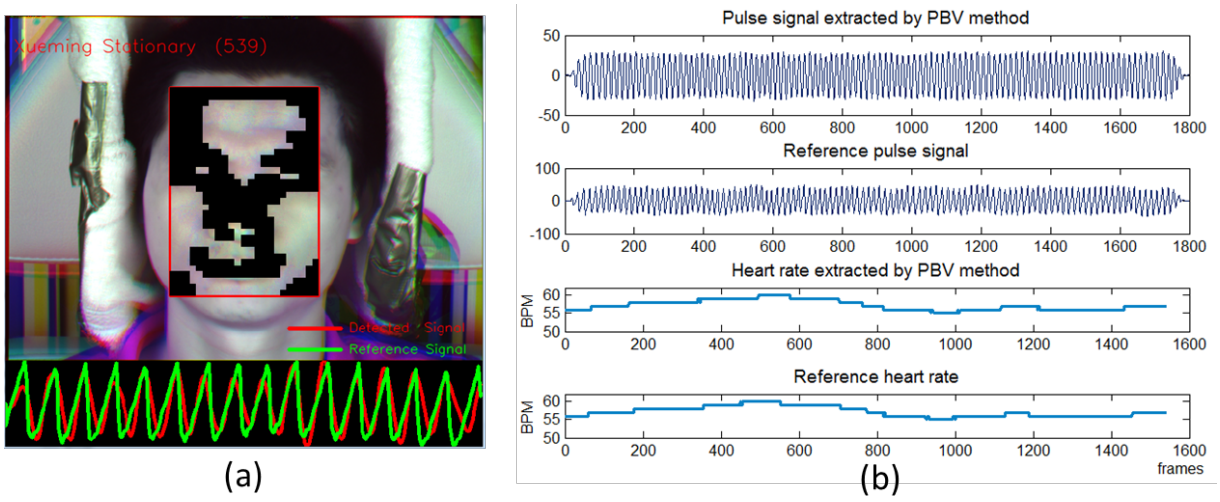

Figure 8. a) Snapshot of the real time pulse signal extraction using OpenCV. The red rectangle on the face specifies the region of interest (ROI), the black pixels within which are discarded by skin classification algorithm. At the bottom, the green signal corresponds to the reference pulse signal, while the red signal corresponds to the pulse signal extracted by the PBV method. b) Overview of the resulting pulse signals and heart rates, both from the PBV method and from the reference.

Figure 9 shows the overall results of all the three evaluation metrics for both monochrome cameras and the color camera. It can be clearly recognized that the PBV method shows great agreements with the reference and the results are slightly better for monochrome cameras than for the RGB camera. The SNRs for the monochrome cameras ranges from 2.5 dB to 9.3 dB and the average SNR over all 41 subjects is around 6.0 dB. For the RGB camera, 7 subjects have a SNR smaller than 3 dB and the average SNR is 4.7 dB. In terms of the ACCU metric, the smallest percentage of accuracy is around 96% and the average ACCU is around 99%, indicating very high precision of the PBV method. Another observation from these two metrics is that the results are skin-type independent because they do not show a trend by melanin levels. The Bland-Altman plots show that the biggest difference from reference heart beats is just around 2 bpm, which further proves the accuracy of the PBV method and the reliability of the PBV signatures. The Pearson Correlation plots in Figure 10 show a perfect correlation between the PBV method and the reference, where x-axis and y-axis correspond to the temporal mean heart rates from reference and from the PBV method, respectively. We can now safely conclude that for stationary subjects, the PBV method performs very well in the infrared spectrum.

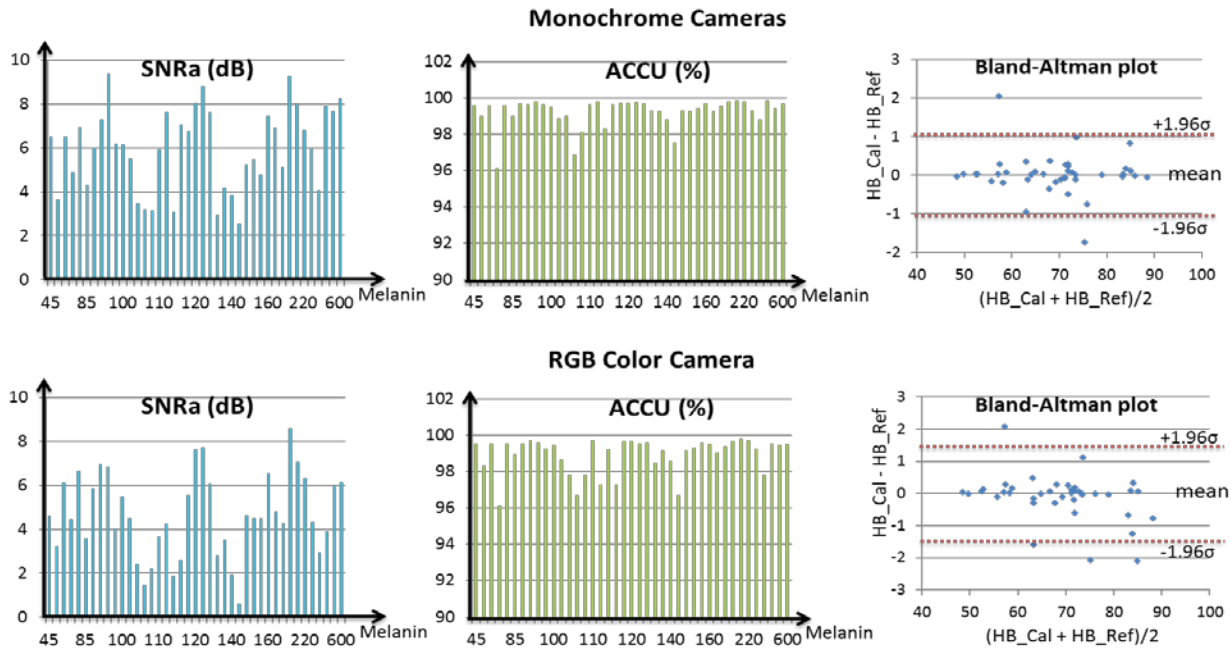

Figure 9. The evaluation results for monochrome cameras (upper row) and the RGB color camera (bottom row) with difference metrics: from left to right, signal-to-noise-ratio (SNRa), accuracy (ACCU) and Bland-Altman plot. All three metrics show a great agreement between our method and the reference.

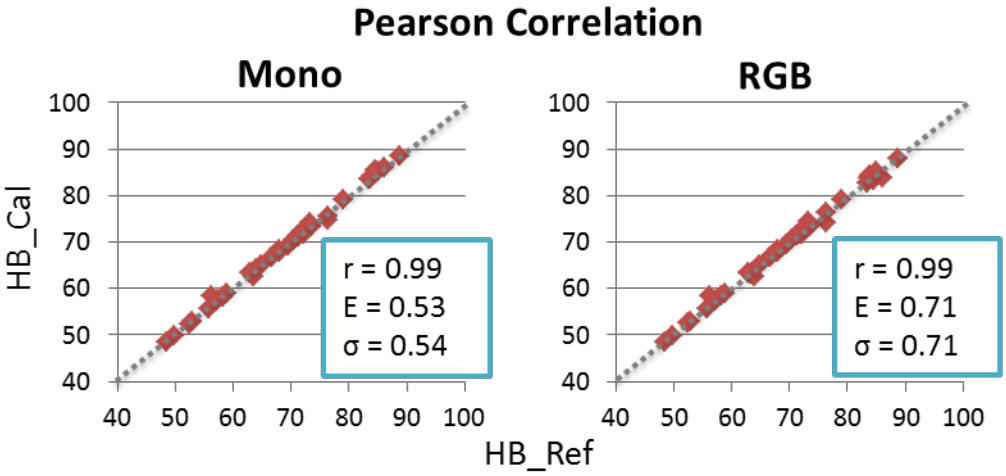

Figure 10. Pearson Correlation plots for monochrome cameras (left) and RGB camera (right) where r corresponds to Pearson correlation, E corresponds to root-mean-square error (RMSE) and σ corresponds to the standard deviation. This figure shows a perfect correlation between our method and the reference.

III. Motion Robustness

In this chapter we test the motion robustness of the PBV method. To this end, we recorded videos of 13 subjects with three types of motions: rotation, scaling and translation. Specifically, rotation means the subject rotates his/her head horizontally and vertically in a random manner. Scaling means the subject moves his/her face closer or further to the cameras. Translation means the subject moves his/her head leftward and rightward while keeping the front face towards the cameras.

The experimental setup is identical to the setup described in Section II. The duration of all recordings was set to 2 minutes 20 seconds. Subjects were asked to keep stationary for the first 20 seconds and then do the head movement for 2 minutes. Each subject therefore had 6 videos of all three types of motions: 3 for the monochrome cameras and 3 for the color camera. Again, we used the same PBV signatures ([0.29 0.61 0.74] for the monochrome cameras and [0.39 0.70 0.60] for the color camera) to extract the pulse signals.

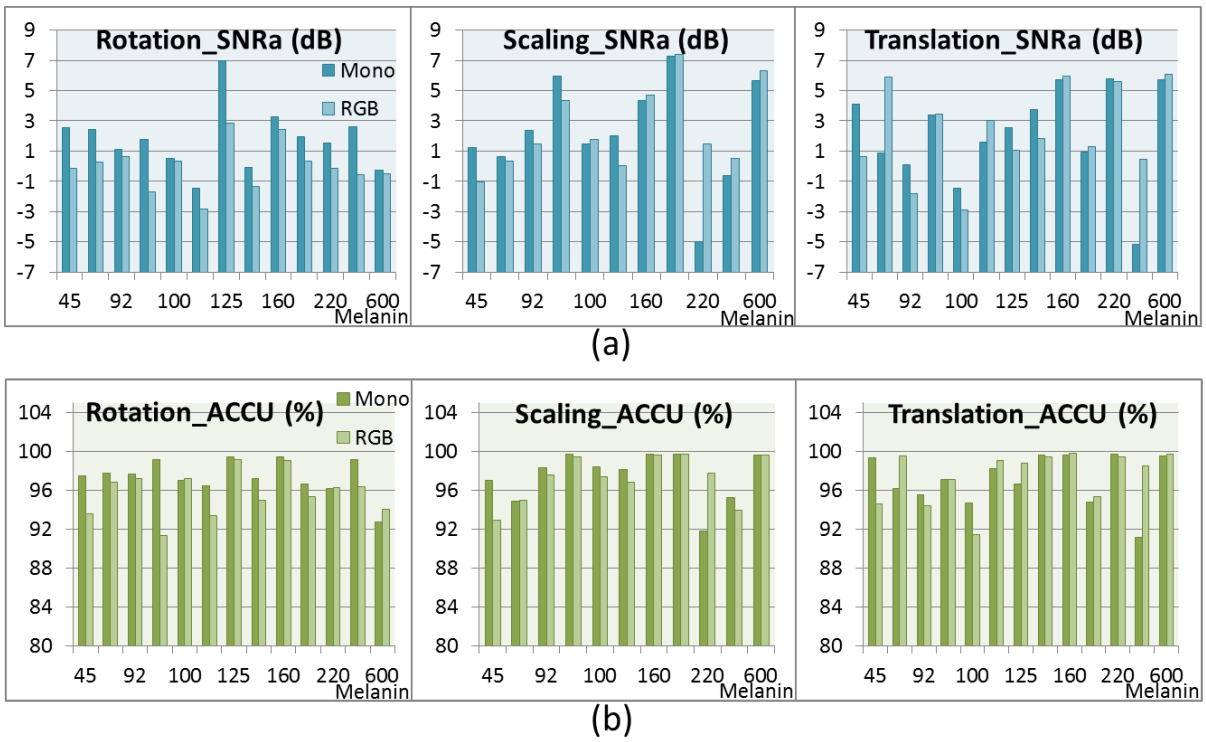

Figure 11. The evaluation results for monochrome cameras and the RGB color camera with three types of motions: from left to right, rotation, scaling and translation. a) Signal-to-noise-ration (SNRa). b) Accuracy (ACCU). The results from monochrome cameras are slightly better than the ones from RGB camera. Under all three motion scenarios, the PBV method achieves a very high accuracy (minimum 91%) compared to the reference.

The overview of the results is shown in Figure 11, where two videos of scaling motion were removed because there were some problems with the reference signals. The average SNRa for rotation, scaling and translation is 1.8 dB, 2.1dB and 2.2 dB with the monochrome cameras, and 0.0 dB, 2.1 dB and 2.3 dB with the RGB camera, respectively. The reason for the comparatively lower performance of videos with rotation motion is that when the subject is rotating the head, some parts of the face will disappear. This implies the tracking ROI is not identical from frame to frame, thereby causing the distortions. For the videos with scaling and translation motions, the subject is always facing towards the camera, in which case the distortions caused by wrong tracking will be minimized. Even with a couple of outliers, the ACCU metric shows that the PBV method achieves the average accuracy of around 97% for all three motions. Although the accuracy is a bit lower compared to the stationary videos, it is considered sufficiently precise. And since the results show no trend by the melanin levels, the performance is found to be skin-type independent.

In theory, since the PBV signature for monochrome cameras has a bigger angle with the motion vector [1 1 1], we expect the monochrome cameras to have a better performance than the RGB camera in videos with motions. This can be somewhat confirmed by Figure 11, although the performances are quite similar. Actually, in practice, since we use three monochrome cameras to record videos, the output videos suffer from image misalignment. Therefore, we need to align the three channels before processing the videos. However, the image alignment method [24] cannot be perfect, particularly when the subject is moving. The resulting misalignment may degrade the performance of our method.

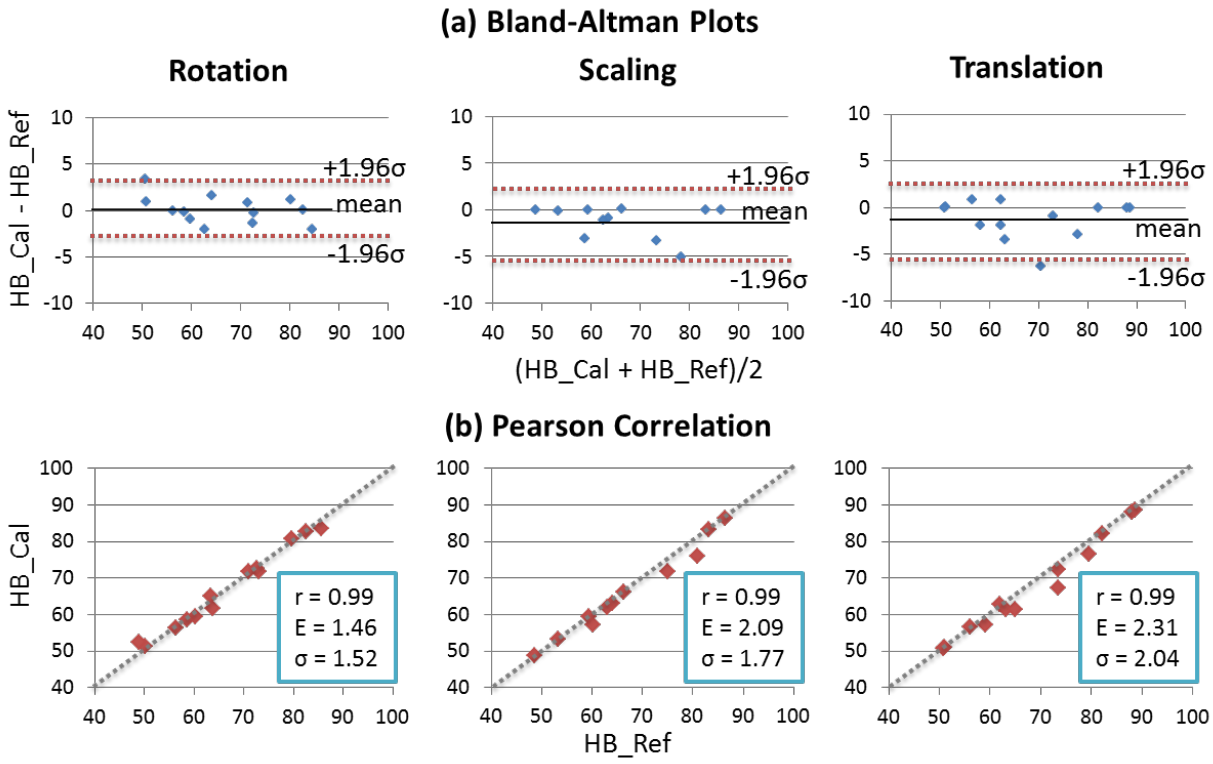

Figure 12. Correlation plots for monochrome cameras with three types of motions: from left to right, rotation, scaling and translation. a) Bland-Altman plots. b) Pearson Correlation plots, where r corresponds to Pearson correlation, E corresponds to root-mean-square error (RMSE) and σ corresponds to the standard deviation, same as in Bland-Altman plots. This figure shows our method does not significantly vary from the reference.

For videos with motions, it is worth noting that the selection of the region of interest (ROI) is an important factor because it can affect the tracking algorithms, thereby changing the results significantly. There is no golden rule for the selection, but from our experience, the ROI should be around half of the head size and should cover the forehead. Moreover, since our videos are captured with a frame rate of 15 frames per second, too quick motions should be avoided.

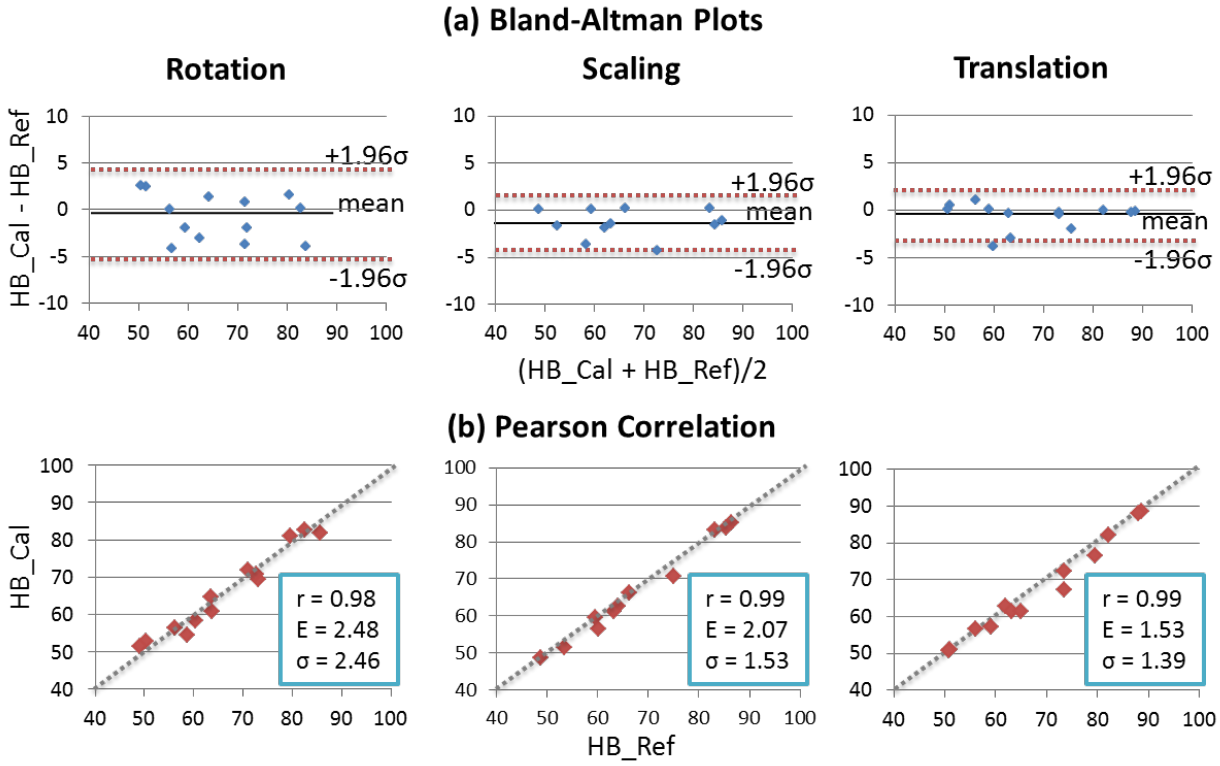

Figure 13. Correlation plots for RGB camera with three types of motions: from left to right, rotation, scaling and translation. a) Bland-Altman plots. b) Pearson Correlation plots, where r corresponds to Pearson correlation, E corresponds to root-mean-square error (RMSE) and σ corresponds to the standard deviation, same as in Bland-Altman plots. This figure shows our method does not significantly vary from the reference.

In Figure 12 for the monochrome cameras and Figure 13 for the RGB camera, it can be clearly seen from both the Bland-Altman and the Pearson Correlation plots that our method is closely correlated with the reference contact sensor under all three types of motions. The correlation is greater than 0.98 while the root-mean-square error (RMSE) is less than 2.5 for all our videos. In conclusion, the PBV method is highly robust to all three types of head movement, e.g., rotation, scaling and translation. The resulting heart rates are nearly identical to the reference.

IV. Improvement

In the previous section, we have already proved the motion robustness of the PBV method under three simple motion scenarios: rotation, scaling and translation. However, in reality subjects may induce all kinds of complex motions, which impose challenge on all rPPG methods. To further improve the motion robustness of the PBV method, one possible and simple way is to achieve the optimal PBV signatures by changing the experimental setup. Since we assume that motion has an equal influence over all color channels, we aim to obtain a blood volume pulse vector $\vec{P}_{bv}$ that has an angle with the motion vector [1 1 1] which is as large as possible to minimize the motion artifacts. According to Eq. (2), given the same hardware used for measurements, e.g., camera and illumination source, the PBV vector $\vec{P}_{bv}$ is stable for subjects with the same skin type. This implies we shall try different cameras and lamps in order to obtain the optimal $\vec{P}_{bv}$. To this end, we have mainly three different approaches: 1) using dedicated infrared LEDs as illumination, 2) using color cameras with different spectral responses, and 3) using three identical monochrome cameras by changing the filters inside. We choose 3) because we do not have many different cameras or LEDs and it is too costly and time-consuming to buy new ones.

A. Simulations

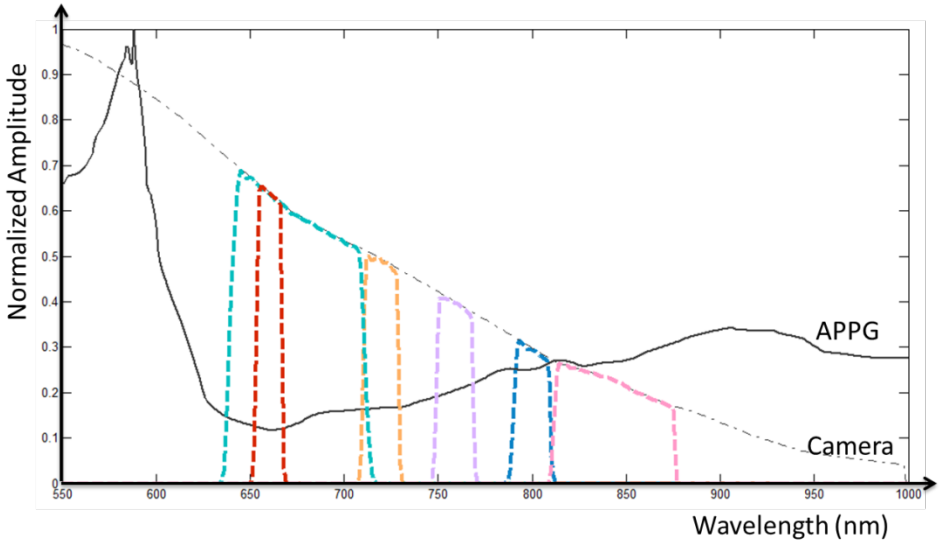

Figure 14. Normalized spectral response of the absolute PPG-amplitude, monochrome camera AVT Stingray F-046B, and six optical filters with center wavelengths at 661 nm, 675 nm, 720 nm, 760 nm, 800 nm, 842 nm, respectively.

Figure 14 shows the overview of the normalized spectral response of the absolute PPG, monochrome cameras and the filters available for our experiments. The sensor used in the monochrome cameras is the the Sony ICX415, which is used in the AVT Stingray F-046B . As mentioned in Section II, the RPPG-amplitude curve from Hülsbusch's model shows very low amplitudes in the infrared spectrum. On the other hand, the APPG-amplitude curve from Corral's

measurement contains a lot of noises and the spectral response of the halogen lamp used is unknown. Neither of these curves is sufficiently accurate for our research. Therefore, scientists at Philips Research performed their own measurements and built different models. For our study, we chose the reliable absolute PPG curve computed from the model developed by W. Verkruysse, a principal research scientist at Philips Research, who also helped set up the experiments. It is worth noting that even this APPG curve is not 100% accurate. With respect to the spectral response of the halogen lamp, we used a typical curve from Internet. Such inaccuracy will lead to distortions between the measured and simulated results, as can be seen from Table 3. The table shows that the measured PBV signature is around 6^o greater for the monochrome cameras and round 3^o smaller for the RGB camera compared to simulated results.

	675 nm	800 nm	842 nm	Angle	Angle_off
Mono_Pbv_PCA_sorted_median	0.29	0.61	0.74	19.34	6.08
Mono_Pbv_Simulated	0.37	0.63	0.68	13.62	

(a)

	Red	Green	Blue	Angle	Angle_off
RGB_Pbv_PCA_sorted_median	0.39	0.70	0.60	12.76	-3.29
RGB_Pbv_Simulated	0.36	0.74	0.57	15.55	

(b)

Table 3. Normalized mean PBV-vectors obtained from measurement and simulation for the monochrome cameras (a) and for the RGB camera (b). Angle corresponds to the angle with the motion vector [1 1 1] and Angle_off corresponds to the angle with the simulated results.

For the absolute PPG curve in Figure 14, a dip at around λ = 660 nm can be clearly recognized and the normalized amplitude starts to rise after that. The dip can partly explain the distortions in Table 3 because the noise was found to be stronger than the PPG signal around the dip [16]. Since our aim is to obtain the PBV vector $\vec{\mathrm{P}}_{\mathrm{bv}}$ with the maximum angle with the motion vector [1 1 1], the heuristic is thereby to select three filters in such a way that the APPG-amplitudes within their pass-bands differ from each other as much as possible. In other words, we should select filters with center wavelengths at around 660 nm, 800 nm and 930 nm, respectively.

Firstly, we attempt to find the theoretically optimal PBV-vector $\vec{\mathrm{P}}_{\mathrm{bv}}$ that may improve the motion robustness. Although we may lack some of the necessary optical filters currently, the results can provide a glimpse into the theoretically maximum angle with the vector [1 1 1]. Therefore, given the filter with center wavelength at 661 nm, we run the simulations in MATLAB in an attempt to find the other two filters accordingly, assuming the passband bandwidths of both filters are 40 nm. The results have confirmed our filter selection heuristic, as shown in Figure 15a. The peak of the angle values is found to be 26^o approximately, with two filters centering around 700 nm and 940 nm. Since we are interested in the infrared spectrum, we do not want two filters in the visible light spectrum (λ <= 700 nm) simultaneously. Then if we move the center wavelength of the filter further into IR spectrum, for example from 700 nm to 740 nm, without changing the other two filters, the angle will decrease to 25^o.

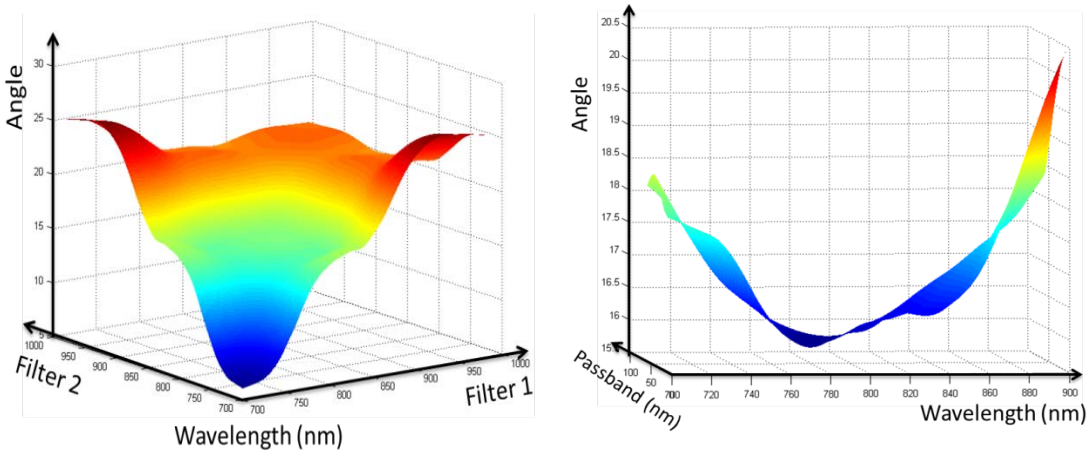

Figure 15. The angle with the simulated PBV-vector and the motion vector [1 1 1]. a) Assuming the filter centering at 661 nm is unchanged, we attempt to find the combination of the other two filters resulting in maximum angle. b) Assuming two filters centering at 661 nm and 842 nm unchanged, we try to find the optimal center frequency and passband width of the third filter.

In addition to the center wavelength λ_c, the filter's bandwidth can also affect the PBV-vector and the corresponding angle with [1 1 1]. For our second simulation, we kept one more filter ($\lambda_c = 842\ nm$ with the maximum cut-off wavelength of 875 nm) unchanged in order to find the optimal λ_c and passband width of the remaining filter. As can be seen from Figure 15b, the minimum angle is found to be around 15.8^o when $\lambda_c = 780\ nm$, and the angle increases when the filter moves towards to either visible light of the infrared spectrum. Another observation is that changing the passband width from 20 nm to 120 nm does not significantly change the angle (maximum 2^o off depending on λ_c).

In order to verify the simulations by experimental results, we shall utilize the available optical filters. We thus performed simulations with all possible combinations of 6 filters under the constraint that at least two filters are centering in the infrared spectrum. The maximum angle is found to be around 17.6^o, when using three filters centering at 661 nm, 720nm and 842 nm, respectively (around 4^o of improvement, compared to the filters we used in Section II and III). The corresponding simulated blood volume pulse vector is [0.37 0.49 0.79].

B. Experimental results

The experimental setup is exactly the same as described in Section II. This time in total 11 healthy volunteers took part in our study. Figure 16 shows that the PBV vectors measured by all 5 methods are quite similar and consistent. The two channels corresponding to filters with center wavelengths of 661 nm and 720 nm contain comparable and relatively small strengths of the PPG signal, because they are both close to the dip of the PPG curve. As can be seen from Table 4, on average, the measured PBV-vectors are around $7^o \sim 9^o$ off from the simulated results. Such distortion is probably caused by the curves we used for simulation. It is worth noting that the angle with the vector [1 1 1] has been improved by more than3^o, which is consistent with the simulated improvement (4^o). This confirmed our assumption of the angle improvement by changing the filters. As before, we still

choose the resulting normalized PBV-vector from the *Median of sorted PCA* method ([0.3719 0.3754 0.8490]) as the PBV signature for later evaluations.

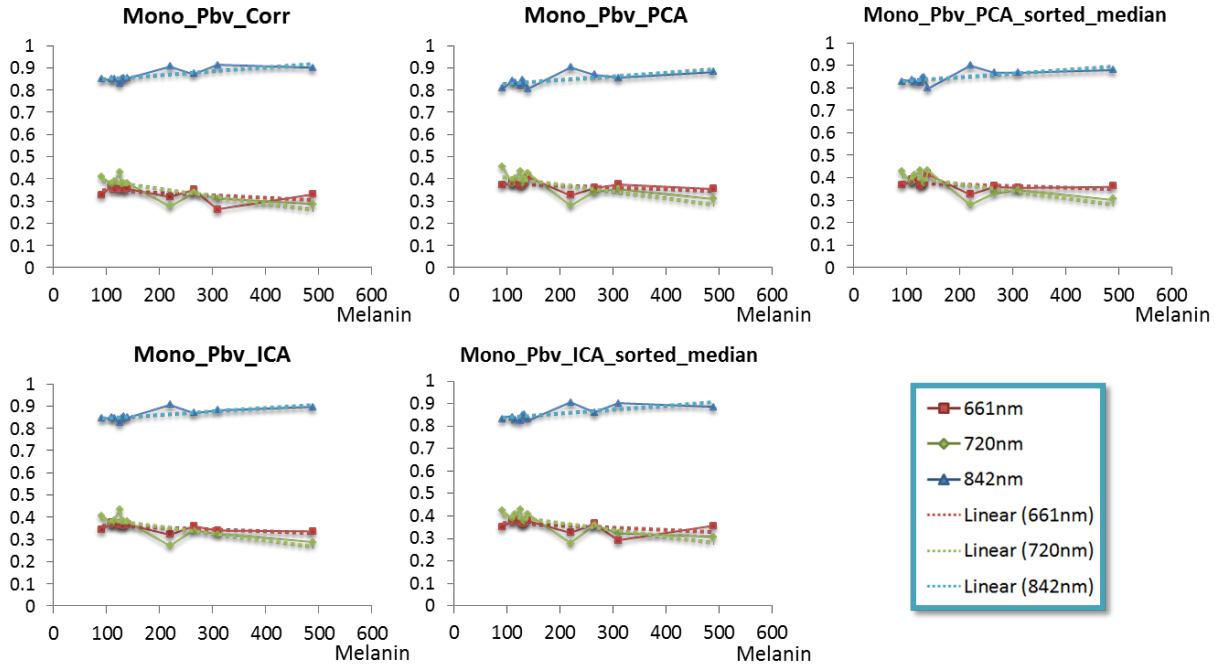

Figure 16. The normalized blood volume pulse vectors extracted from monochrome cameras AVT Stingray F-046B, plotted by measured melanin values. The bandpass optical filters used have center wavelengths of 661 nm, 720 nm and 842 nm, respectively. The methods used are Simple Correlation, Correlation with PCA, Median of sorted PCA, Correlation with ICA, and Median of sorted ICA.

	661 nm	720nm	842 nm	Angle	Angle_off
Mono_Pbv_Corr	0.34	0.36	0.87	25.02	8.80
Mono_Pbv_PCA	0.37	0.38	0.85	22.62	7.01
Mono_Pbv_PCA_sorted_median	0.37	0.38	0.85	22.84	7.27
Mono_Pbv_ICA	0.36	0.36	0.86	24.32	8.46
Mono_Pbv_ICA_sorted_median	0.36	0.37	0.86	23.65	7.70
Mono_Pbv_Simulated	0.37	0.49	0.79	17.64	

Table 4. Normalized mean PBV-vectors measured with different methods for monochrome cameras. Angle corresponds to the angle with the motion vector [1 1 1] and Angle_off corresponds to the angle with the simulated results.

Since this PBV signature was acquired over a relatively small dataset of 11 subjects, we intended to verify its reliability by using it as input for the PBV method to extract pulse signals. The overall results of our three evaluation metrics are shown in Figure 17. The SNRs range from 3.8 dB to 9.0 dB and the average SNR over all 11 subjects is around 6.4 dB. The ACCU metric shows the percentage of accuracy is higher than 98% for all subjects. The Bland-Altman plot indicates an excellent correlation between the extracted heart rates and the reference, where the difference is less than 1 bpm. The

results are indeed slightly better than in our first set of experiments shown in Section II, but since the sample is small, it is unfair to draw a conclusion regarding the improvement. Instead, since all of the three metrics show a nearly perfect performance of the PBV method, we believe the PBV signature is sufficiently reliable.

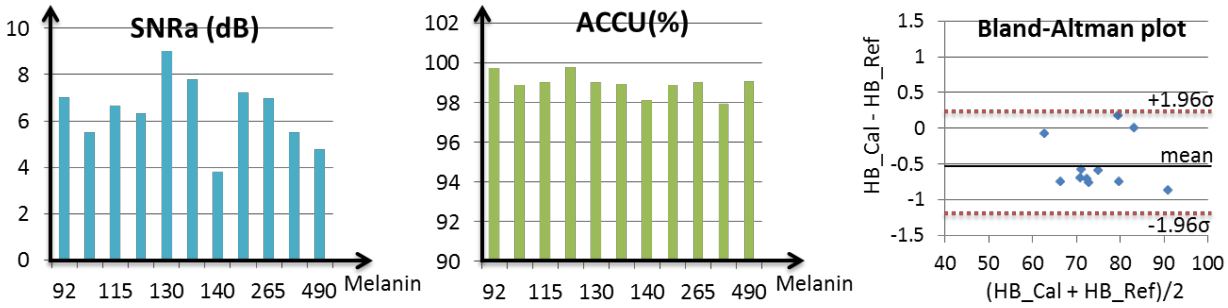

Figure 17. The evaluation results for monochrome cameras with difference metrics: from left to right, signal-to-noise-ratio (SNRa), accuracy (ACCU) and Bland-Altman plot.

Finally, an overview of all three PBV signatures obtained from simulation and measured from our dataset using the *Median of sorted PCA* method is shown in Table 5. The improvement of the angle with the motion vector [1 1 1] by changing the filters is 3.5^o. In section III we already observed that the performance of both monochrome cameras and RGB camera are assembling each other, even though their angles have a difference of 6.6^o. From our experience, we believe this new combination of filters will not lead to noticeable improvements on motion robustness so that it is unnecessary to continue the experiment on it. But if the measured improvements are consistent with the simulated results, the maximum angle from measurement using the optimal combination of filters (with center wavelengths of 661 nm, 740 nm and 940 nm) is expected to be around 30^o, which is 11^o greater than the one obtained with filters centering at 675 nm, 800 nm and 842 nm. We expect using these filters will greatly improve the motion robustness of the PBV method.

	Red	Green	Blue	Angle	Subjects
RGB_Pbv_Measured	0.39	0.70	0.60	12.76	41
RGB_Pbv_Simulated	0.36	0.74	0.57	15.55	
	675 nm	800 nm	842 nm	Angle	Subjects
Mono_Pbv_Measured	0.29	0.61	0.74	19.34	41
Mono_Pbv_Simulated	0.37	0.63	0.68	13.62	
	661 nm	720nm	842 nm	Angle	Subjects
Mono_Pbv_Measured	0.37	0.38	0.85	22.84	11
Mono_Pbv_Simulated	0.37	0.49	0.79	17.64	

Table 5. Overview of PBV signatures obtained from simulation and measured by the *Median of sorted PCA* method for RGB camera and monochrome cameras. Angle corresponds to the angle with the motion vector [1 1 1]. Size is the number of subjects in the dataset, from which the PBV signature is obtained.

C. Discussion about the discrepancy between simulation and measurements

As can be seen from Table 5, some discrepancies exist between measured and simulated results, and we believe the discrepancies are mostly attributable to the inaccurate curves used in the model. Specifically, five curves used as input to the model are color channel response $H_{chi}(\lambda)$ acquired from the documentation of the manufacturer, illumination spectrum $I(\lambda)$ measured from the lamps in our experimental setup, skin reflectance $\rho_s(\lambda)$, the RPPG-amplitude curve computed by Philips scientists, and the halogen lamp spectrum $I_h(\lambda)$ for APPG calculation. Among these five curves, the latter two are found to have a big impact on the simulated results. This is because the former three curves $H_{chi}(\lambda)$, $I(\lambda)$, $\rho_s(\lambda)$ of the AC and DC components in Eq. (2) will compensate each other.

Firstly, we start with the halogen lamp spectrum $I_h(\lambda)$ and its impact on the simulated results. According to Eq. (1), ideally, $I_h(\lambda)$ should be measured from the halogen lamp that was used for the APPG calculation. However, we could not get the detailed information about the halogen lamps from Philips. Therefore, we searched the Internet for it and used the one from [25], as shown in Figure 18a. It is worth noting that there exist many different spectrum curves, such as the one in Figure 18b [26]. To show their impact on the simulated results, we used these two spectrums as input to calculate the PBV signature, the results of which are summarized in Table 6. It can be clearly seen that different halogen lamp spectrum curves can lead to very different simulated results. Since we could not guarantee the spectrum curve chosen for simulations is the actual one used for APPG calculation, we conclude that such uncertainty will inevitably result in some discrepancies between simulation and measurement.

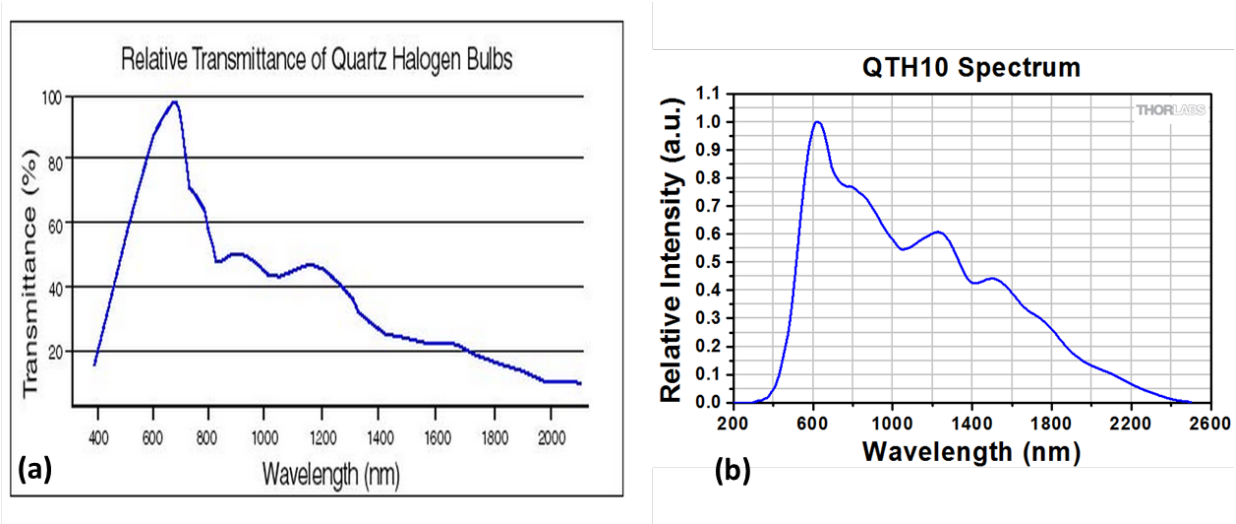

Figure 18. Relative transmittance response of two different halogen lamps. a) The one chosen for the simulations. b) QTH10 selected for comparison.

	Red	Green	Blue	Angle
RGB_Pbv_Simulated	0.36	0.74	0.57	15.55
Mono_Pbv_Simulated_QTH10	0.40	0.80	0.44	18.25
	675 nm	800 nm	842 nm	Angle
Mono_Pbv_Simulated	0.37	0.63	0.68	13.62
Mono_Pbv_Simulated_QTH10	0.25	0.60	0.76	21.56
	661 nm	720nm	842 nm	Angle
Mono_Pbv_Simulated	0.37	0.49	0.79	17.64
Mono_Pbv_Simulated_QTH10	0.27	0.36	0.89	28.60

Table 6. PBV signatures computed from simulations with two different halogen lamp spectrums.

Another factor that can cause the discrepancies is the PPG-amplitude curve. Both Figure 2 and Figure 14 show that all PPG curves show a dip at around λ = 660 nm, where the noise is found to be stronger than the signal [16]. When we used the optical filter with center wavelength around this dip (675 nm or 661 nm), the results would be harmed by noise. And since the PBV-vector $\vec{P}_{bv}$ is normalized, the inaccurate component of this channel will affect the other two, thereby causing variations to the results. The variation is found to be smaller for the RGB camera than for the monochrome cameras because the long-pass filter inside the RGB camera transmits the entire spectrum from 650 nm to 1000 nm, thus minimizing the noise and leading to more consistent results.

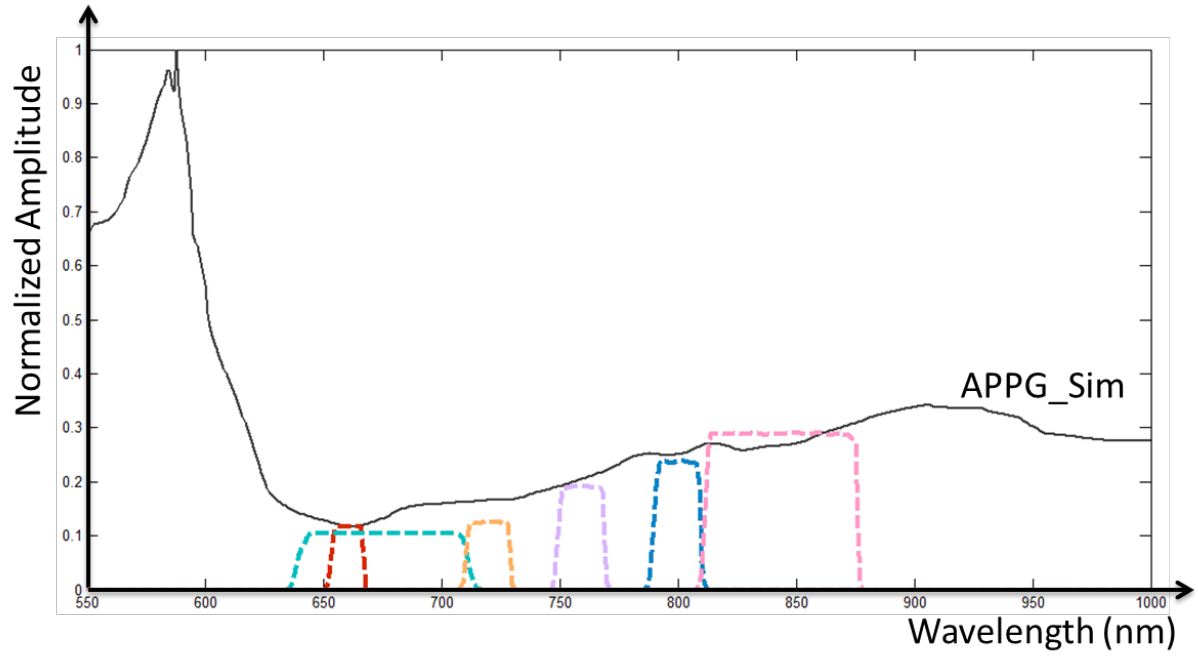

Figure 19. Normalized spectral response of the absolute PPG-amplitude used in simulations, and the six coarse sections of the PPG curve derived from our measured PBV signatures. The six sections correspond to the six optical filters with center wavelengths at 661 nm, 675 nm, 720 nm, 760 nm, 800 nm, 842 nm, respectively.

Assuming the halogen lamp spectrum $I_h(\lambda)$ used in our simulation and our measured PBV signatures are correct, we can accordingly derive a "true" APPG curve and compare it with the one used in the simulation. Since we have six optical filters, we can only derive a coarse APPG curve with six sections, each of which corresponds to the respective filter. And we have to assume the APPG amplitude is constant over the passband of the filters, which makes the derived APPG curve even coarser, as shown in Figure 19. However, it does show a mismatch between the original and the derived curves and it confirms our assumption that an inaccurate APPG curve can lead to discrepancies between simulated and measured results.

But more importantly, even though some discrepancies do exist, we can still observe a consistent trend in Table 5. For instance, the angle difference between simulation and measurement does not vary much for the monochrome cameras with different filters ($5.2^o\ and\ 5.6^o$). This indicates the improvement of the angle from measurement is consistent as expected from simulation. As a result, the simulation serves as a useful guidance for future improvement.

V. Conclusions and Future Work

In the thesis, we have studied the state-of-the-art rPPG method – PBV method – in detail and explored its application in the infrared spectrum. It is assumed that there exists a 'signature' of the variations of blood volume representing the relative pulsatilities of the normalized color channels of the video camera. The assumption had been proved in the visible light spectrum by de Haan *et al* [14]. Yet the answer whether this method can also be used in the infrared spectrum is unknown. Based on the five methods proposed by Mark van Gastel and Gerard de Haan, we computed the PBV signatures from raw videos: *simple correlation*, *Correlation with PCA, Median of sorted PCA, Correlation with ICA* and *Median of sorted ICA* in the IR spectrum.

By experimenting on 41 stationary subjects with both monochrome cameras and an RGB color camera, we have obtained stable PBV vectors with all five methods. The theoretically best method – *Median of sorted PCA* indeed resulted in PBV vectors with the least variation. The corresponding normalized mean PBV vectors were thereby defined as the PBV signatures. To verify its reliability and evaluate the PBV method, we used these PBV signatures to derive pulse rates from the 41 stationary subjects. The result has shown that the method is nearly identical to the reference contact sensor.

We used the same PBV signatures to evaluate the motion robustness on 13 subjects with three types of head movement: rotation, scaling and translation. Since the PBV signature for monochrome cameras has a bigger angle with the motion vector [1 1 1], we expected that the monochrome cameras would outperform the RGB camera, but the results did not show a noticeable difference because the performance of the monochrome cameras will be somewhat hampered by the image misalignment. Overall, the experimental results have shown a great agreement between our method and the reference, the correlation between them is greater than 0.98.

In a further attempt to improve motion robustness, we aimed to achieve the optimal PBV signatures by changing the optical filters in the monochrome cameras so that the resulting PBV signature would have a maximum angle with the vector [1 1 1]. The theoretically optimal combination of filters was found to have center wavelengths of 661 nm, 740 nm and 940 nm. Yet with limited filters in practice, filters centering at 661 nm, 720 nm and 842 nm shall increase the angle by 3^o. The corresponding PBV signature was measured on 11 stationary subjects and was proved reliable by experimental results.

We conclude that the PBV method can be applied in the infrared spectrum to extract comparable pulse signals to the commercial reference contact sensor. It is also highly robust to certain motions. In future, we can explore more approaches to obtain the PBV signature with maximum angle with the motion vector, such as changing the optical filters (the one with center wavelength of 940 nm), illumination sources (specific IR LEDs) or cameras. Before we have the necessary hardware, we may improve the methods to obtain the PBV signatures with less variation or optimize the ROI tracking algorithms in OpenCV. Specifically, the down scaling size of the ROI is empirically found to have a significant influence over the performance, and the existing skin classifier is not sufficiently accurate to distinguish between skin and non-skin pixels. These are the areas we can work in the future.

References

[1] Johns Hopkins Medicine. "*Vital signs (body temperature, pulse rate, respiration rate, blood pressure)*". Baltimore, US: Johns Hopkins Health Library. Information published online, Available: http://www.hopkinsmedicine.org/healthlibrary/conditions/cardiovascular_diseases/vital_signs_body_temperature_pulse_rate_respiration_rate_blood_pressure_85,P00866/, retrieved February, 2014.

[2] Politano L, Palladino A, Nigro G, Scutifero M, Cozza V. "*Usefulness of heart rate variability as a predictor of sudden cardiac death in muscular dystrophies*". Acta Myol 27:114-22. 2008.

[3] Kawase, M., Komatsu, T., Nishiwaki, K., Kimura, T., Fujiwara, Y., Takahashi, T., & Shimada, Y. "*Heart rate variability during massive hemorrhage and progressive hemorrhagic shock in dogs*". Canadian Journal of Anesthesia, 47(8), 807-814. 2000.

[4] Chen, W. L., & Kuo, C. D. "*Characteristics of heart rate variability can predict impending septic shock in emergency department patients with sepsis*". Academic emergency medicine, 14(5), 392-397.2007.

[5] "Heart rate variability", Available: http://en.wikipedia.org/wiki/Heart_rate_variability, retrieved June 2014.

[6] A. B. Hertzman, "*Photoelectric plethysmography of the fingers and toes in man*," Exp. Biol. Med., vol. 37, no. 3, pp. 529–534, 1937.

[7] M. Huelsbusch and V. Blazek, "*Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI*," Proc. SPIE, vol. 4683, pp. 110–117, 2002.

[8] C. Takano and Y. Ohta, "*Heart rate measurement based on a time-lapse image*," Med. Eng. Phys., vol. 29, pp. 853–857, 2007.

[9] Verkruysse, W., Svaasand, L. O., & Nelson, J. S. "*Remote plethysmographic imaging using ambient light*". Optics express, 16(26), 21434-21445. 2008.

[10] Poh, M. Z., McDuff, D. J., & Picard, R. W. "*Non-contact, automated cardiac pulse measurements using video imaging and blind source separation*". Optics Express, 18(10), 10762-10774. 2010.

[11] Lewandowska, M., Ruminski, J., Kocejko, T., & Nowak, J. "*Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity*". In Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on (pp. 405-410). 2011.

[12] Balakrishnan, G., Durand, F., & Guttag, J. "*Detecting pulse from head motions in video. In Computer Vision and Pattern Recognition (CVPR)*", 2013 IEEE Conference on (pp. 3430-3437). 2013.

[13] G. de Haan and V. Jeanne, "*Robust pulse rate from chrominance-based rPPG*," Biomedical Engineering, IEEE Transactions on, vol. 60, no. 10, pp. 2878–2886, 2013.

[14] G. de Haan and A. van Leest, "*Improved motion robustness of remote-PPG by using the blood volume pulse signature*," to be published in 2014.

[15] M. Hülsbusch, "*An image-based functional method for opto-electronic detection of skinperfusion*", PhD-thesis, RWTH Aachen, department of EE, (in German) 2008.

[16] Martinez, Luis F. Corral, Gonzalo Paez, and Marija Strojnik. "*Optimal wavelength selection for noncontact reflection photoplethysmography*." International Commission for Optics (ICO 22). International Society for Optics and Photonics, 2011.

[17] M. Thomas, "*Large Scale Study on Remote Photoplethysmography*", Graduation Report resulting from internship at Philips Research Eindhoven, ISEN-Brest, 2011.

[18] G. Bradski, "*The opencv library*," Dr. Dobb's Journal of Software Tools, 2000.

[19] Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2012). "*Exploiting the circulant structure of tracking-by-detection with kernels*," In Computer Vision–ECCV 2012 (pp. 702-715). Springer Berlin Heidelberg, pp. 702-715, 2012.

[20] Farnebäck, Gunnar. "*Two-frame motion estimation based on polynomial expansion*." Image Analysis. Springer Berlin Heidelberg, pp. 363-370, 2003.

[21] "*filtfilt - Zero-phase digital filtering*", Available: http://www.mathworks.nl/help/signal/ref/filtfilt.html, retrieved June 2014.

[22] J.F. Cardoso, "*MATLAB implementation of source separation of complex signals with JADE*", Available: http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m, retrieved June 2014.

[23] W. Wang, S. Stuijk and G. de Haan, "*Exploiting Spatial-redundancy of Image Sensor for Motion Robust rPPG*", to be published in 2014.

[24] "*Registering Multimodal MRI Images*", Available: http://www.mathworks.nl/help/images/registering-multimodal-mri-images.html, retrieved June 2014.

[25] "*Lab 10 INVESTIGATION OF FACTORS AFFECTING THE HILL REACTION*", Available: http://openwetware.org/wiki/BISC110/F10:_Series_3_Experiment_10_Hill_Reaction_Variables, retrieved June 2014.

[26] "*QTH10(/M) Quartz Tungsten-Halogen Lamp – User Guide*", Available: http://www.thorlabs.de/thorcat/TTN/QTH10-Manual.pdf, retrieved June 2014.