Robust pulse-rate from chrominance-based rPPG

Gerard de Haan and Vincent Jeanne

***Abstract*—Remote photoplethysmography (rPPG) enables contactless monitoring of the blood volume pulse using a regular camera. Recent research focused on improved motion robustness, but the proposed blind source separation techniques (BSS) in RGB color space show limited success. We present an analysis of the motion problem, from which far superior *chrominance-based* methods emerge. For a population of 117 stationary subjects, we show our methods to perform in 92% good agreement ($\pm1.96\sigma$) with contact PPG, with RMSE and standard deviation both a factor of two better than BSS-based methods. In a fitness setting using a simple spectral peak detector, the obtained pulse-rate for modest motion (bike) improves from 79% to 98% correct, and for vigorous motion (stepping) from less than 11% to more than 48% correct. We expect the greatly improved robustness to considerably widen the application scope of the technology.**

I. Introduction

PHOTOPLETHYSMOGRAPHY (PPG) is an optical technique to monitor various vital signs, like pulse-rate, respiratory rate and blood oxygenation, first described in the 1930s [1] and highly popular because of its non-invasiveness. Essentially, PPG detects the optical absorption variations of the human skin due to blood volume variations during the cardiac cycle.

Earlier work has shown that these variations can also be measured at a distance leading to remote-PPG (rPPG) [2], [3]. Publications have even shown successful rPPG using a regular color video camera in ambient light conditions [4], [5], [6], [7]. rPPG is a highly relevant development for cases where contact has to be prevented because of extreme sensitivity, e.g. neonates, skin-damage, or when an increased unobtrusiveness is required/desired (surveillance, fitness).

The main concern with rPPG is robustness to subject motion. Recent research has brought some improvement, but significant motion renders current algorithms useless. At this point we aim to contribute by analyzing how motion enters the pulse-signal and deriving far superior motion robust rPPG algorithms from the analysis.

Essentially all motion robust rPPG techniques profit from the fact that the variations in optical absorption of the human skin depend on the wavelength used, as shown in Figure 1. Motion of the skin relative to the sensor, on the other hand, mostly affects the light reflected or transmitted by the skin *regardless* the wavelength. Huelsbusch and Verkruijsse, in 2008, found the PPG signal had different relative strength in the three color channels of a video camera pointed at human skin [8], [4]. Huelsbusch first exploited this difference to achieve motion robustness by separating the noise and the PPG signal into two independent signals built as a linear combination of two color channels [8]. One combination approximated the clean pulse-signal, the other the motion artifact, and the energy in the pulse-signal was minimized to optimize the combination. Poh et al. extended this work proposing a linear combination of all three color channels defining three independent signals with Independent Component Analysis (ICA) using non-Gaussianity as the criterion for independence [5]. Lewandowska et al. varied this concept defining three independent linear combinations of the color channels with Principal Component Analysis (PCA) [6]. With both Blind Source Separation (BSS) techniques, the component that carries the pulse-signal is a priori unknown. Commonly, the selection assumes the pulse-signal shows the strongest periodicity. Consequently, periodic motion as in a fitness setting will render the selection useless. Moreover, a fairly long observation interval is required to have sufficient resolution in the frequency domain, which prohibits adaptation to quickly changing statistics.

In Section II of this paper, we shall analyze how motion enters the pulse-signal. From this analysis new techniques for rPPG emerge that shall be shown superior to all earlier methods both in SNR and motion robustness. In Section III, we provide the assement details, the results of which are shown for 117 *stationary* subjects over a broad range of skin-types in Section IV, and for some subjects *exercising* in a gym to test motion robustness in Section V. Finally, we draw our conclusions in Section VI.

II. Analysis

For the following analysis, we assume an area of skin is illuminated by a light source and registered with an RGB video camera. We assume blood volume changes, due to heartbeat, in human skin lead to color changes in the reflected light. The intensity of a given pixel in image number i in color channel $C \in \{R, G, B\}$ registered by the camera, can be modeled as:

$$C_i = I_{Ci}(\rho_{Cdc} + \rho_{Ci}), \quad (1)$$

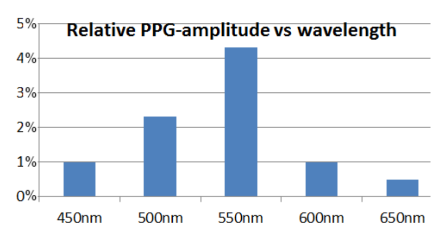

Fig. 1. As reported in [9], the amplitude of the ppg-signal in light reflected from the skin varies as a function of the wavelength, showing a strong peak around 550nm and a dip around 650nm.

Gerard de Haan and Vincent Jeanne are with Philips Innovation Group, Research, Eindhoven, The Netherlands, e-mail: (G.de.Haan@Philips.com, Vincent.Jeanne@Philips.com).

Manuscript received January 15th, 2013; revised May 6th, 2013.

where I_{Ci} is the intensity of the light source integrated over the exposure time of the camera in image i for color channel C, ρ_{Cdc} is the stationary part of the reflection coefficient of the skin in color channel C, while ρ_{Ci} is used to indicate the zero-mean time-varying fraction caused by the pulsation of the blood volume.

Consequently, the amplitude of the heartbeat induced color variation is proportional to the intensity I_{Ci} of the light source in each of the color channels. To produce a pulse-signal that is independent of the presumed stationary color of the light source and its brighness level, we can normalize each color channel C by dividing its samples by their mean over a temporal interval:

$$C_{ni} = \frac{C_i}{\mu(C_i)}, \tag{2}$$

where $\mu(C_i)$ can be a running average centered around image i, or an average of an overlap-add processing interval that includes image i. In either case, the average is preferably taken over a number of images such that the interval contains at least a pulse-period.

Normalization, however, cannot prevent influence of intensity variations *during* the normalization period. A typical situation where this occurs is when the skin surface moves with respect to the light source[1]. If we assume that such intensity modulations are equal for all channels, a ratio of two normalized color channels would not be affected by motion. The pulsatility of the blood would still be available in this ratio provided that the pulsatility due to blood volume changes is different in the individual color channels. Given the pulsatility as a function of wavelength exhibits a strong peak in green and dips in red [9], [10], as illustrated in Figure 1, the ratio of normalized green and red would make a motion robust pulse signal S_i:

$$S_i = \frac{G_{ni}}{R_{ni}} - 1 = \frac{G_i}{R_i} . \frac{\mu(R_i)}{\mu(G_i)} - 1. \tag{3}$$

We shall refer to this method as $RoverG$. We note that this method is related to a method suggested by Huelsbusch [8], where a weighted difference of red and green is suggested. The relation between ratios and differences shall be elaborated later in this section.

In practice, the $RoverG$ method works less perfect, since the light reflected from the skin consists of 2 components as described by the dichromatic reflection model [11]:

- A diffuse-, or body-, reflection component, which has traveled through the skin and shows its color including variations thereof due to the cardiac cycle; and
- A component that is directly reflected from the surface of the skin, the specular reflection component, which shows the color of the illuminant and no pulse-signal.

If we include specular reflection in our model (Eq. (1)), we get:

$$C_i = I_{Ci}(\rho_{Cdc} + \rho_{Ci} + s_i), \tag{4}$$

[1]Motion may also change the focus of a pixel to a different location on the skin, but the color change that results from this can be minimized by averaging over a large group of pixels, or the application of motion-compensation. We shall not elaborate the issue further in this paper.

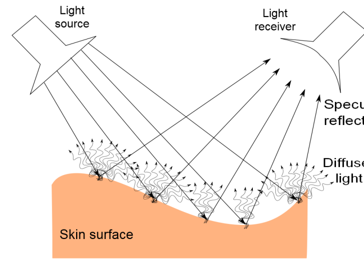

Fig. 2. Illustration of specular and diffuse reflection. Through scattering of the light inside the skin, the diffuse reflection changes color with the blood volume of the skin, whereas the specular reflection exhibits the color of the light source and is not affected by blood volume changes.

where s_i is the additive specular reflection contribution. The specular reflection component s_i is identical for all color channels, whereas the stationary part of the skin reflection, ρ_{Cdc}, is different for the individual color channels C, with $\rho_{Rdc} > \rho_{Gdc} > \rho_{Bdc}$ [12]. Consequently, the registered color depends on the specular reflection fraction in the total reflected light. The relative contribution of specular and diffuse reflections, which together make the observed color, depend on the angles between camera, skin and light source. Therefore they vary over time with motion of the person in front of the camera, and create a weakness in the proposal using normalized red over green, as the additive specular component is not eliminated in the ratio. Figure 2 illustrates the situation.

Recognizing this weakness, we see a possible improvement by adding the third color channel. If we, initially, assume white light, we note that the specular reflection affects all channels by adding an identical (white light) specular fraction to their respective diffuse reflection component. This implies that we can eliminate the specular reflection component by using *color difference*, i.e. *chrominance*, signals. From three color channels, e.g. RGB[2], we can build two orthogonal chrominance signals, e.g. $X = R - G$ and $Y = 0.5R + 0.5G - B$[3]. Again the variations due to the blood volume changes in the skin will likely be different, while motion affects both chrominance signals identically. A ratio of the two would be an interesting candidate rPPG algorithm that we shall reference to as $XoverY$[4]:

$$S = \frac{X_n}{Y_n} - 1 \tag{5}$$

Although this algorithm uses normalized color difference signals, X_n and Y_n, the color channels themselves, R, G, B, are not normalized and the algorithm will be imperfect for non-white illumination. To enable correct functioning with colored light sources, we investigated *skin-tone standardization*. We found from our large scale experiment described in Section III, that good results could be obtained, for the whole range of skin-types, assuming a fixed skin-tone where the normalized skin tone, $[R, G, B]/\sqrt{R^2 + G^2 + B^2}$, is assumed to be the

[2]To simplify later equations we dropped the sample index, i.e. from here on we write R instead of R_i.

[3]It is equally possible to use the U and V channels of a YUV color-space, which are slightly different from the proposed vectors

[4]For simplicity, we re-use S for the resulting pulse signal from all methods, although clearly they are different signals

TABLE I
ANGLE (DEGREES) BETWEEN STANDARDIZED AND ACTUAL SKIN-TONE SHOWS THAT THE DEVIATIONS DEPENDS ON THE PHOTO-TYPE BUT REMAIN SMALL.

Photo-type:	1	2	3	4	5
Angle:	2.6^o	2.7^o	3.6^o	5.4^o	1.8^o

same for everyone under white light:

$$[Rs, Gs, Bs] = [0.7682, 0.5121, 0.3841]. \tag{6}$$

We now can correct for potentially non-white illumination by first dividing the individual color channels by their means and next multiply $[R_n, G_n, B_n]$ with $[0.7682, 0.5121, 0.3841]$, i.e. the standardized RGB channels results as:

$$R_s = 0.7682R_n \; G_s = 0.5121G_n \; B_s = 0.3841B_n. \tag{7}$$

The result is an algorithm that can work correctly regardless the color of the illuminant:

$$S = \frac{X_s}{Y_s} - 1, \tag{8}$$

with

$$\begin{aligned} X_s &= \tfrac{R_s - G_s}{0.7682-0.5121} = 3R_n - 2G_n \text{ and} \\ Y_s &= \tfrac{R_s + G_s - 2B_s}{0.7682+0.5121-0.7682} = 1.5R_n + G_n - 1.5B_n. \end{aligned} \tag{9}$$

In Table I we show the angle between the actually measured vectors $[R, G, B]/\sqrt{R^2 + G^2 + B^2}$ for different skin-types, under white light, and the standardized skin vector $[Rs, Gs, Bs] = [0.7682, 0.5121, 0.3841]$. The results corroborate the underlying assumption that skin tones have roughly identical coordinates in RGB-space under white illumination.

Although the methods of Huelsbusch, Poh et al. and Lewandowska et al. may seem quite different as they use a linear combination of the individual sensor signals rather than a ratio, we shall show that this difference is actually very small.

To support this statement, we re-write Eq. (8):

$$\log(1 + S) = \log(\tfrac{X_s}{Y_s}) = \log(X_s) - \log(Y_s). \tag{10}$$

From a Taylor expansion of the logarithm:

$$\log(x) = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + ... \tag{11}$$

we can see that, since all arguments of logs in Eq. (10) are close to 1 and re-using Eq. (7), we may approximate Eq. (8) by:

$$S \approx X_s - Y_s = 1.5R_n - 3G_n + 1.5B_n. \tag{12}$$

The pulse-signal resulting from the methods proposed by Huelsbusch, Poh et al. and Lewandowska et al. can be written as a linear combination of the individual color channels:

$$S = c_1R_n + c_2G_n + c_3B_n. \tag{13}$$

The main difference between these methods is in the calculation of the coefficients c_i, while Poh et al. and Lewandowska et al. need additional heuristics to select the correct component. In contrast, the algorithm that emerges from our analysis has fixed coefficients that do not have to be estimated with advanced statistical methods and requires no further heuristics to distinguish the pulse-signal from other components present in the RGB channels. We shall therefore refer to this algorithm as $Fixed$.

For a final sophistication of our proposed algorithm, we return to Eq. (12) and recognize that our small blood volume pulse signal S results as a difference of the two signals X_s and Y_s the variations in which may be strong compared to the pulse-signal. If the skin-tone standardization is slightly off, the result will be that the variations in X_s and Y_s will not have identical amplitudes. We recognize though, that we can correct for this using:

$$S = X_f - \alpha Y_f, \tag{14}$$

with

$$\alpha = \frac{\sigma(X_f)}{\sigma(Y_f)}, \tag{15}$$

where $\sigma(X_f)$ is the standard deviation of X_f, and for best results we used X_f and Y_f, which are the band-passed filtered versions of X_s and Y_s. This allows for a minimization of the large in-band disturbances in the output signal with simple statistics[5]. We shall refer to this algorithm as $X_s min \alpha Y_s$.

To illustrate the algorithm in normalized RGB-space, we use Eq. (9), to re-write Eq. (14):

$$S = 3(1 - \tfrac{\alpha}{2})R_f - 2(1 + \tfrac{\alpha}{2})G_f + \tfrac{3\alpha}{2}B_f \tag{16}$$

where R_f is the band-passed filtered version of R_n etc.

III. ASSESSMENT DETAILS

In Sections IV and V, we shall present the results of a comparison of the various methods resulting from our analysis in the previous section and some state-of-the-art methods. The purpose of this benchmark is

- To assess the accuracy of rPPG over a large population of 117 stationary subjects, recorded by Thomas [13]
- To compare the signal quality (SNR) of the chrominance-based methods emerging from our analysis with the recent BSS-based methods
- To assess the motion robustness of all mentioned methods by comparing pulse-rates obtained from exercising subjects in a fitness.

The current section provides the details of these assessments.

A. Study setup static subjects

We used a 1024x752 pixels, 8 bit, global shutter RGB CCD camera (type USB UI-2230SE-C of IDS Gmbh) operated at 20 pictures/sec and focused at the subject's face using a flexible C-mount lens (Tamron 12VM412ASIR) maximizing the amount of facial pixels in the image. The duration of the video recording was set to one minute and uncompressed data was stored. All recordings were made in a controlled environment using professional studio illumination and subjects were asked

[5]Note that this minimization does NOT introduce the risk of eliminating the pulse-signal itself in case there are no disturbing motions, since the pulse signal is in anti-phase for X_s and Y_s, while the motion is in phase

TABLE II
FITZPATRICK PHOTO-TYPE DISTRIBUTION OVER THE STUDY POPULATION AS ESTIMATED FROM THE EXPERIENCE MELANIN VALUES.

Estimated photo-type:	1	2	3	4	5	6
Number in test data:	4	4	34	47	24	5

to sit and relax for two minutes prior to the recording to ensure a stable pulse-rate. Also, they were asked to remain stationary for the duration of the recordings. In parallel with the video, we synchronously recorded the raw pulse-oximeter data from a transmissive pulse-oximeter finger clip of Contec Medical Systems, model CMS50E, using the USB protocol available on the device. To extract the pulse-rate from the output signals of the different methods, we simply performed a peak detection in the frequency domain using a 512 point FFT on the Hanning windowed signals. The outputs are pre-processed using a FIR band-pass filter with cutoff frequencies 40-240 BPM to select the pulse frequencies. The raw PPG signal obtained from the reference contact sensor is processed exactly the same to eliminate possible effects from post-processing applied in the reference sensor.

B. Recruitment process static subjects

In total 117 healthy volunteers took part in the study. Informed consent was obtained from each subject, and the study was approved by the Internal Committee Biomedical Experiments of Philips Research. Thomas aimed at having all skin types represented, although skin types are not equally distributed. She assessed the skin type by measuring the melanin content on the subject's face using a skin pigmentation analyzer, SPA 99, from Courage & Khazaka. These melanin measurements can be loosely linked to the Fitzpatrick photo-types according to the manual of the device. Table II shows the distribution of these loosely estimated photo-types over the study population. No restrictions were applied to the subjects with respect to alcohol and/or caffeine intake, smoking habits, etc.

C. Pre-processing static subjects

Even if the recording condition ensures minimal lighting variation in the environment, possible noise in the signal can be introduced by the subject's movements. To further minimize the impact of unintended motion a segment, starting at $i = i_s$, of 500 consecutive pictures exhibiting the smallest amount of inter-frame motion was selected from the longer video sequence as follows:

$$i_s = \text{argmin}_n \sum_{i=n}^{n+499} (abs(I_i(\vec{x}) - I_{i+1}(\vec{x}))) \tag{17}$$

where $I_i(\vec{x})$ is the pixel intensity at location $\vec{x} = (x, y)$ in picture number i of the video sequence, defined as:

$$I_i(\vec{x}) = \frac{R_i(\vec{x}) + G_i(\vec{x}) + B_i(\vec{x})}{3} \tag{18}$$

Then, the extraction of the raw signals from the frame segment required by the selected methods consists in:

- Applying a face detector as introduced by Voila & Jones [14] to define a region of interest (ROI).
- Applying a simple skin selection process, which produces a skin-mask, inside the ROI. This process is applied to remove all pixels containing facial hairs and facial features that pollute the rPPG signal. The average of the skin pixel values is the output of this step.

Repeating these steps on each frame of the video sequence we obtain the raw RGB signals that are used as basis for the rPPG analysis of the different methods.

D. SNR-metric

To obtain a quality metric from the signals produced by the different methods we use signal-to-noise ratio (SNR) analysis. We compute the ratio of the energy around the fundamental frequency plus the first harmonic of the pulse-signal and the remaining energy contained in the spectrum defined by:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{f=30}^{240} (U_t(f)\hat{S}(f))^2}{\sum_{f=30}^{240} (1 - U_t(f))\hat{S}(f))^2} \right), \tag{19}$$

where $\hat{S}(f)$ is the spectrum of the pulse-signal, S, f the frequency in beats per minute (BPM), and $U_t(f)$ a binary template window as defined in Figure 3.
Due to the very controlled recording with stationary subjects the spectra are relatively clean. For a given video segment we use the highest frequency peak detected in the *reference* signal as the fundamental frequency and select the first harmonic accordingly. This is done to avoid any bias in the metric in case a rPPG spectra exhibit a highest peak that is different from the actual reference peak.

E. Benchmark algorithms

To benchmark our chrominance-based rPPG proposals, we compare the output with the state-of-the-art using the most recent rPPG-methods of Poh et al. and Lewandowska et al. as our benchmark algorithms. These two methods are BSS-based using ICA (hence method name: ICA) and PCA (method name: PCA), respectively. The components are band-pass filtered and an heuristic is applied to select the proper component. In the approach of Poh et al., the raw temporal traces, for color channel C, obtained from a skin region are first detrended using a smoothness priors approach (smoothing

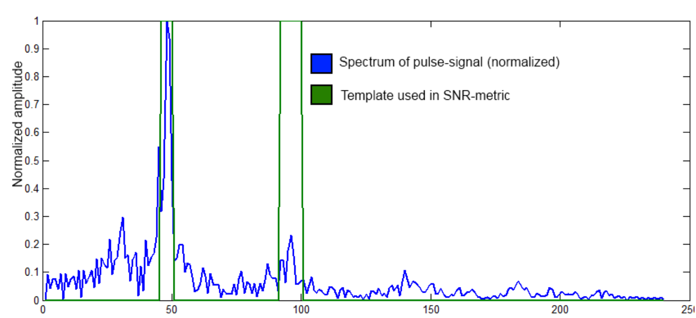

Fig. 3. The SNR calculation uses a template passing 5 bins in the 512 bin spectrum, centered around the contact sensor pulse-rate (10 bins around the first harmonic) to allow for heart-rate variability. The SNR is measured by the energy ratio of the components inside and outside the template.

Fig. 4. Fitness devices and camera view as used in our motion robustness test. From left to right: a stationary bike, a stepping device, and the camera view of a subject on the stepping device.

parameter = 10 and cutoff frequency 0.89 Hz) and normalized:

$$C_{ni} = \frac{C_i - \mu(C_i)}{\sigma(C_i)}, \tag{20}$$

where C_i is the raw temporal signal of a color channel and $\mu(C_i)$ and $\sigma(C_i)$ are the mean and standard deviation of the color channels, respectively. The resulting three normalized signals are decomposed into independent source signals using ICA based on the joint approximate diagonalization of eigenmatrices algorithm (JADE) [15].

Since ICA returns the independent components in random order, Fourier analysis is used to find the component exhibiting the highest peak in its normalized power spectrum. This component is used as the output pulse-signal after filtering it with a 5-point averaging filter and a band-pass filter selecting the frequency components between 40 and 240BPM.

In the approach of Lewandowska et al., the raw RGB temporal traces obtained from a skin region are first filtered using a FIR band-pass filter with cutoff frequencies 40-240BPM. This set of filtered signals is then decomposed into 3 uncorrelated source signals using PCA. To select the proper component as the pulse-signal, the same heuristics as with ICA is used.

F. Assessment of motion robustness

To test motion robustness, we set up a second experiment in which we evaluate resulting output signals obtained from subjects exercising in a gym. To include moderate and strong motion, we asked a subject to exercise on a stationary bike and a stepping device, respectively. A photograph of these devices is shown in Figure 4, together with a screenshot from the camera used to register the pulse-rate during the stepping exercise.

We assumed that in such uncontrolled environments with a mixture of daylight-fluorescent light and significant motion,

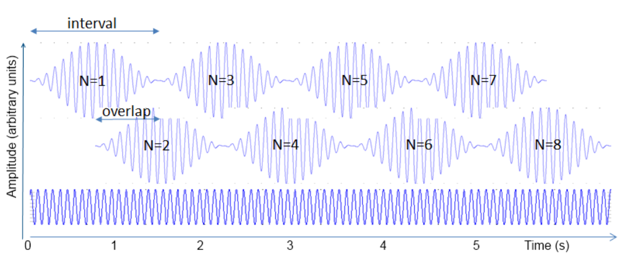

Fig. 5. Illustration of the overlap-add procedure. Every interval an optimized pulse-signal is calculated and multiplied with a Hanning window. Half the interval length later this is repeated and the pulse output signal (bottom) results as the sum of these overlapping pieces.

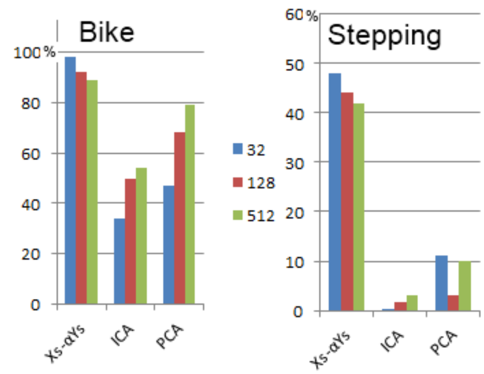

Fig. 6. Percentage of time that the frequency peak corresponds to pulse-rate for overlap-add interval of 32 up to 512. The chrominance-based methods perform better for shorter intervals. Methods based on BSS, however perform best with longer intervals. With serious motion (stepping device), this may be the reason BSS-based methods break down completely.

all methods could profit from a normalization over an interval significantly shorter than the length of the exercise. Also the methods that apply some form of optimization (ICA, PCA, $Xsmin\alpha Y_s$) may profit from allowing them to adapt the optimization during the exercise. Consequently, there could be an advantage in separately normalizing / optimizing partially overlapping time intervals during the exercise, and glueing the resulting pieces together in an overlap-add fashion using Hann windowing on individual intervals:

$$S_i = \sum_N wh_{N,i} S_{N,i}, \tag{21}$$

where for interval number $N = \lfloor 2i/interval \rfloor$, $S_{N,i}$ is the optimized signal from pictures in the interval, i.e. $i \in [(N-1)interval/2 + 1 : (N+1)interval/2]$, while $wh_{N,i}$ is the Hann windowing function centered in interval N and zero outside the interval:

$$wh_{N,i} = 0.5 - 0.5cos(2\pi i/interval). \tag{22}$$

The process is illustrated in Figure 5.

When using the overlap-add processing, the optimal optimization / normalization interval length has to be determined. To this end, we ran an initial experiment in which we varied the interval length from 32 ...up to 512 picture periods, i.e. between 1.6 and 25.6 seconds. Also in our motion robustness assessment, a simple peak-detector in the frequency domain was used, as described in Section III A, but we used a shorter sliding Fourier window of 256 picture periods, i.e. about 12 seconds, as we expect the pulse-rate to change more quickly. The resulting percentage of time this peak corresponds to the actual pulse-rate for the exercise on the stationary bike and different interval lengths is shown in Figure 6.

Figure 6 shows that for the stationary bike, our chrominance-based method $X_s min \alpha Y_s$ performs best with a relatively short interval length of 32 picture periods, i.e. about 1.6 seconds. This behavior was typical for all chrominance methods and allows the algorithms to adapt more quickly to changing distortion statistics. Shorter interval lengths typically decrease the performance of the methods based on BSS. This is likely caused by the heuristic to select the correct

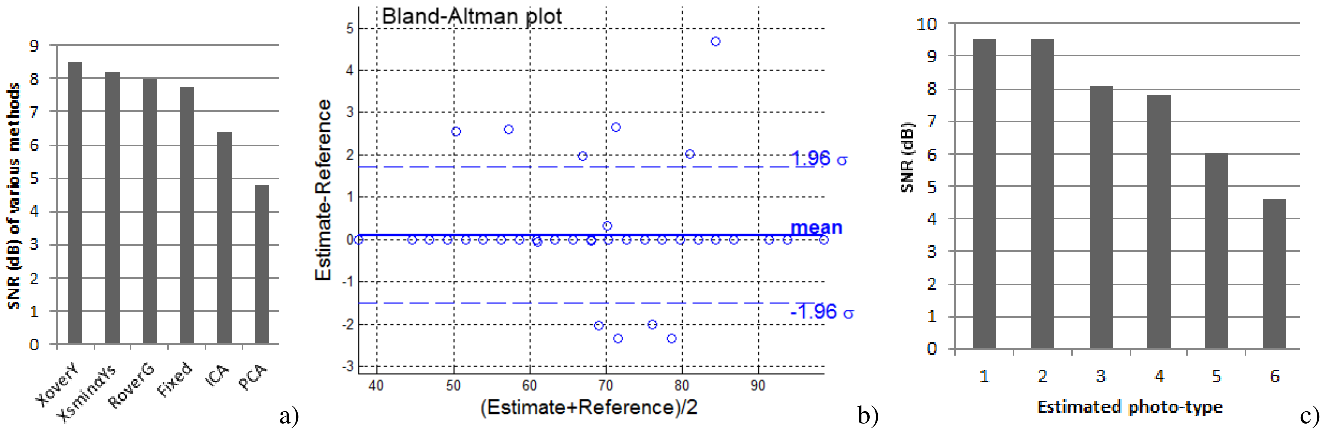

Fig. 7. a):The obtained SNR for the various methods obtained from our database of 117 static persons. b): Bland-Altman plot indicating 92% good agreement ($\pm 1.96\sigma$) between rPPG and contact PPG (shown for $XoverY$ method). c): The SNR as a function of the estimated photo-type shows that increased melanin content of the skin leads to a reduced fraction of the diffusely reflected light from the skin and consequently a decreased SNR of the rPPG signal.

component, which fails for short intervals as the resolution of the spectrum, calculated to find the periodic signal, drops.

IV. Results static subjects for range of skin-types

As shown in Table III, the pulse-rates obtained by the various rPPG methods are very similar to the one obtained by the reference contact sensor. The Pearson correlation between the two measurements is high with $r = 1$ (0.97) for the best (worst) result ($P < 0.005$). The best linear fit, using LMSE optimization, gives a slope of B=1.00 (0.96) for the best (worst) method. The standard deviation observed between the two sets is $\sigma = 0.8$ (2.6), and the RMSE is $E = 0.4$ (1.1) BPM for best (worst) method. Figure 7b shows the Bland-Altman plot of the two measurement sets for the best method. The figure shows 92% good agreement ($\pm 1.96\sigma$) between the contact and remote PPG measurements. Also the accuracy appears to be independent of the pulse-rate of the subjects, which suggests the pulse-rate range coverage of this technique is as broad as the one obtained with contact PPG sensors.

Referring to Table III and Figure 7, we conclude that the results correspond to our expectations. The use of color difference signals, $XoverY$ is a little better than $RoverG$, since it eliminates the effect of specular reflections. Since we only used white illumination of the skin in our experiment the method, $Fixed$, that uses skin-tone standardization can only do worse. We note that this loss is modest, which confirms the fairness of the underlying assumption. Our final sophistication, algorithm $X_s min\alpha Y_s$ can largely recover this small loss introduced by skin standardization.

TABLE III
Results of all rPPG methods correlate very well with the contact sensor signal for static subjects, although the average error and standard deviation are higher for the BSS-based methods. The table shows Pearson correlation r, slope of linear fit B, standard deviation σ, and RMSE, ϵ.

	$RoverG$	$XoverY$	Fixed	$X_s min\alpha Y_s$	ICA	PCA
r:	1.00	1.00	0.99	1.00	0.97	0.99
B:	1.00	1.00	0.99	0.99	0.97	0.96
σ:	1.1	0.8	1.2	0.9	2.6	1.8
ϵ:	0.5	0.4	0.5	0.5	1.1	0.9

The resulting SNR for the methods discussed in Section II, averaged over all 117 subjects in our experiment, is shown in Figure 7a. Although all methods can provide the accurate pulse-rate in this controlled setting with stationary subjects, the SNR-quality of the signals is clearly better for the chrominance-based methods. Figure 7c illustrates for a single method, $XoverY$, that the SNR also depends on the skin type. The difference observed between the two extremes of the Fitzpatrick scale is about a factor of two. This decrease in SNR for darker skin-types makes sense, as the higher melanin content absorps part of the diffusely reflected light that carries the PPG-signal, while the specular reflection is not reduced.

V. Results for exercising subjects

The resulting pulse-rate from our simple peak-detector for the exercising subjects is shown in Figures 8 and 9. For an interval length of 1.6 seconds (32 picture periods), the highest peak in the output spectrum of algorithm $X_s min\alpha Y_s$ corresponds within 3BPM to the measured pulse-rate for more than 98% of the time when exercising on the bike. This compares favorably with the BSS-based algorithms ICA (PCA), which achieve around 54 (79)% provided we allow them with a much longer optimization interval of 25 seconds (512 picture periods).

The stepping sequence turned out to be much more challenging, as expected, and for the best method, $X_s min\alpha Y_s$, the highest spectral peak corresponds to the actual pulse-rate 48% of the time. The performance of the BSS-based methods does not exceed 11%, regardless the length of the interval, possibly because a long interval length allows no fast adaptation, while a short interval length leads to failing component selection. Since these percentages can react sensitively to small variations of the spectrum, Figure 10 shows the achieved signal-to-noise ratio of the pulse signal for all tested methods. This SNR confirms the favorable motion robustness of the chrominance based methods compared to the earlier BSS-based methods, particularly for the stronger motion of the stepping device.

To get a further impression of the quality of the pulse-signals from the individual methods, we show spectrograms for the bike and the stepping exercise in Figure 11. Again

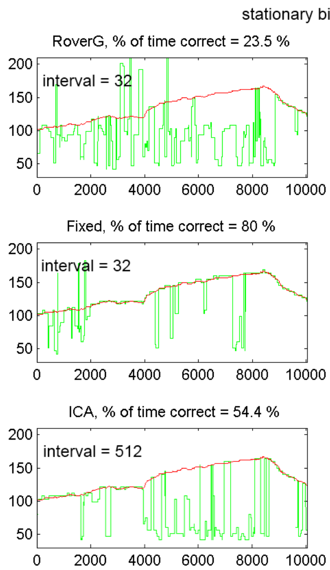

Fig. 8. Momentary pulse-rate (BPM) found with the various methods, for an interval length of 32 or 512 on the bike video, using a simple peak detection in the frequency domain (green line) and the ground-truth obtained from a chest band (red line). The horizontal axis shows the picture-number, 10,000 corresponds to 5 minutes after start (20Hz camera).

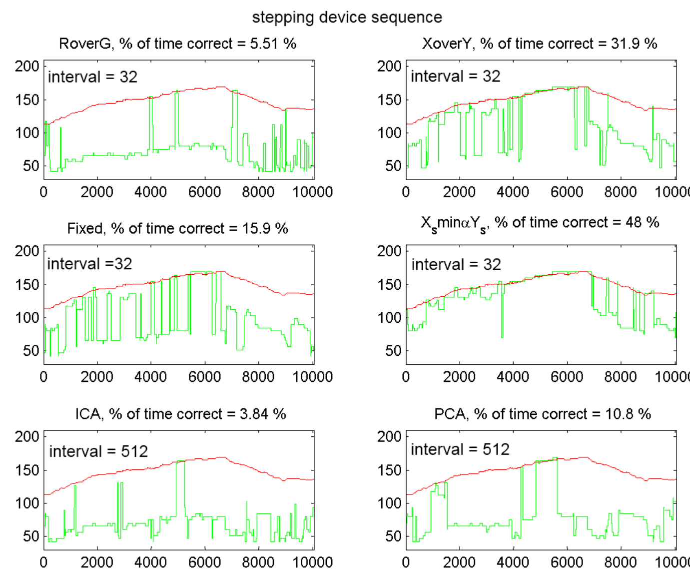

Fig. 9. Momentary pulse-rate (BPM) found with the various methods, for an interval length of 32 or 512 on the stepping video, using a simple peak detection in the frequency domain (green line) and the ground-truth obtained from a chest band (red line). The horizontal axis shows the picture-number, 10,000 corresponds to 5 minutes after start (20Hz camera).

we used the optimal interval length for each algorithm. This optimum was 32 picture periods for all chrominance-based methods and 512 for the BSS-based methods. From our assessment we draw the following conclusions:

- The very basic $RoverG$ method gives poor results, as predicted in our analysis section likely due to normalization errors as specular reflection is not taken into account.
- The methods using chrominance signals, $XoverY$, $Fixed$ and $X_s minaY_s$ give the cleanest spectra with our final design, $X_s min\alpha Y_s$, often showing the pulse-rate as the strongest frequency component.
- The failure of the BSS-based methods had to be expected as all exercise typically causes a periodic motion inside the pulse-rate frequency band. Consequently, these methods cannot reliably detect which component carries the pulse-signal. This is evident already for the stationary bike which exhibits only moderate motion.
- The chrominance-based methods give a clearly superior performance on short overlap-add interval lengths of 1.6 second, while the BSS-based methods perform better on long intervals of 25 seconds. The longer interval length is considered a drawback, as it affects the latency of the method.
- The spectrum of $X_s min\alpha Y_s$ seems cleanest, which indeed leads to a much improved score even with our very basic peak-detection algorithm to establish the pulse-rate. We expect more advanced pulse-rate extraction can further improve this score.

VI. Conclusions

In this paper, we have analyzed remote PPG (rPPG) using a color video camera. We showed why an attempt to achieve motion-robustness with a ratio of two normalized color signals is problematic, due to the unpredictable normalization errors resulting from specular reflections at the skin surface, absent in contact PPG.

This put us on the track of using color difference signals in which this specular reflection component is eliminated, assuming white illumination. Elaborating this track, we derived a number of possible algorithms that separate the blood volume pulse-signal from motion-induced distortions in a deterministic fashion.

We benchmarked these chrominance-based methods with algorithms that have been proposed in earlier research on this topic and are based on blind source separation (BSS) with additional heuristics to select the proper signal from the

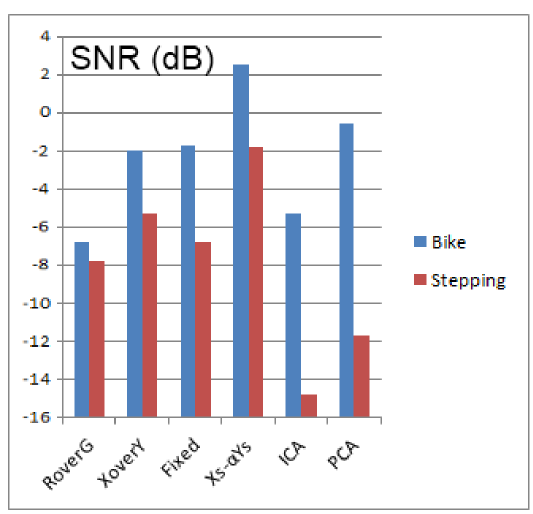

Fig. 10. The signal-to-noise ratio of the pulse signal for the bike and stepping sequence compared. The interval length was selected to be optimal for the individual methods.

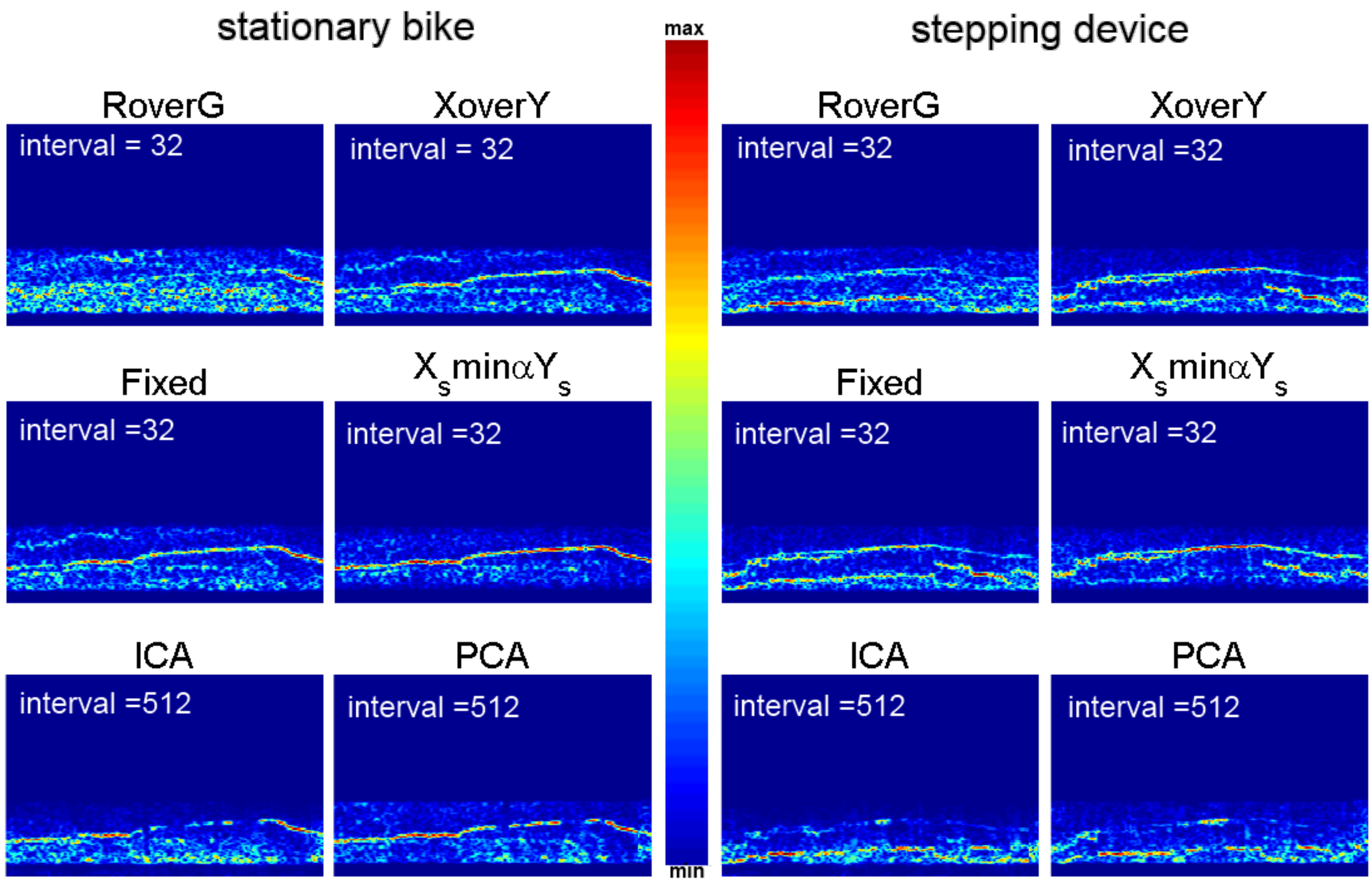

Fig. 11. Spectrograms obtained, for optimal interval length per method, from the various methods for the exercise bike and stepping sequences. The color red indicates the highest amplitude. Time is on the horizontal axis, ranging from 0 to 5 minutes, the vertical axis shows the frequency from 0 to 600 BPM. Note that in the stepping sequence the pulse-rate and a motion frequency are very close and sometimes coincide (top line is the pulse-rate), while often the motion peak has a higher amplitude.

resulting components.

We found an important advantage of our chrominance-based approach is that it eliminates this rather weak component selection heuristic, which can be expected to fail with periodic motion from exercise. As a related advantage, we found that the chrominance-based methods can more quickly adapt to changing conditions and can perform well with shorter latency. We also showed that the inherent drawback resulting from the "white illumination assumption", can be successfully eliminated by a "skin-tone standardization".

To evaluate the proposals, we have analyzed the accuracy of the alternative rPPG measurement techniques over a large population of 117 subjects. We have demonstrated that remote PPG is providing pulse-rates in 92% good agreement ($\pm 1.96\sigma$) with a contact PPG sensor. Pearson's correlation coefficient was very high, $r = 1.00$ for the best chrominance-based rPPG method and $r = 0.97$ for the worst, ICA-based, method. The low error, $RMSE = 0.4$ (1.1) for the best (worst) method, between rPPG and the reference contact-PPG sensor shows interesting prospect for future applications.

Analyzing the measurements obtained for different skin types, we showed that pulse-rate extraction using rPPG can be performed robustly regardless of the skin type, $\sigma = 0.8$ over all skin types for the best chrominance-based method. However, the signal-to-noise energy-ratio of the pulse-signal decreased with the melanin content of the skin from roughly $9.5dB$ for the lightest down to $4.5dB$ for the darkest skin.

In a second experiment, we tested the motion robustness of several methods, moving the camera to the gym and analyzing the pulse-rates obtained from a person exercising on a stationary bike and a stepping device, respectively.

This experiment confirmed the expected problem with the heuristics required by the BSS-based methods, as they could not reliably distinguish the pulse-signal and the periodic motion distortion. This already occurred with quite modest motion, on the stationary bike, where our best chrominance-based method showed the pulse-rate as the strongest spectral peak more than 98% of time, where ICA and PCA scored at best 79%. For more vigorous motion, on the stepping device, our best methods still showed the pulse-rate as the highest spectral peak more than 48% of the time, while ICA and PCA completely failed with scores below 4% and 11%, respectively.

Acknowledgements

The authors are grateful for the help and contributions of their Philips colleagues Ingmar van Dijk, Ihor Kirenko, Herman Kloosterman, Arno van Leest and Wim Verkruijsse. Also, we like to thank our reviewers for their thorough feedback on the manuscript.

REFERENCES

[1] A. B. Hertzman, "Photoelectric plethysmography of the fingers and toes in man", Exp. Biol. Med. 37(3), pp. 529-534 (1937).

[2] M. Huelsbusch and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI", Proc. SPIE 4683, pp. 110-117 (2002).

[3] C. Takano, Y. Ohta, "Heart rate measurement based on a time-lapse image", Medical Engineering & Physics 29 (2007), pp. 853-857.

[4] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light", Opt. Express 16(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. Express 18(10), pp. 10762-10774 (2010).

[6] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring Pulse Rate with a Webcam - a Non-contact Method for Evaluating Cardiac Activity", in Proc. FedCSIS, 2011, pp. 405-410.

[7] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise", Journal of Biomedical Optics 16(7), pp. 077010 (July 2011), doi: 10.1117/1.3602852.

[8] M. Huelsbusch, "Ein bildgestuetstes, funktionelles Verfahren zur optoelektronischer Erfassung der Hautperfusion", PhD-thesis, RWTH Aachen, 28-Jan. 2008, Fakultaet fuer Elektrotechnik un Informationstechnik, pp.70.

[9] J.A Crowe and D. Damianou, "The wavelength dependence of the photoplethysmogram and its implication to pulse oximetry", Engineering in Medicine and Biology Society, 14th Annual International Conference of the IEEE Oct. 29-Nov. 1, 1992, Volume: 6, pp. 2423 2424.

[10] Luis F. Corral, Gonzalo Paez, Marija Strojnik, "Optimal wavelength selection for non-contact reflection photoplethysmography", 22nd Congress of the International Commission for Optics: Light for the Development of the World, edited by Ramn Rodrguez-Vera, Rufino Daz-Uribe, Proc. of SPIE Vol. 8011, 801191.

[11] S. Tominaga, "Dichromatic reflection models for a variety of materials", COLOR research and application, 19, pp. 277-285, 1994.

[12] Y. Kanzava, Y. Kimura, and T. Naito, "Human Skin Detection by Visible and Near-Infrared Imaging", MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, Japan.

[13] M. Thomas, "Large Scale Study on Remote Photoplethysmography", Graduation Report resulting from internship at Philips Research Eindhoven, ISEN-Brest, Oct., 20, 2011.

[14] P. A. Viola, M. J. Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2): 137-154 (2004).

[15] J.-F.Cardoso, High-order contrasts for independent component analysis", Neural Comput., vol. 11, pp. 157-192, 1999.

Gerard de Haan received B.Sc., M.Sc., and Ph.D. degrees from Delft University of Technology in 1977, 1979 and 1992, respectively. He joined Philips Research in 1979 to lead research projects in the area of video processing/analysis. From 1988 till 2007, he has additionally teached post-academic courses for the Philips Centre for Technical Training at various locations in Europe, Asia and the US. In 2000, he was appointed "Fellow" in the Video Processing & Analysis group of Philips Research Eindhoven, and part-time "Full-Professor" at the Eindhoven University of Technology.

He has a particular interest in algorithms for motion estimation, video format conversion, image sequence analysis and computer vision. His work in these areas has resulted in 3 books, 2 book chapters, more than 160 scientific papers and 130 patent applications, and various commercially available ICs. He received 5 Best Paper Awards, the Gilles Holst Award, the IEEE Chester Sall Award, bronze, silver and gold patent medals, while his work on motion received the EISA European Video Innovation Award, and the Wall Street Journal Business Innovation Award.

Gerard de Haan serves in the program committees of various international conferences on image/video processing and analysis, and has been a "Guest-Editor" for special issues of Elsevier, IEEE, and Springer.

Vincent Jeanne received the M.Sc degree in Electrical Engineering from Institut Superieur dElectronique et du Numerique (I.S.E.N), France, in 2006. In 2006, he joined the Video Processing & Analysis group of Philips Research Laboratories. He worked on the development of real time high performance computer vision algorithms, mainly focused on object detection and recognition using machine learning techniques. Since 2008 his focus is on Contactless Vital Signs Monitoring using Video Camera where he has several responsibilities going from developing new algorithms in research projects, transferring research knowledge to development projects as well as leading research projects. His work in these areas has resulted in 10 patent applications, 7 scientific papers and one commercial product: the Philips Vital Signs Camera App.