

Motion robust remote-PPG in infrared

Mark van Gastel, Sander Stuijk and Gerard de Haan

Abstract—Current state-of-the-art remote PPG (rPPG) algorithms are capable of extracting a clean pulse-signal in ambient light conditions using a regular color camera, even when subjects move significantly. In this study, we investigate the feasibility of rPPG in the (near)-infrared spectrum, which broadens the scope of applications for rPPG. Two camera setups are investigated, one setup consisting of three monochrome cameras with different optical filters, and one setup consisting of a single RGB camera with a visible light blocking filter. Simulation results predict the monochrome setup to be more motion robust, but this simulation neglects parallax. To verify this, a challenging benchmark dataset consisting of 30 videos is created with various motion scenarios and skin-tones. Experiments show that both camera setups are capable of accurate pulse-extraction in all motion scenarios, with an average SNR of +6.45 and +7.26 dB respectively. The single camera setup proves to be superior in scenarios involving scaling, likely due to parallax of the multi-camera setup. To further improve motion robustness of the RGB camera, dedicated LED-illumination with two distinct wavelengths is proposed and verified. This paper demonstrates that accurate rPPG measurements in infrared are feasible, even with severe subject motion.

Index Terms—Infrared, remote photoplethysmography, vital signs monitoring

I. INTRODUCTION

THE (cardiac) pulse-signal is one of the most important physiological signals used by medical professionals for the diagnosis and tracking of the patient’s medical condition. Photoplethysmography (PPG) is a low-cost optical technique for detecting arterial pulsations non-invasively. The technique was first described by Hertzman [1] in the 1930s. PPG is based on the principle that blood volume changes vary the optical density of the skin over a vascular region, because of the differences in light absorption between blood and the surrounding tissue. Nowadays, PPG is applied ubiquitously in hospital settings, where a contact sensor is typically attached to a finger/toe/ear, or patched to the skin [2]. The contact sensor comprises a light source emitting light to the skin surface, and a photo-detector capturing the light reflected or transmitted from/through the skin. The PPG waveform consists of a pulsatile component, superimposed on a slowly varying component. The pulsatile component shows changes in the blood volume that occur between the systolic and diastolic phases of the cardiac cycle. The slowly varying component of the PPG waveform corresponds to the detected transmitted or reflected optical signal from the tissue, and

depends on the structure of the tissue and the average blood volume of both arterial and venous blood.

Over the last decade, it has been shown that blood volume variations can also be measured at a distance, leading to remote PPG (rPPG). This is highly attractive for cases where direct contact with the skin has to be prevented (e.g. neonates, subjects with skin-damage) or unobtrusiveness is desired (e.g. surveillance, fitness). Humphreys *et al.* [3] was able to successfully extract the PPG signal with a monochrome CMOS camera placed at 40 cm from the finger, and a LED with a wavelength of 800 nm as illumination. Verkrusse *et al.* [4] proved the feasibility of rPPG under ambient light conditions using a regular color camera. They observed that the green color channel features the strongest plethysmographic signal, corresponding to an absorption peak of oxy-hemoglobin.

More recently, rPPG methods using multiple wavelengths have been proposed [5]–[7]. The motivation is to improve the robustness to subject motion, which is the main concern with (r)PPG. With a single wavelength, no distinction between pulse-induced intensity variations and variations caused by motion exist. Multiple channels with different mixtures of the pulse-induced intensity variations allows to distinguish between both. Hülsbusch separated the noise and the PPG signal into two independent signals built as a linear combination of the DC-normalized red and green color channel [8]. The energy in the pulse-signal was minimized as an optimization criterion. Poh *et al.* [6] and Lewandowska *et al.* [7] proposed to construct the pulse-signal as a linear combination of all three normalized color channels. To find this linear combination, they employed blind-source-separation (BSS) techniques, ICA and PCA respectively. Since it is a priori unknown which of the components comprises the pulse-signal, the periodic nature of the pulse-signal is used for component selection in both methods. However, this heuristic selection criterion fails when strong periodic subject movements are present, e.g. in a fitness setting.

De Haan *et al.* [9] eliminated the component selection criterion by constructing a linear combination of the color channels orthogonal to the main distortions: intensity-variation and specular reflection, assuming a standardized normalized skin-color. This chrominance-based method outperforms the BSS methods for videos recorded in a gym with subjects exercising with vigorous motion.

Instead of making assumptions about the distortions and skin-tone or the periodicity of the pulse-signal, De Haan *et al.* [10] proposed to use the unique signature of the blood volume pulse to extract the pulse-signal. This ‘signature’ is derived

M. van Gastel and S. Stuijk are with the Electronic Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: (m.j.h.v.gastel@tue.nl, s.stuijk@tue.nl).

G. de Haan is with the Philips Innovation Group, Philips Research, Eindhoven, The Netherlands, and with the Electronic Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: (g.de.haan@philips.com).

Manuscript received November 04, 2014; revised December 10, 2014; accepted January 06, 2015

from physiology and optics, and used to create a method, the ‘‘PBV-method’’, to extract the pulse-signal. Essentially, the PBV-method suppresses all variations not aligned with the signature of the blood volume pulse. Experimental results show a large improvement in motion robustness compared to earlier methods, and we therefore recognize the PBV-method as the current state-of-the-art method for motion robust remote pulse-extraction.

In this paper, we aim at further extending the application range of the PBV-method by investigation the feasibility of motion-robust rPPG in the near-infrared (NIR) part of the light spectrum, enabling us to extract the pulse-signal in full darkness. This is not trivial since the relative PPG amplitude is significantly reduced in NIR compared to visible light. Because also the spectral response of the camera has decreased in NIR, safety regulations regarding the maximum radiation levels of the illuminant have to be observed to not cause any risk to the subject, e.g. eye damage.

First, the optimal wavelengths for motion-robust rPPG in NIR are investigated, using three monochrome cameras with different optical filters. Predictions for the blood volume pulse vector are made, which are consequently verified by large scale experiments. Next, the feasibility of using a single camera in NIR is investigated, by replacing the IR blocking filter of a regular RGB camera with a visible light blocking filter. To further improve motion robustness, spectrally optimized LED-illumination is proposed and experimentally verified. In Section II-D the framework of the pulse-extraction algorithm is presented. The proposed methods are evaluated on a challenging dataset consisting of 30 videos of subjects with different skin-tones which perform various motion scenarios, as described in Section II-E. The results are presented in Section III, with a discussion on the results in Section IV. Finally, in Section V, our main conclusions are drawn.

II. MATERIALS AND METHODS

To investigate the feasibility of robust rPPG in IR, current state-of-the-art algorithms in visible light are employed, which are adapted for operating in IR. A modified version of the framework of Wang *et al.* [11] with multiple parallel rPPG sensors, is combined with the PBV pulse-extraction method of De Haan *et al.* [10]. The next subsection provides a more detailed description of the PBV-method to clarify the design considerations and simulations in the subsequent subsections.

A. PBV method

De Haan *et al.* showed that the minute optical absorption changes caused by blood volume variations in the skin occur along a very specific vector in a normalized RGB-space [10]. This unique blood volume ‘signature’ enables robust rPPG pulse-extraction that minimizes the contribution to the pulse-signal of color variations with other signatures. Compared to the motion robust chrominance-based pulse-extraction method by De Haan *et al.* [9], no assumptions about the distortion signals have to be made. Instead, the known ratios of the relative PPG amplitudes in the normalized color channels, \vec{P}_{bv} ,

are employed to discriminate between the pulse-signal and distortions. The relative PPG amplitudes in the normalized color channels are defined as: $\sigma(\vec{C}_{(in)})$, where $i \in \{R, G, B\}$ and σ denotes the standard deviation. The color channels are normalized by: $\vec{C}_{(in)} = \frac{1}{\mu(\vec{C}_{(i)})} \vec{C}_{(i)} - 1$, where μ corresponds to the (temporal) mean value. More details about why the pulse vector is known, are provided in the continuation of this subsection.

We assume that the pulse-signal \vec{S} can be constructed as a linear combination of the three normalized color channels:

$$\vec{S} = \vec{W} \mathbf{C}_N, \quad (1)$$

where \vec{W} , dimensions 3×3 , is the weighting matrix with $\vec{W} \vec{W}^T = 1$, and \mathbf{C}_N has dimensions $3 \times N$, where N indicates the number of samples in the time-window.

Since the ratios of the relative PPG amplitudes in the color channels of the camera are known, the aim is to find the weights, \vec{W} , that construct the pulse-signal \vec{S} , for which the correlation with the normalized color channels equals \vec{P}_{bv} :

$$\vec{S} \mathbf{C}_N^T = k \vec{P}_{bv} \Leftrightarrow \vec{W} \mathbf{P}_{BV} \mathbf{C}_N \mathbf{C}_N^T = k \vec{P}_{bv}, \quad (2)$$

and therefore the weights \vec{W}_{PBV} can be calculated using:

$$\vec{W}_{PBV} = k \vec{P}_{bv} \mathbf{Q}^{-1} \quad \text{with} \quad \mathbf{Q} = \mathbf{C}_N \mathbf{C}_N^T, \quad (3)$$

where scalar k is chosen to assure that \vec{W}_{PBV} has unit length. To employ the PBV-method and extract the pulse-signal, the ratios of the relative PPG amplitudes of the normalized channels compiled in the normalized blood volume pulse vector \vec{P}_{bv} have to be known.

Let us summarize the prediction of the pulse vector from physiology and optics following [10]. The *relative* PPG amplitude as function of the wavelength λ , $\sigma(PPG(\lambda))/\mu(PPG(\lambda))$, has been *modeled* by Hulsbusch [8]. Corral [12] *measured* the *absolute* PPG spectrum, $\sigma(PPG(\lambda))$, using a tungsten-halogen lamp as illumination which emits radiation in the visible and NIR section of the light spectrum. The relative PPG (*RPPG*) can be related to the absolute PPG by:

$$PPG(\lambda) = \rho_s(\lambda) I_h(\lambda) RPPG(\lambda), \quad (4)$$

since the light-source and the skin determine the baseline component of the absolute PPG spectrum. Here $\rho_s(\lambda)$ and $I_h(\lambda)$ represent the skin reflectance spectrum and the emission spectrum of the tungsten-halogen illumination, respectively. The PPG spectra of Hulsbusch and Corral are displayed in Figure 1, the skin reflectance spectra and the light spectra in Figure 2. The ratios of the relative PPG amplitudes in the three channels of a camera, described by the blood volume pulse vector \vec{P}_{bv} , can be predicted by:

$$\vec{P}_{bv} = \begin{bmatrix} \int_{\lambda=400}^{1000} H_{C1}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} PPG(\lambda) d\lambda \\ \int_{\lambda=400}^{1000} H_{C1}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} \rho_s(\lambda) d\lambda \\ \int_{\lambda=400}^{1000} H_{C2}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} PPG(\lambda) d\lambda \\ \int_{\lambda=400}^{1000} H_{C2}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} \rho_s(\lambda) d\lambda \\ \int_{\lambda=400}^{1000} H_{C3}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} PPG(\lambda) d\lambda \\ \int_{\lambda=400}^{1000} H_{C3}(\lambda) \frac{I(\lambda)}{I_h(\lambda)} \rho_s(\lambda) d\lambda \end{bmatrix}. \quad (5)$$

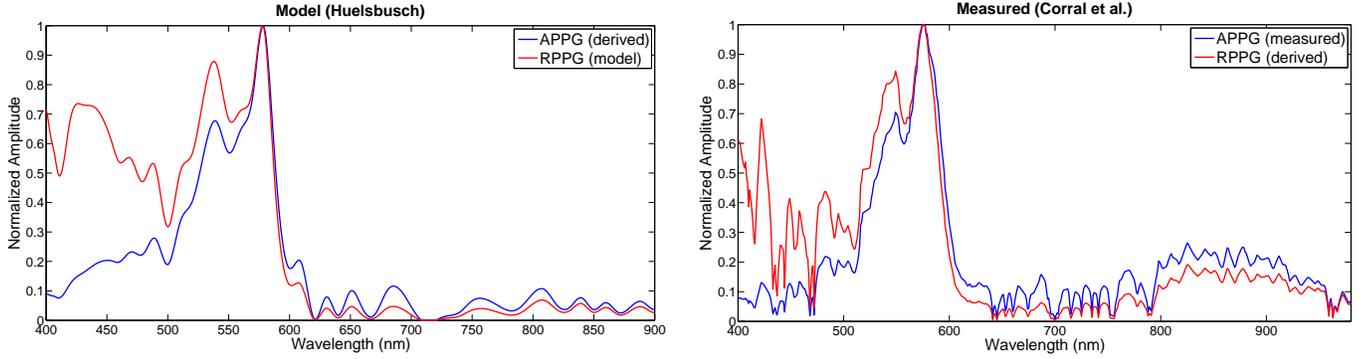


Fig. 1: a) The modeled relative PPG spectrum by Hülsbusch [8] and the derived absolute PPG spectrum, b) The measured absolute PPG spectrum of Corral [12] and the derived relative PPG spectrum. All spectra have been scaled to 1 for their peak locations.

Here $H_{C1,C2,C3}$ are the responses of the three channels respectively and $I(\lambda)$ is the spectrum of the illuminant.

Kanzawa *et al.* [13] measured the skin reflectance of 50 subjects in the visible and NIR section of the light spectrum. The subjects in [13] have different skin-melanin concentrations and their skin-color is classified into three categories; “bright”, “mongoloid” and “dark”, which we have interpreted as skin-types II, III and V, according to Fitzpatrick’s scale [14], and we shall use this interpretation in the continuation of this paper.

Since motion affects all color channels equally under uniform white illumination, the normalized vector describing the motion induced color variations is: [0.58, 0.58, 0.58], further referred to as ‘motion vector’ (not to be confused with a vector describing displacement). For motion robustness, the inner-product between this motion vector and the pulse vector \vec{P}_{bv} should be small to be able to discriminate between signals in the direction of the pulse vector and signals which have a different orientation.

Predictions of the pulse vector in visible light with an RGB camera performed by De Haan [10], showed that the PPG amplitude spectrum of Hülsbusch provides more accurate predictions compared to Corral’s spectrum. The simulated pulse vector using Hülsbusch’s deviated only 4° from the measurements, while simulations using Corral’s curve were 7° off. We repeat the predictions in NIR with both spectra for the two camera setups, which are later compared to the measured pulse vectors to verify which spectrum provides the most accurate predictions. This PPG spectrum is subsequently employed for the simulations of the dedicated LED-illumination described in Section II-C.

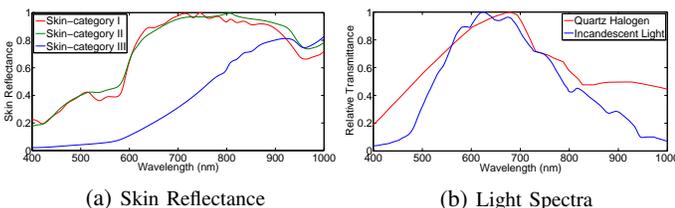


Fig. 2: a) Skin reflectance spectra of three skin-categories measured by Kanzawa [13], b) Transmittance spectra of a tungsten halogen lamp and the incandescent light bulbs in the experimental setup.

In order to compare the predicted results with the measured pulse vectors, a large dataset comprising recordings of 40 participants with skin-pigmentation concentration levels ranging from 45 to 600 (on a scale of values from 0-999), is created. The participants are asked to sit still with their head fixed by a head-rest to prevent motion affecting our measurements. Informed consent is obtained for each participant prior to the recordings. After recording, a rectangular bounding-box is manually annotated and tracked by the CSK algorithm of Henriques *et al.* [15]. The spatial means of all pixels within the ROI are calculated for every frame. By concatenating these values, traces for every camera channel in the setup are constructed.

To acquire \vec{P}_{bv} , the ratios of the relative PPG amplitudes in the channels have to be measured. First, the traces of the camera channels are mean-centered normalized within a time-window by:

$$\vec{C}_{(i)n} = \frac{1}{\mu(\vec{C}_{(i)})} \vec{C}_{(i)} - 1, \text{ for } i = 1, 2, 3, \quad (6)$$

where the vectors have length 64 and $\mu(\vec{C}_{(i)})$ corresponds to the (temporal) mean of the vector. Next, the normalized channel traces are band-pass filtered, [0.6-3] Hz, to eliminate noise. A pulse-signal is constructed by performing PCA on the filtered channel traces, where potential involuntary motion and noise present in the traces are separated from the pulse-signal. By using an overlap-add procedure with a Hanning window on the time-windowed traces of 64 samples, traces for the entire recording time of 120 seconds are constructed. Finally, \vec{P}_{bv} is obtained after normalization of the inner-products between the constructed pulse-signal and the three channels.

B. Monochrome Cameras

By employing appropriate optical filters, desired light wavelengths can be exposed to the sensor of a monochrome camera where other wavelengths are blocked. Since the goal is to attain motion robustness, filters have to be selected such that the inner-product between \vec{P}_{bv} and the motion vector is minimal. This criterion is analogous to maximizing the angle between both vectors, with a maximum of 90° corresponding to an inner-product of zero. To determine which filter

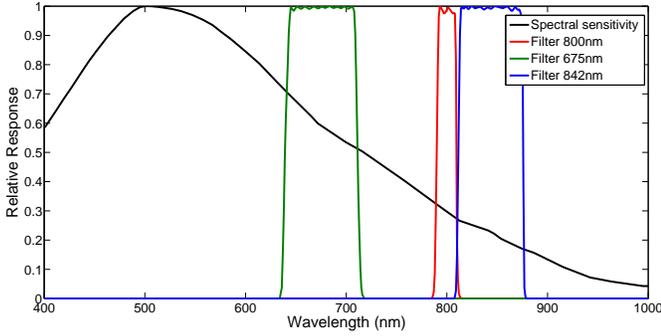


Fig. 3: Spectral response of the monochrome CCD camera, type Marlin F046B, with the filter characteristics of the three applied optical filters of Semrock Inc.

TABLE I: Pulse vector simulations for the setup with the three monochrome cameras.

Hülbsbusch				
Skin-type	PBV ₈₀₀	PBV ₆₇₅	PBV ₈₄₂	Angle (degrees)
II	0.69	0.57	0.44	10.1
III	0.69	0.57	0.44	10.1
V	0.69	0.57	0.44	10.2
Average	0.69	0.57	0.44	10.1

Corral				
Skin-type	PBV ₈₀₀	PBV ₆₇₅	PBV ₈₄₂	Angle (degrees)
II	0.55	0.30	0.78	20.2
III	0.55	0.31	0.78	19.3
V	0.55	0.33	0.77	17.8
Average	0.55	0.31	0.78	19.1

combination provides the best motion robustness, simulations with all possible filter combinations are performed, assuming a pass-band of 50 nm for all three filters. The simulation results show that an angle of more than 23° between \vec{P}_{bv} and the motion vector can be achieved when a combination of filters with center wavelengths of 670, 750 and 830 nm is employed. Limited by the available optical filters in the laboratory, our best approximation lead us to use filters with center wavelengths of 675, 800 and 842 nm, as illustrated in Figure 3.

As described in the previous subsection, the blood volume pulse vector \vec{P}_{bv} can be predicted by Equation (5), where $H_{C(1,2,3)}(\lambda)$ is here corresponding to the product of the spectral response of the camera and the response of the applied optical filter. In Table I, the predicted pulse vectors for the monochrome cameras are displayed together with the angles between the pulse and motion vectors, assuming incandescent illumination.

An overview of the measured pulse vectors for the setup with the monochrome cameras is displayed in Figure 4. The results show that pulse vectors are quite stable over the entire range of skin pigmentation levels. This was also to be expected, because the skin reflectance spectrum is largely uniform in the NIR section of the light spectrum [13]. The average pulse vector for the three monochrome cameras with optical filters of 800, 675 and 842 nm respectively, is:

$$\vec{P}_{bv} = [0.61, 0.29, 0.74], \quad (7)$$

which has an angle of 19.0° with respect to the motion vector.

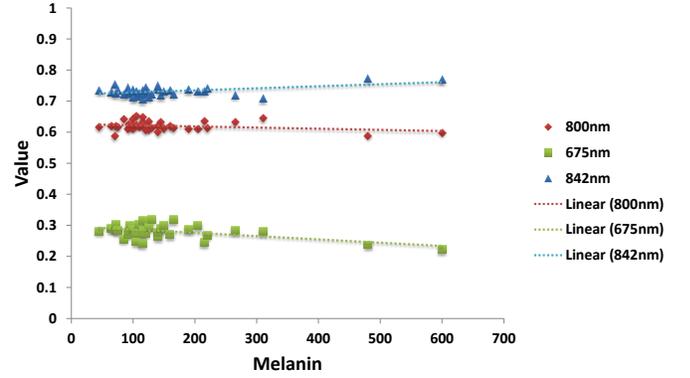


Fig. 4: Pulse vector estimation for the three monochrome cameras in NIR.

It can be seen that Corral's PPG spectrum provides a more accurate prediction of the blood volume pulse vector compared to Hülbsbusch's PPG spectrum. Pulse vector simulations using Corral's curve are 4° off, while simulations using Hülbsbusch's curve are 24° off when compared to the measured pulse vector (7). This large discrepancy may be caused by Hülbsbusch focussing primarily on visible wavelengths by modelling the PPG spectrum, while Corral *et al.* actually measured the PPG amplitude for wavelengths up-to 980 nm.

C. RGB Camera

An RGB camera samples the visible light spectrum [400-700] nm, using a Bayer pattern to achieve color selectivity [16]. Wavelengths longer than 700 nm are blocked by an IR-blocking filter. As explored by De Haan *et al.* [10], the blood volume pulse vector for an RGB camera has an angle of approximately 19° with respect to the motion vector in visible light.

Since the sensor of an RGB camera is sensitive for wavelengths in IR, typically up-to 1000 nm, rPPG with a color camera seems possible also in NIR. The most obvious possibility to apply it in IR, is to replace the Bayer color field array (CFA) for a CFA which samples the light spectrum for wavelengths in NIR, [700-1000] nm. However, this operation is rather expensive and difficult to realise.

We considered it more interesting to just use a regular RGB

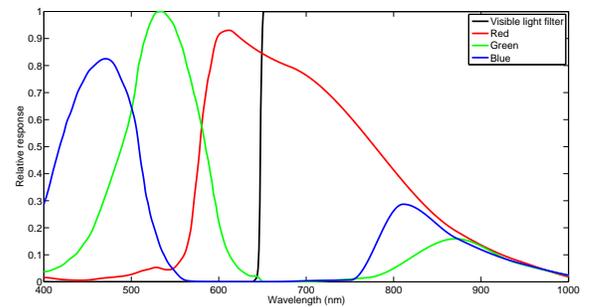


Fig. 5: Measured spectral response of the Marlin F046C CCD color camera and the filter response of the visible light blocking filter attached to the camera.

TABLE II: Pulse vector simulations for the setup with the RGB camera.

Hülsbusch				
Skin-type	PBV _R	PBV _G	PBV _B	Angle (degrees)
II	0.62	0.50	0.61	5.2
III	0.62	0.50	0.61	5.2
V	0.62	0.50	0.61	5.3
Average	0.62	0.50	0.61	5.2

Corral				
Skin-type	PBV _R	PBV _G	PBV _B	Angle (degrees)
II	0.38	0.67	0.63	12.8
III	0.40	0.67	0.63	12.1
V	0.41	0.66	0.63	11.0
Average	0.40	0.67	0.63	11.9

camera with a visible light blocking filter replacing the IR blocking filter. Although not clearly specified in most camera specifications, the color channels of an RGB camera have responses for NIR wavelengths. However, it is ambiguous how auspicious the blood volume vector for this configuration is in terms of motion robustness. To predict the performance of an RGB camera in IR, the spectral response of the RGB camera in the wavelength range from 400 to 900 nm is measured using the Lambda 800 photospectrometer from PerkinElmer[®], with a spectral resolution of 10 nm. The spectral response of the Marlin F046C CCD camera of Allied Vision Technologies GmbH is visualized in Figure 5.

As described in subsection II-A, the blood volume pulse vector \vec{P}_{bv} can be predicted by Equation (5), where $H_{C(1,2,3)}(\lambda)$ is here corresponding to the product of the spectral response of the camera color channel and the response of the applied visible light blocking filter. In Table II, the predicted pulse vectors for the RGB camera are displayed.

An overview of the measured pulse vectors for the RGB camera is displayed in Figure 6. Similar to the results with the monochrome cameras, the pulse vector is quite stable over the entire range of skin-pigmentation levels.

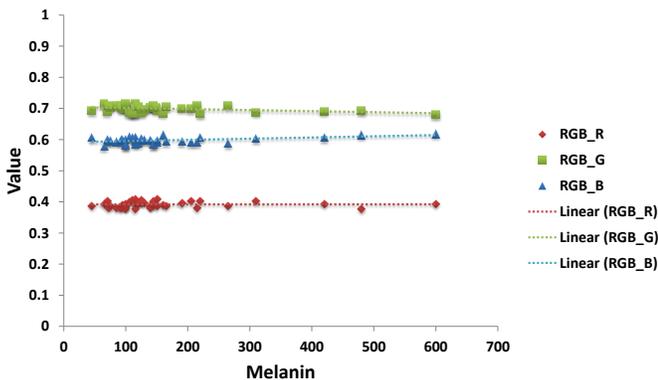


Fig. 6: Pulse vector estimation for RGB camera in NIR.

The average pulse vector for the RGB camera is:

$$\vec{P}_{bv} = [0.39, 0.70, 0.60], \quad (8)$$

which has an angle of 12.7° with respect to the motion vector.

Similar to the monochrome cameras setup, Corral's curve provides the most accurate simulation results. For the RGB

camera, predictions using Corral's curve are 2° off, where predictions using Hülsbusch's curve are 17° off.

Hardware Improvements

Since the pulse vector is influenced by the light spectrum, it may be possible to improve motion robustness by selecting a specific light spectrum. This can be achieved by employing dedicated LEDs, whose emission spectra are more bandlimited compared to incandescent light bulbs. To verify which combination of LEDs yields the best motion robustness, simulations are performed for all wavelengths passed by the visible light blocking filter.

The results of simulations for the RGB camera are visualized in Figure 7. Here, the center wavelengths are on the horizontal axis, and the power ratio between the LEDs are on the vertical axis. For the simulations, the absolute PPG spectrum of Corral *et al.* is employed. The emission spectra of the LEDs are modeled by a Gaussian distribution function. Simulation results show that the angle between the pulse and the motion vector can be increased from 12.7° (obtained with incandescent light) to more than 18°. As can be observed from the simulations, a combination of LEDs with center wavelengths of 660 and 940 nm results in this favorable pulse vector. To verify that the simulated pulse vector is corresponding with the actual pulse vector, a LED illumination unit with the optimal wavelength combination is constructed, which is described in Section II-E. For fair comparison, the \vec{P}_{bv} prediction is repeated with the actual emission spectra of the LEDs in the illumination unit, instead of the model used to determine the optimal wavelength combination. The predicted pulse vector has an angle of 18°, close to the measured pulse vector which has an angle of more than 17°.

As the optimal LED combination requires a 660 nm LED, the illumination is still visible for the human eye. In order to demonstrate the capability of the RGB camera in full darkness, $\lambda > 700$ nm, the 660 nm LED is replaced by a LED with a center wavelength of 760 nm. Again, predictions for the pulse vector with the actual emission spectra of the LEDs

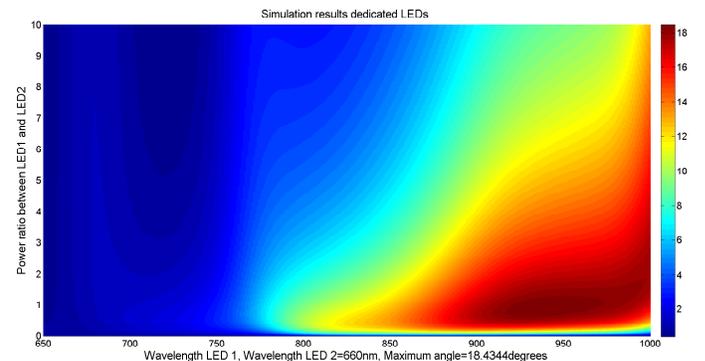


Fig. 7: Simulations RGB camera using dedicated LEDs. By selecting LEDs with center wavelengths of 660 and 940 nm, an angle of more than 18 degrees with respect to the motion vector can be achieved, an improvement of more than 5 degrees compared to incandescent light.

are performed, which show an angle of more than 13° with respect to the motion vector, similar to incandescent light. The predicted pulse vector is validated by the measured pulse vector obtained from recordings, which shows a deviation of less than 1° . These results confirm the idea that by selecting specific light spectra with LEDs, enhanced motion robustness for an RGB camera can be achieved and a considerable angle is achievable even in full darkness.

D. System Framework

Let us now introduce the system framework, visualized in Figure 8. The framework is a modified version of the framework proposed by Wang *et al.* [11]. In this paper, we will address the differences between our framework and the framework proposed by [11].

Image Registration: When multiple cameras are employed to capture the scene, image registration is required to align the frames of the cameras. In order to register the three monochrome channels, a 2D affine transformation involving translation, rotation, scaling and shearing is employed. The transformation matrix, \mathbf{M} , is determined based on the first 100 frames of the recording and applied to the entire duration of the recording. This transformation can be written as:

$$\mathbf{p}' = \mathbf{M}\mathbf{p} \Leftrightarrow \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (9)$$

where $\begin{bmatrix} x \\ y \end{bmatrix}$ are the original pixels locations, and $\begin{bmatrix} x' \\ y' \end{bmatrix}$ are the pixel locations after the 2D affine transformation. The elements (t_x, t_y) in the transformation matrix indicate translation, and (a, b, c, d) indicate the product of rotation, scaling and shearing operations.

Skin Classification: Skin classification is performed with an one-class SVM classifier. Instead of using the intensity normalized RGB and YCrCb as feature descriptors as proposed by Wang *et al.* [11], the plain pixel values of the three channels are adopted as feature description since they yield better results in NIR.

Pulse Extraction: After skin classification, pulse-extraction is performed for each of the multiple signals using the PBV-method of De Haan *et al.* [10]. This method requires a blood volume pulse vector, \vec{P}_{bv} , which is obtained from the large scale experiments as described earlier. Wang *et al.* [11] employed the chrominance-based method [9] instead of the PBV method for their experiments in visible light conditions.

E. Experiments

In this section the performed experiments are described, together with a description of the adopted evaluation metrics and details about the implementation of the algorithm.

1) *Experimental Setup:* A schematic representation of the experimental setup is visualized in Figure 9. Participants are asked to sit on a chair, looking into the cameras in front of them, which are placed at a distance of 2.5 meters

from the head. The four cameras, three monochrome and one RGB, record the scene simultaneously and their data is transmitted over a FireWire connection to an acquisition PC with LabView, where uncompressed video-data is stored. The three monochrome cameras in the setup are the Marlin F046B and the color camera is the Marlin F046C, all of Allied Vision Technologies GmbH. All cameras have 25 mm lenses, a frame rate of 15 fps, a resolution of 640x480 pixels and 8 bits depth. Before recording, the cameras are focussed and manually aligned. For reference, a Philips IntelliVue X2 patient monitor with pulse oximetry finger probe is attached to the subject, where the reference pulse-signal is transmitted to the connected acquisition PC.

Two light units consisting of incandescent light bulbs with diffusers are placed at both sides of the chair at a distance of 1 meter. For the experiments with dedicated LEDs, the units are placed at a distance of 60 cm. A diffuser is placed at a distance of 2 cm from the LEDs to attain a homogeneous light spectrum. We verified with a radiometer (type: LT1700 of International Light Technologies) that even at maximum intensity the LED units remain a factor of 20 below the irradiance safety limit in NIR.

2) *Benchmark Dataset:* To the best of our knowledge, no dataset of rPPG recordings with ground truth data is available. Therefore, we created our own benchmark dataset with incandescent illumination to compare both camera setups. Incandescent light is employed to prevent that potential inhomogeneities in the light spectrum influence the results. The 6 participants in the videos range in age between 22 and 30 years. The study is approved by the Internal Committee Biomedical Experiments of Philips Research, and the informed consent is obtained from each subject. The melanin content of the skin is measured with a skin pigmentation analyzer (model: Mexameter[®] MX 18 MDD of Khazaka electronic GmbH), with the measurement-probe located at the backside of the fore-arm. The melanin indices are loosely linked to three skin-types (II,III,V), according to the Fitzpatrick scale [14]. A reference sensor is attached to a finger of the participant and connected to the acquisition PC.

To evaluate the motion robustness of the algorithm, five different motion scenarios are performed by the subjects: stationary, translation, rotation, scaling and mixed. For the stationary scenario, the head of the subject is fixed in a head rest and he/she is asked to remain stationary during the recording. The translation motion scenario consists of repetitive horizontal and vertical head translations, where repetitive horizontal and vertical head rotations are performed for the rotation motion scenario. The scaling motion scenario consists of repetitive head movements to and from the cameras, where in the mixed motion scenario all previously described motions are executed randomly. The length of each recording is 120 seconds, where the recording starts 1 minute after the participant entered the setup to ensure a stable heart rate.

3) *Evaluation Metrics:* This study adopts the performance metrics used in [9], SNR and PERC. For a detailed description of both metrics, we refer to [9]. Additionally, the correspondence between the pulse-rate extracted from the rPPG pulse-signal and the pulse-rate extracted from the reference

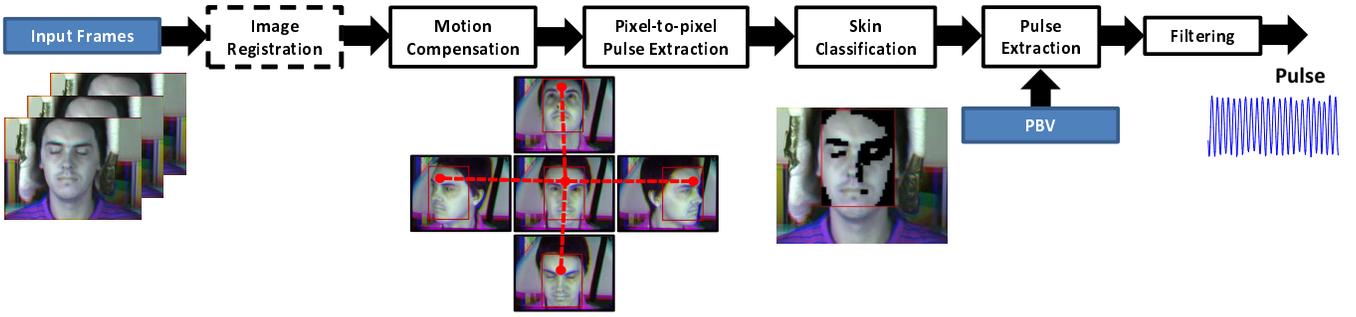


Fig. 8: System framework for motion robust pulse-extraction in NIR. Image registration is only required for the multi-camera setup.

sensor is evaluated. The discrepancy is expressed in the mean-absolute-error (MAE) and the root-mean-squared-error metrics (RMSE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |PR(i) - PR_{ref}(i)|, \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (PR(i) - PR_{ref}(i))^2}, \quad (11)$$

where PR and PR_{ref} are obtained by using a peak-detector in the frequency domain using a sliding Fourier window. All four metrics use a window of 150 samples (10 seconds) to allow for a varying pulse-rate.

Furthermore, correlation and Bland-Altman plots are included to show the agreements of the instantaneous pulse-rate between the rPPG and reference PPG-sensor. Finally, Analysis of Variance (ANOVA) is applied on SNRa (average SNR) values to analyse the significance of difference between methods under certain categories (i.e. camera setups and skin-types).

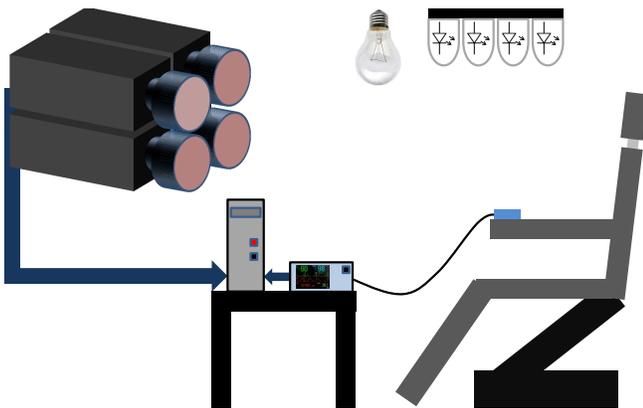


Fig. 9: Schematic overview of the experimental setup. The subjects are seated on an adjustable chair. Four cameras, three monochrome and one color, capture the scene simultaneously and transmit their data to the acquisition PC over FireWire, where video-data is stored uncompressed. For the experiments with incandescent light, illumination units are placed at a distance of 1 meter, where for the experiments with LED illumination the illumination units are placed at a distance of 60 cm. A reference pulse-signal is acquired by the pulse oximetry finger probe, which is connected to a patient monitor.

4) *Implementation:* The proposed algorithm is implemented in Java using the OpenCV 2.4.8 library and executed on a laptop with a Intel Core i5 2.60 Ghz processor and 8 GB RAM. In the first frame, a rectangular ROI indicating the face is initialized manually. For fair comparison, all system parameters are identical for the evaluation of the entire dataset. The values of the evaluation metrics are calculated offline using Matlab.

III. RESULTS

The results of the two camera setups gained on the benchmark video sequences are summarised in Table III. The correlation and Bland-Altman plots of both setups are displayed in Figure 10. Plots of the performed one-way ANOVA are visualized in Figure 11. As observed by Wang *et al.* [11] and confirmed by the results of this dataset, gender is not the key factor which needs to be investigated in the dataset, since the differences between male and female from the same skin-type are rather small. Therefore, the results are averaged over both genders.

A. Stationary scenario

The results show that both camera setups perform similar in scenarios without subject motion. Although a slightly higher SNRa is achieved by the monochrome cameras, 11.2 dB versus 10.1 dB for the RGB camera, the differences between the extracted pulse rates and corresponding metrics are negligible. The correct pulse rate is extracted for almost 99.8 percent of the duration of the recordings.

B. Motion scenarios

In videos with head motions, both camera setups perform worse compared to the scenario without head motion. However, even for the most challenging mixed motion scenario, worst-case, PERC is still 63%, with an SNRa of 2.3 dB. In general, the RGB camera performs better in motion scenarios, where the difference is most prominent in scenarios involving scaling. Overall, no significant difference, p -value = 0.32 (> 0.05), in performance is observed between both camera setups.

C. Differences between skin-types

For both camera setups, no significant difference, p -value = 0.17 (> 0.05), in performance is observed between the three skin-types.

TABLE III: Results for both camera setups gained on benchmark video sequences (averaged over genders).

Video	Monochrome Cameras				RGB Camera			
	SNRa (dB)	PERC (%)	MAE (BPM)	RMSE (BPM)	SNRa (dB)	PERC (%)	MAE (BPM)	RMSE (BPM)
Skin-type II Stationary	11.8	100	0.22	0.68	9.38	100	0.26	0.69
Skin-type II Translation	6.77	94.0	1.00	2.05	9.17	93.9	0.71	1.72
Skin-type II Rotation	3.89	78.1	1.71	2.82	4.31	75.7	2.22	3.86
Skin-type II Scaling	3.48	73.7	3.22	5.00	4.79	81.6	1.82	3.24
Skin-type II Mixed	2.27	63.1	3.08	4.52	2.25	67.7	2.66	4.13
Skin-type III Stationary	10.4	99.3	0.52	1.43	9.72	99.3	0.53	1.18
Skin-type III Translation	6.46	86.6	1.83	2.53	7.87	89.4	1.23	1.83
Skin-type III Rotation	5.29	87.3	1.68	2.58	7.96	92.3	1.62	2.43
Skin-type III Scaling	5.03	84.9	1.33	2.52	6.87	88.0	1.24	2.23
Skin-type III Mixed	5.81	84.6	2.05	2.73	5.97	82.4	1.62	2.77
Skin-type V Stationary	11.4	100	0.19	0.52	11.1	100	0.21	0.53
Skin-type V Translation	8.03	96.8	0.47	1.02	9.01	95.4	0.32	0.73
Skin-type V Rotation	8.07	97.5	0.36	0.89	7.90	94.1	0.67	1.28
Skin-type V Scaling	3.60	73.8	2.68	4.90	6.48	83.8	1.24	2.28
Skin-type V Mixed	4.37	78.7	2.63	4.25	6.20	88.6	1.64	3.02
Average	6.45	86.6	1.53	2.56	7.26	88.8	1.20	2.13

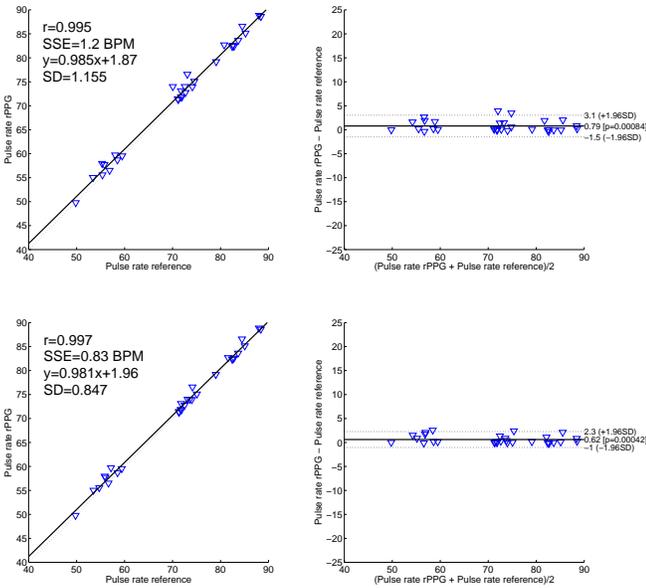


Fig. 10: Correlation and Bland-Altman plots for the setup with the monochrome cameras (top) and RGB camera (bottom). Here r indicates the Pearson correlation, SSE the sum of squared errors, y the linear fit and SD the standard deviation.

IV. DISCUSSION

Both camera setups show the feasibility of rPPG in NIR, regardless of the reduced PPG amplitude compared to visible

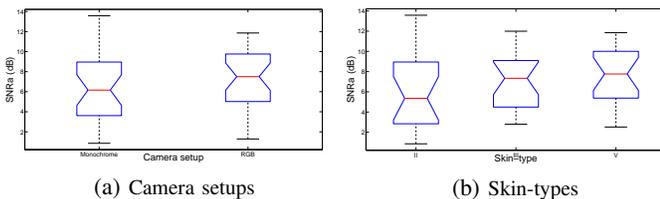


Fig. 11: Statistical comparison using ANOVA. The plots display the median (red bar), standard deviation (blue box), minimum and maximum (black bar) of the SNRa values.

light. Despite the more advantageous angle between the pulse vector and the motion vector in the monochrome cameras setup, the RGB camera shows comparable performances for challenging motion scenarios. For recordings containing scaling movements, the RGB camera even slightly outperforms the monochrome cameras, although the difference is not significant (p -value = 0.22 (> 0.05)). This can be explained by the effects of parallax in the monochrome cameras setup. Here, the cameras are registered by an affine transformation which is determined based on the first 100 frames of the recording. When head motions to and from the cameras are performed, the transformation becomes inaccurate and the performance drops.

Based on measurements on 40 subjects, the pulse vector showed to be quite stable over the entire range of skin-pigmentation levels. This was to be expected because of the fairly uniform skin reflectance spectrum in NIR. The skin-tone invariance of the algorithm is confirmed by ANOVA on the SNRa values from both setups.

Performed simulations and experiments show that the angle between the pulse and motion vector can be increased when LEDs with different wavelengths are applied instead of incandescent light. However, without a correct diffuser, different skin locations are exposed to different illumination spectra. Consequently, also different relative PPG amplitudes may occur in the three color channels of the RGB camera. Since the PBV pulse-extraction method assumes the same ratio of relative PPG amplitudes in the color channels over the entire skin area, the performance of the algorithm is expected to reduce when inhomogeneities are introduced.

V. CONCLUSIONS

This paper shows the feasibility of motion robust (cardiac) pulse detection in NIR. Current state-of-the-art methods developed for rPPG in visible light, are adopted for use in NIR. Simulations verified by large scale experiments, show that a setup consisting of three monochrome cameras with different optical filters is favorable in terms of motion robustness compared to a single RGB camera setup, where the IR-blocking filter is replaced by a visible light blocking filter. Experimental

results on 30 challenging benchmark video sequences with incandescent light show that both setups are capable of accurate pulse-extraction and their performance is comparable for all skin-types. In general, the RGB camera provides slightly better results, a MAE of 1.20 compared to a MAE of 1.53 for the monochrome cameras, where the difference is most prominent in the scaling motion scenario, likely induced by the effects of parallax. Since a single-optics setup is preferable, simulations with dedicated LEDs are performed to further improve motion robustness of the RGB camera, leading to a nearly similar pulse vector in terms of motion robustness compared to the monochrome cameras setup. When full darkness is desired, the dedicated NIR illumination should result in similar motion-robustness compared to incandescent light.

VI. ACKNOWLEDGEMENT

The authors would like to thank Ihor Kirenko, Wim Verkruijsse, Patriek Bruins, Mukul Rocque, Mohammed Meftah, Harry Sterken, Hugo Cornelissen and Jean Schleipen of Philips Research for their support. Furthermore, we would like to thank Xueming Lin for his assistance by constructing the dataset and all the volunteers who participated in the experiments.

REFERENCES

- [1] A. B. Hertzman and C. R. Speakman, "Observations on the finger volume pulse recorded photo-electrically," *Am. J. Physiol.*, vol. 119, pp. 334–335, 1937.
- [2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, 2007.
- [3] K. Humphreys, C. Markham, and T. Ward, "A CMOS camera-based system for clinical photoplethysmographic applications," *Proceedings of SPIE*, vol. 5823, pp. 88–95, 2005.
- [4] W. Verkruijsse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.
- [5] M. Hülsbusch and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using pPGL," *Proc. SPIE*, vol. 4683, pp. 110–117, 2002.
- [6] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.
- [7] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity," pp. 405–410, 2011.
- [8] M. Hülsbusch, "An image-based functional method for opto-electronic detection of skin-perfusion (in German)," Ph.D. dissertation, RWTH Aachen, 2008.
- [9] G. de Haan and V. Jeanne, "Robust pulse-rate from chrominance-based rPPG," *IEEE Trans. on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [10] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological Measurement*, vol. 3, no. 9, pp. 1913–1926, 2014.
- [11] W. Wang, S. Stuijk, and G. de Haan, "Exploiting spatial-redundancy of image sensor for motion robust rPPG," *IEEE Trans. on Biomedical Engineering*, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2014.2356291>
- [12] L. F. Corral Martinez, G. Paez, and M. Strojnik, "Optimal wavelength selection for noncontact reflection photoplethysmography," *22nd Congress of the Int. Commission for Optics*, vol. 8011, pp. 801 191–7, 2011.
- [13] Y. Kanzawa, T. Naito, and Y. Kimura, "Human skin detection by visible and near-infrared imaging," *IAPR Conference on Machine Vision Applications*, vol. 12, pp. 14–22, 2011.
- [14] T. Fitzpatrick, "The validity and practicality of sun-reactive skin types i through vi," *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871, 1988.
- [15] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," *ECCV 2012*, vol. 7575, pp. 702–715, 2012.
- [16] B. E. Bayer, "Color imaging array," Patent US 3 971 065, 1976.



Mark van Gastel received his M.Sc. in Electrical Engineering from the Eindhoven University of Technology in 2013. He is currently a PhD candidate in the Department of Electrical Engineering at Eindhoven University of Technology in cooperation with Philips Research Eindhoven. His research focuses on signal processing and computer vision with a particular interest in remote vital signs monitoring.



Sander Stuijk received his M.Sc. (with honors) in 2002 and his Ph.D. in 2007 from the Eindhoven University of Technology. He is currently an assistant professor in the Department of Electrical Engineering at Eindhoven University of Technology. He is also a visiting researcher at Philips Research Eindhoven working on bio-signal processing algorithms and their embedded implementations. His research focuses on modelling methods and mapping techniques for the design and synthesis of predictable systems with a particular interest into bio-signals.



Gerard de Haan received BSc, MSc, and PhD 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 taught 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 "Full-Professor" at 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, 170 scientific papers and more than 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.