Modified RGB Cameras for Infrared Remote-PPG

Modified RGB Cameras for Infrared Remote-PPG

Wenjin Wang, and Albertus C. den Brinker

***Abstract*—Multi-wavelength cameras play an essential role in remote photoplethysmography (PPG). Whereas these are readily available for visible light, this is not the case for near infrared (NIR). We propose to modify existing RGB cameras to make them suited for NIR-PPG. In particular, we exploit the spectral leakage of the RGB channels in infrared in combination with a narrow dual-band optical filter. Such camera modification is simple, cost-effective, easy to implement, and it is shown to attain a pulse-rate extraction performance comparable to that of multiple narrow-band NIR cameras.**

***Index Terms*—Vital signs monitoring, photoplethysmography, biomedical sensing, camera, RGB, near infrared.**

I. Introduction

CAMERA-based remote photoplethysmography (remote-PPG) enables contactless measurement of the cardiac pulse by detecting the pulse-induced subtle color changes of the human skin [1]. It can be used as a fundamental technology to measure various human vital signs (e.g. pulse rate (variability), respiratory rate, blood oxygen saturation, pulse transit time, etc.) for camera-based contactless health monitoring [2], [3]. The implementation of remote-PPG in near infrared (NIR) promises attractive applications in darkness, as it involves invisible light that is less obtrusive to the human eye. It is typically important for clinical applications that require long-term continuous monitoring of patients (24/7) including nights [4], such as in the Intensive Care Unit (ICU) [5], Neonatal Intensive Care Unit (NICU) [6], Coronary Care Unit (CCU) [7], and Sleep Center [8]. It is also important for assisted-living applications, such as elderly and baby care at home [9] to detect myocardial infarction or sleep disorders (e.g. apnea); and driver monitoring [10] in automotive to detect emergencies (e.g. cardiac arrest or stroke), etc. In view of all these applications, we consider NIR as the mode of most practical relevance.

Extracting the PPG signal from the camera video signals is much more difficult in NIR than in RGB. This is mainly due to two physiological reasons as explained in [11]: (i) low pulsatile strength (AC/DC) in the NIR spectrum and (ii) low pulsatile contrast between the NIR-wavelength channels. On the other hand, skin-tone is a less deteriorating factor for PPG extraction in NIR, because infrared light can penetrate deeper into the skin and skin melanin has less spectral absorption in this range [12], i.e. the difference between bright skin and dark skin is reduced. Various algorithmic solutions have been proposed to extract the PPG signal in NIR, with a particular focus on improving the motion robustness. These include (i) *single-wavelength NIR-PPG* [13], [14], which uses signal processing techniques (e.g. filtering, decomposition) to denoise the single-channel signal from the monochrome infrared camera; and (ii) *multi-wavelength NIR-PPG* [11], [15], which uses multi-channel combination (or color space projection) to eliminate non-pulsatile disturbances (e.g. motion). This is similar to the PPG extraction in RGB [16]–[18] that projects multi-channel color signals onto a single direction orthogonal to distortions. It has been shown [11] that multi-wavelength NIR-PPG has more degrees of freedom to suppress distortions and thus has more potential to give rise to a robust solution, i.e. it especially shows considerable improvements over the single-wavelength approach when using physiological principles of blood pulsation to design the channel combination function [11].

The use of multi-wavelength NIR-PPG requires a multi-wavelength NIR camera. Such cameras suited for ubiquitous vital signs monitoring (or video health monitoring) are hard to find nowadays (in view of performance and price) and there is no indication that this will change in the near future. For example, most surveillance night-vision cameras (e.g. IP cameras) are monochrome in infrared (with an active single-wavelength IR light source), which do not serve the purpose of multi-wavelength NIR-PPG that requires at least two NIR-wavelength channels for pulse extraction. However, there are indeed a few options built and prototyped in the lab:

(i) *Multi-/hyper-spectrum cameras*, which measure a wide range of spectral bands beyond the visible light. Whether they can be used for remote-PPG depends on their specifications. For example, we explored the option of using the high-end Silios camera that has eight infrared bands for PPG extraction. But due to the strong cross-talk between the bands (i.e. spectral overlap), we observed very limited independent information between the multi-band signals which indicates a serious drawback for PPG extraction (see Appendix A). It is possible to consider other multi-/hyper-spectrum cameras that may meet the requirements (e.g. narrow bands, high sensitivity, low sensor noise, etc.), but they are usually very expensive and only suited for research purpose, thus not realistic for large-scale deployment in health monitoring applications.

(ii) *Multi-camera systems*, which use multiple monochrome cameras with narrow-band IR filters centered at different wavelengths to mimic a multi-wavelength NIR camera. This has been prototyped in [8], [11], [15], demonstrating functional performance for measuring pulse rate and blood oxygen saturation and yielding reasonably good results. However, the setup involving multiple camera sensors and optics is still expensive and bulky (see Fig. 4 (a)), not flexible/lightweight for deployment/installation in practical settings (e.g. cockpit or incubator). Most importantly, the parallax between the different cameras can cause serious problems (e.g. artifacts of

W. Wang and A. C. den Brinker are with the Philips Research, Eindhoven, The Netherlands, e-mail: (wenjin.wang@philips.com).

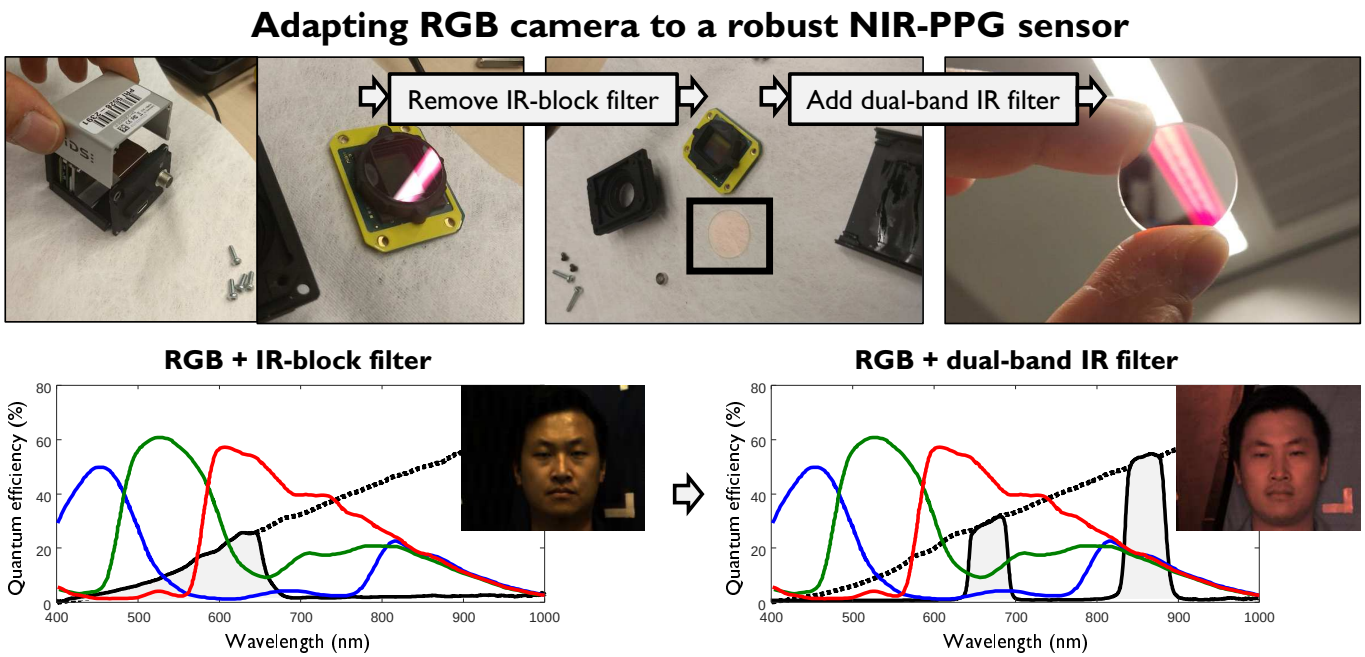

Fig. 1. The flowchart of adapting an RGB camera to a dual-band NIR camera for robust PPG measurement. It has two main steps: (i) open the RGB camera and remove the IR-block filter on top of the sensor; (ii) place in a dual-band IR filter (i.e. a narrow dual-band IR filter sensitive at 660 nm and 850 nm is demonstrated in our case). This modification leads to a shift of the sensed spectrum and thus a change of the DC image shown in the figure. The IR-block and dual-band filter responses combined with the incandescent light are represented by the solid black lines.

color shifts due to body motion or imperfect image alignment) for vital signs monitoring that measures subtle skin color changes. This is a typical bottleneck for the applications with a short subject-to-camera distance such as in automotive.

(iii) *RGB2NIR systems*, which convert an RGB camera into a multi-wavelength NIR camera by using its spectral leakage in infrared [19]. Replacing the IR-block filter in the RGB camera with a visible-light block filter, we can shift the sensed spectrum from visible to invisible. It is a cheaper and more lightweight solution than (i) and (ii). However, it has significant cross-talk problems since the spectral responses of the RGB filters largely overlap in the infrared part, especially at longer wavelengths. This is a common characteristic for consumer-grade RGB cameras, where the reduced NIR-channel independency restricts its performance for PPG extraction.

The principal goal of this work is to investigate and discuss the camera hardware solutions in NIR, specifically answer the question: *given the same piece of software/algorithm, can we modify the camera optics to improve the robustness of measurement?* Such an improvement is fundamental and general for vital signs monitoring. Although there is no perfect/off-the-shelf camera solution for multi-wavelength PPG monitoring in NIR, we still consider RGB2NIR to be a more appropriate/competitive option for pulse-rate monitoring in near-future applications (e.g. more cost-effective). Therefore, we use it as the basis to develop our solution. Since the major limitation of RGB2NIR is the strong cross-talk in NIR, we propose to use a dual-band IR cut filter to reduce the spectral redundancy/overlap between the RGB channels in NIR (see the flowchart of modification in Fig. 1). This modification increases the wavelength independency between the channels in NIR and thus improves the AC/DC color characteristics (i.e. relative pulsatility contrast) used for PPG extraction. The proposed solution, called "dual-band RGB2NIR", has been benchmarked against the latest multi-camera system that has six different wavelength combinations (from 1-wavelength to 4-wavelength). The results show that the proposed dual-band RGB2NIR modification outperforms the multi-camera system with 1-3 wavelengths and has a performance close to the 4-wavelength setting (in a condition where parallax has been reduced between the cameras). It is considered to be a motion-robust hardware solution in NIR, while it is lightweight, cost-effective, and easy to implement.

The remainder of this paper is structured as follows. In Section II, we describe the proposed dual-band RGB2NIR solution. In Sections III and IV, we use a benchmark to verify the proposition. Finally, in Section V, we draw the conclusions.

II. Method

Unless stated otherwise, we use the following mathematical conventions throughout the paper. Italic characters denote scalars. Boldface characters denote vectors and matrices; vectors are row vectors. Lastly, $\mathbf{1}$ denotes a row vector containing only ones.

To assess the performance of remote-PPG in NIR with a modified RGB camera, we experimented with the simple modification along the lines set by the earlier feasibility study [19]. Using an RGB camera means that the sensitivity of the RGB channels in NIR needs to be sufficiently high, such that the NIR signals are not lost in the quantization noise. Thus, not all regular RGB cameras are suitable for this purpose, such as the ones with low quantum efficiency in NIR. Secondly, operating in NIR implies that the distortions associated with specular

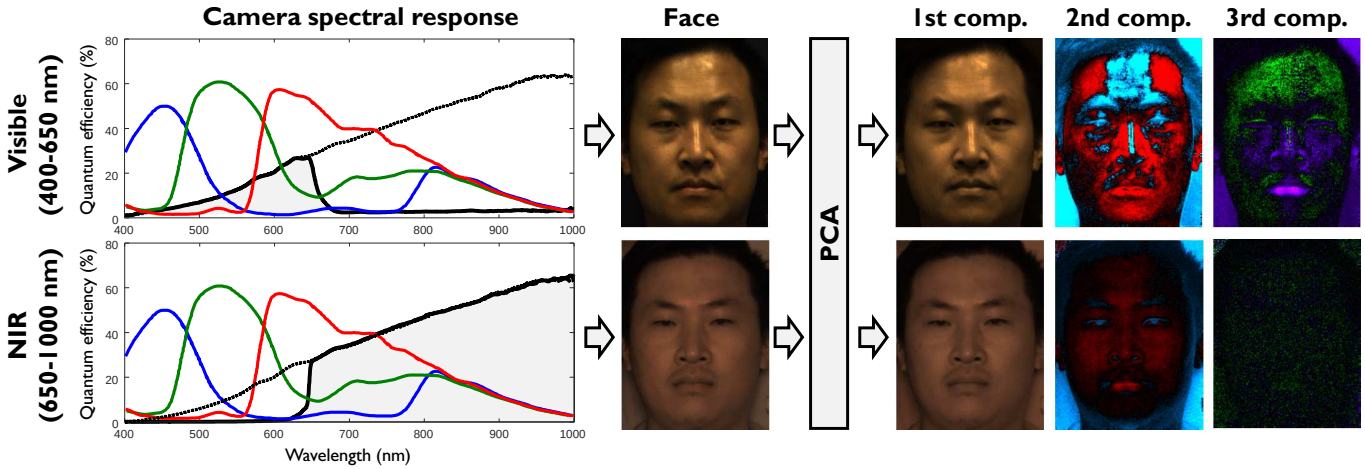

Fig. 2. The spectral responses of the RGB camera (camera type: IDS UI-3200SE-C-HQ, Sony IMX304LQR-C sensor) in the visible range (400-650 nm) and NIR range (650-1000 nm). Their DC images and associated PCA components factorized in the color space are also shown. The PCA images are created by back-projecting the PCA projections of image pixels to the original color space. The block filter responses combined with the incandescent light (dashed black lines) are represented by the solid black lines.

reflections are virtually absent, because infrared light penetrates deeper into the skin and exhibits much less mirror-like specular reflections as typically appear in the visible light [11]. According to [17], the challenge is thus to differentiate the pulse from motion-induced intensity variations while these are essentially the changes in the skin-tone (DC) direction in the original color space (or $\mathbf{1}$-direction in the DC-normalized color space). In principle, ignoring the specular reflection and sensor noise, one could use a 2-wavelength camera to achieve robust PPG measurement in NIR.

Consider the camera spectral responses as shown in Fig. 2. The RGB filters are peaked at different wavelengths in visible (400-650 nm) and the RGB channels roughly subdivide the wavelength in this range into three parts. This does not hold for the NIR range (650-1000 nm). Looking at the R and B curves in the infrared part, the plot suggests that if one would consider the B and R-G channels, the band would be subdivided into two parts. Obvious combination with the G channel to create three sub-bands like in the visible domain is absent, i.e. the R and G curves have similar spectral coverage. This intuition is supported by the Principle Component Analysis (PCA) performed on the image pixels (DC) recorded by the camera in visible and NIR (see Fig. 2). The physical meaning of such decomposition in the color space is: the first/primary component is related to the intensity level; the second and third components, orthogonal to the intensity, are related to the chromaticity (i.e. sudo-chromaticity in NIR) [20]. It is clear from the example that all PCA directions in visible contain meaningful spectral projections (i.e. full rank), whereas only the first two PCA components in NIR contain spectral information, i.e. the third component holds residual noise (almost no spectral information). This means that the investigated RGB camera in the NIR domain is essentially a 2-band system. The PCA example also supports our earlier statement that the specular reflection/component that is common in the visible light is largely absent in NIR, i.e. the second PCA component in the visible domain has clear specular reflection at the forehead and yet not in the NIR domain. Hence, this example corroborates the notion that PPG extraction in NIR is essentially defined as separating the intensity variation component and pulsatile variation component.

The camera spectral responses in Fig. 2 trigger us to improve this setup further: there is a large part of spectrum overlap between the RGB filters in NIR (> 815 nm, see details in Fig. 2), meaning that the RGB channels sense the same optical information. If we could reduce the spectral redundancy, the channel independency will be improved, thus the relative pulsatile contrast between channels can be increased (for better PPG extraction) [18]. An intuitive option is using two narrow-band cut filters to exclude the redundant long-wavelength tail and target the extremes of the remaining part (short NIR and end-mid NIR). It will still end up with a 2-channel system but the channels will be more independent than before. The rationale of restricting ourselves to a dual-band filter is that the underlying RGB2NIR system is already a 2-wavelength system, i.e. a single narrow band will reduce the number of channels, while using > 2 narrow bands does not bring extra benefit.

The question that remains is how to choose the two wavelengths for the dual-band IR filter. In principle this can be determined as follows. As components in an optical system, we take into account the source with spectrum $I(\lambda)$, the reflecting skin tissue $\rho(\lambda, t)$, the optical filter $F(\lambda)$ and the camera sensitivity $S_i(\lambda)$ (i.e. i-th channel). Following [17] we have for the i-th color channel:

$$C_i(t) = \int_\lambda I(\lambda)\rho(\lambda, t)F(\lambda)S_i(\lambda)d\lambda, \tag{1}$$

where λ is the spectral wavelength and t is the time. All time-dependencies are absorbed in the reflection ρ where we will consider blood volume changes and motion as its origins.

For the dual-band filter, we introduce a system consisting of two narrow-bands modeled as:

$$F(\lambda) = \delta(\lambda_1)\Delta_1 + \delta(\lambda_2)\Delta_2, \tag{2}$$

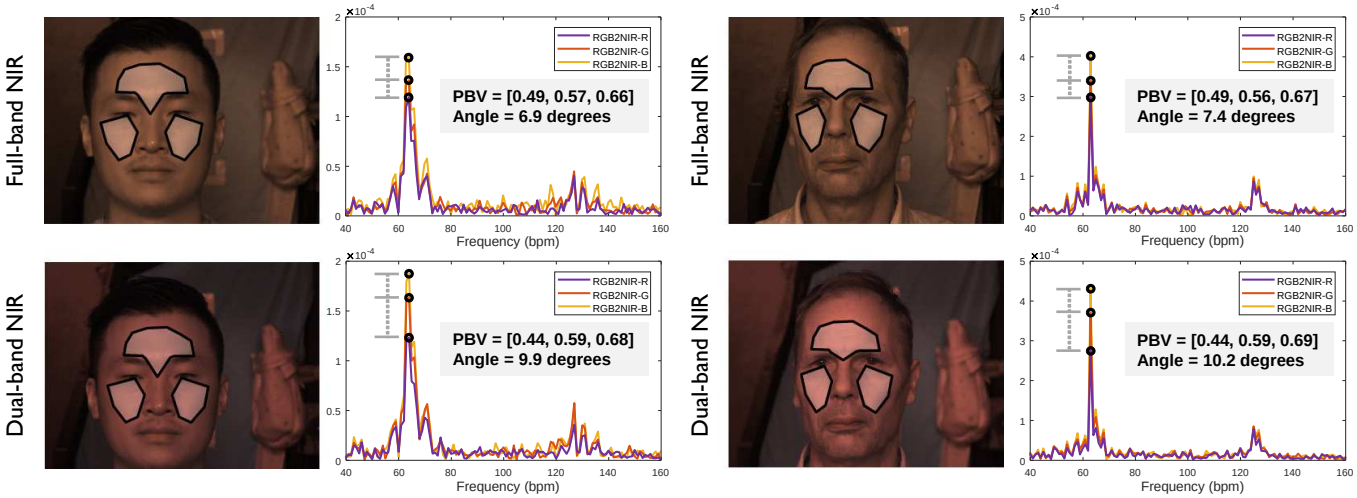

Fig. 3. The pilot measurements of the PBV vector and its angle w.r.t. the direction of $\mathbf{1}$ for both the full-band NIR (650-900 nm) and dual-band NIR (660 nm and 850 nm). The frequency spectra of the multi-channel infrared signals are measured from the averaged skin-pixel values from the subject's forehead and cheeks (with strong skin pulsatility) by leveraging a facial landmark tracker [21]. We mention that the colors of frequency spectra plots (of RGB2NIR) in this figure do not resemble the meanings of RGB.

where δ is the Dirac function; Δ_k $(k = 1, 2)$ is the band-width of the filter; and without loss of generality, we take $\lambda_1 < \lambda_2$. Using this model, the integral in (1) reduces to a sum:

$$C_i(t) = \sum_{k=1,2} I(\lambda_k)\rho(\lambda_k, t)F(\lambda_k)S_i(\lambda_k)\Delta_k. \tag{3}$$

For the skin reflection, we take the dichromatic reflection model introduced by [17]:

$$\rho(\lambda_k, t) = \rho_0(\lambda_k)(1 + m(t) + r(\lambda_k)P(t)), \tag{4}$$

where $m(t)$ and $P(t)$ represent the (zero-mean) motion signal and PPG signal, respectively (i.e. AC components); r is the relative PPG amplitude; ρ_0 is the steady, time-averaged skin reflection (i.e. DC component).

In order to be independent of the illumination strength and skin color (DC), we construct the DC-normalized camera signals [17]:

$$\tilde{C}_i(t) = \frac{C_i(t)}{\mu(C_i(t))} - 1, \tag{5}$$

where $\mu(\cdot)$ indicates time-averaging. Combining (3) and (5), we obtain:

$$\tilde{C}_i(t) = m(t) + \frac{\sum_{k=1,2} I(\lambda_k)\rho_0(\lambda_k)r(\lambda_k)S_i(\lambda_k)\Delta_k}{\sum_{k=1,2} I(\lambda_k)\rho_0(\lambda_k)S_i(\lambda_k)\Delta_k} P(t). \tag{6}$$

Considering the time-dependent color vector $\mathbf{C}(t) = [\tilde{C}_1(t), \tilde{C}_2(t), \tilde{C}_3(t)]$ we see that the motion is operating along the direction of $\mathbf{1}$ (i.e. intensity variation direction in the DC-normalized color space), while the pulse is operating along a direction dependent on the choices for λ_k $(k = 1, 2)$. To attain the best performance of source separation, it is desired that the color variation direction associated with $P(t)$ (i.e. blood volume pulse vector, abbreviated as the PBV vector [18]) deviates as much as possible from $\mathbf{1}$ [11].

We experimented with simple models for the involved quantities, the particulars can be found in Appendix B. The essential result is that for the full infrared-band setup (RGB2NIR) and light/dark skin types, the angle between PBV vector and $\mathbf{1}$ is about 7 degrees, the use of a dual-band filter increases this angle by several degrees, if the two bands are indeed positioned at the low and high wavelength range separately.

However, there are two practical limitations to this simulation. The first issue is that the derivations do not take the (lower) limits of the camera quantum efficiency into account: at low efficiencies, the sensor noise may dominate. The second issue is that the results depend on the relative PPG amplitude curve and the angle turns out to be sensitive to the exact details of this curve. Therefore, we first tried to support these theoretical results by pilot measurements to show the increased angle.

The pilot experiment measures the PBV vector and its angle w.r.t. $\mathbf{1}$ in the DC-normalized color space for both the full-band and dual-band RGB2NIR system, where the dual-band filter was set at 660 nm and 850 nm (band-width 25 nm) (based on the modeling in Appendix B). The light source was an incandescent lamp and two participants were measured as the showcase. It was found that by using the dual-band filter (see Fig. 3), the angle between PBV vector and $\mathbf{1}$ is increased from 6.9-7.4 degrees to 9.9-10.2 degrees, which is an 38-44% increase in angle. More specifically, the major contrast improvement is between the R channel and G/B channel. The R channel becomes more independent of the other two channels when using the narrow dual-band filter, particularly due to the lower band at 660 nm (e.g. makes it more purely red). This is in line with our expectation. In the next section, we shall benchmark the dual-band RGB2NIR modification against a multi-camera system that uses multiple monochrome cameras with narrow-band infrared filters. The latter one is considered to be the reference system as it demonstrated robust pulse-rate monitoring in NIR [11].

The essential idea of using two narrow bands to reduce the spectral redundancy/overlap in the NIR range of an RGB camera to increase the NIR-channel independency is general, but the exact dual-band selection is not unique, since it depends on the characteristics (e.g. spectral shape) of the light

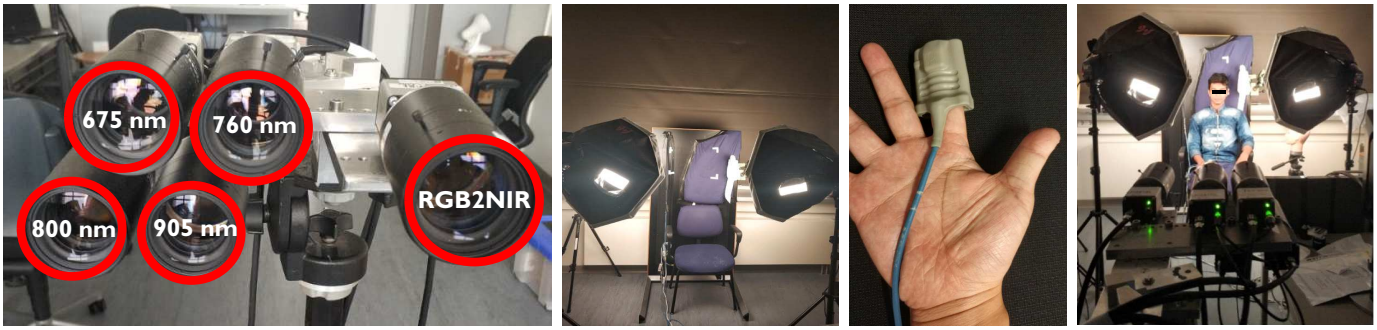

Fig. 4. Illustration of the (a) camera setup, which includes two camera systems: dual-band RGB2NIR system (right side) and 4-wavelength multi-camera system (left side); (b) incandescent illumination unit; (c) reference unit with a finger pulse-oximetry, and (d) actual recording condition with a subject.

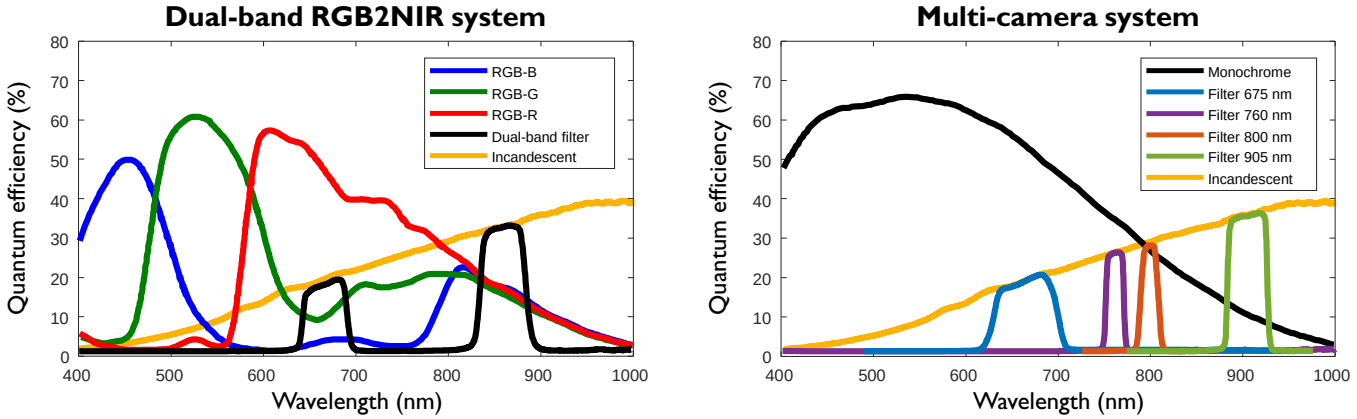

Fig. 5. The wavelength spectra of different optical components in the dual-band RGB2NIR system (e.g. RGB camera filter, dual-band IR filter) and multi-camera system (e.g. monochrome camera filter, narrow-band IR filters). The 4-wavelength multi-camera system can generate different wavelength combinations. The block-filter responses shown in this figure are in combination with the incandescent light.

source. In this paper, we demonstrate a way of selecting a dual-band filter "660-850 nm" for the used RGB camera with the illumination of incandescent light, and such a way of selection (by modeling) can be replicated for other RGB cameras and light sources.

III. Experimental Setup

This section presents the experimental setup for the benchmark. First, we introduce the benchmark dataset. Next, we discuss the protocols for comparing different camera solutions. Finally, we present the evaluation metrics.

A. Benchmark dataset

This paper investigates camera optics for remote-PPG monitoring in NIR, especially aimed at comparing the proposed dual-band RGB2NIR system with the multi-camera system (reference). We create a benchmark dataset based on the following components.

• **Dual-band RGB2NIR system** This is our proposed solution that modifies an RGB camera to a 2-wavelength NIR camera by replacing the IR-block filter with a dual-band IR filter. The RGB camera type is: IDS UI-3200SE-C-HQ (Sony IMX304LQR-C, CMOS sensor, global shutter). The used dual-band IR filter is: MidOpt DB-660/850 Dual Bandpass filter active at 660 nm and 850 nm (∼25 nm bandwidth).

• **Multi-camera system** This is the reference setup demonstrated by [11], [15], which uses four monochrome cameras with narrow-band IR filters centered at 675 nm (∼20 nm), 760 nm (∼20 nm), 800 nm (∼20 nm) and 905 nm (∼33 nm) to mimic a 4-wavelength NIR camera. The monochrome camera type is: IDS UI-3200SE-M-GL (Sony IMX304LLR-C, CMOS sensor, global shutter). It has a similar specification as the used RGB camera for fair comparison (i.e. sensors are from the same manufacturer and same model).

The wavelength spectra of the cameras and filters used by the dual-band RGB2NIR and multi-camera systems are shown in Fig. 5. Both setups use the same type of lens for their camera sensors: Tamron 1.1" 50 mm F/1.8 (Model M111FM50). The cameras are placed on a tripod (see Fig. 4 (a)) in a 2×2 array positioned around 3 m in front of the subject to reduce the parallax[1] present in the multi-camera system. The distance between the lenses of two adjacent cameras is 6 cm, leading to a parallax angle of 11.4° between at the viewing point. With the used focal length, it results in approximately 30-40% skin area in a video frame. All auto-adjustment functions

[1]Parallax is the displacement in the apparent position of an object viewed along different optical paths. To reduce the parallax, image planes of two (adjacent) cameras are first roughly aligned before recording using a real-time overlay feedback. Next, image planes of two cameras are registered using a similarity transformation model consisting of translation, rotation and scaling. The registration model is estimated from the initial frames of two cameras based on their epipolar geometry.

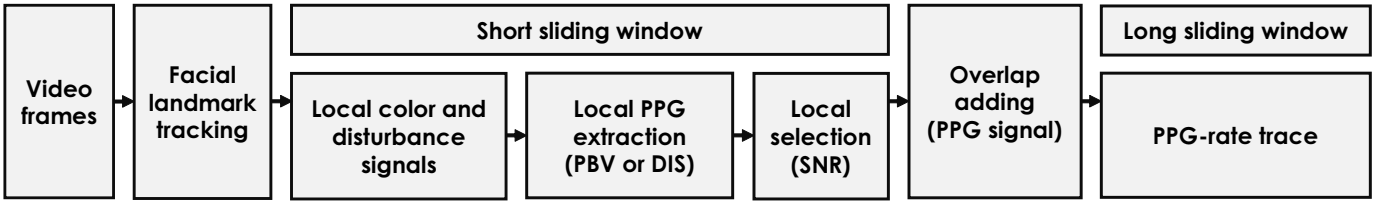

Fig. 6. The architecture of the camera-based vital signs monitoring system used for pulse-rate extraction [11]. The input is the video frames and the output is the pulse-rate trace. It includes a facial landmark tracker [21], a short sliding window for extracting the local color and disturbance signals (i.e. disturbance source is used for DIS), a core PPG-extraction algorithm (PBV or DIS) to extract local PPG signals, a Signal-to-Noise-Ratio (SNR) based metric to select local candidates, and an overlap-add procedure to generate a long pulse signal based on the pulse intervals measured from the short sliding window. In the end, a long sliding window is used to generate the pulse-rate trace. The pulse rate is derived in the frequency domain by using the frequency index of the maximum spectral peak of the pulse signal within the long sliding window.

(e.g. auto-focus, auto-gain, auto-white-balance, auto-exposure) of the camera are turned off during the recording. All videos are recorded in *uncompressed* video format (with 1280×960 pixels, 8 bit depth) at a constant frame rate (15 frames per second).

• **Light source** The illumination sources are two incandescent light fixtures consisting of 9 lamps each. Each fixture is powered at $220\,\text{V}\times1.2\,\text{A} = 264\,\text{W}$, which supplies sufficient energy for the NIR sensing. Two fixtures are placed on the two sides of the subject with 45° angle, providing a homogeneous illumination condition (see Fig. 4 (b)). The spectrum of the used incandescent lamp is shown in Fig. 5.

• **Ground-truth** The ground-truth is the raw PPG signal recorded by a finger-based transmissive pulse oximetry (Model: Philips IntelliVue X2, see Fig. 4 (c)) synchronized with the video acquisition using time stamps. The PPG reference is electrically decoupled from the video recording system. The reference raw PPG signal is processed by a 4-th order Butterworth band-pass filter with cutoff frequencies [0.6, 3] Hz in a zero-phase forward and reverse filtering mode for reducing out-band components.

• **Subjects** A total of 15 healthy adult subjects (aged between 25 and 55 years), with different skin tones categorized from type I to VI according to the Fitzpatrick scale, participated in recordings (8 subjects in type I-II, 4 subjects in type III-IV, and 3 subjects in type V-VI). This study has been approved by the Internal Committee Biomedical Experiments of Philips Research, and informed consent has been obtained from each subject. The recording consisted of 7 phases: 4 stationary phases with 3 motion phases in between. Each phase was approximately one minute, resulting in recordings of 6-7 minutes long. In the odd-numbered phases, the participant was instructed to be stationary. In the other phases, the participant was instructed to perform free head movements (i.e. leading to a mixture of rotation, translation and scaling in the recorded images) and facial expressions (like smiling and talking). There was no instruction for the subject to perform a single type of motion like periodic head rotation, as this is not particularly relevant in practice. The recording scenario is shown in Fig. 4 (d).

B. Benchmark protocols

We benchmarked the proposed dual-band RGB2NIR system with the reference multi-camera system. As for the benchmark, we derived six wavelength combinations from the multi-camera setup (from 1-wavelength to 4-wavelength)[2]: 905 nm, 760-905 nm, 800-905 nm, 675-800-905 nm, 760-800-905 nm, 675-760-800-905 nm. The reason for comparing the dual-band RGB2NIR with different wavelength combinations of the multi-camera system is to help us understand/target the position of our proposition in the reference system, i.e. is it similar to the setting of 2-wavelength or 3-wavelength? It also helps us to investigate the requirement and limitation of the multi-camera system, i.e. what is the minimum number of wavelengths/cameras required for a robust pulse-rate monitoring and what are the wavelength specifications?

We used a (basic) end-to-end vital signs monitoring architecture prototyped in [11] to measure the pulse-rate from videos. The processing architecture is illustrated in Fig. 6 and its settings are kept the same as [11] (by default). It includes various image/signal processing steps that have been validated in prior art, such as using facial landmark detection and tracking to correct skin motion [21], spatial redundancy (or multi-site PPG extraction) to improve the signal quality [22], band-pass filtering to eliminate out-band distortions, core PPG-extraction algorithm (e.g. DIS) to eliminate in-band distortions [11], etc. In this architecture, two different core PPG-extraction algorithms (PBV [18][3] and DIS [11][4]), which have shown robust performance in NIR, are used as alternatives for pulse-signal extraction. The reason is to ensure the conclusions drawn on the camera setup are not dependent on the processing method. We stress that the benchmark is not intended to compare the algorithms, but to obtain results for comparing camera hardware solutions.

All processing steps have been implemented in Matlab R2017b (using Signal Processing Toolbox) and run on a laptop with an Intel Core i7 processor (2.70 GHz) and 8 GB RAM.

[2]For the 2-wavelength setup, we selected the cameras next to each other for the combination instead of the cameras on the diagonal places, to reduce the influence of parallax. For the 3-wavelength setup, we used the camera in the middle as the bridge to transfer the registration of diagonal cameras, instead of registering two diagonal cameras directly.

[3]Blood Volume Pulse Signature (PBV) is a method that uses the relative PPG amplitude between camera wavelength channels as a signature to extract the pulse from the input camera signals by one-step least squares regression.

[4]Discriminative Signature based Pulse Extraction (DIS) is an evolved version of PBV that extends the input wavelength signals to include motion signals for pulse extraction. It extends the blood volume pulse signature to a discriminative signature with additional zero-entries to explicitly suppress the motion signals in the least squares regression.

TABLE I
AVERAGE METRIC VALUES OBTAINED BY PBV AND DIS ON THE BENCHMARK DATASET RECORDED BY SEVEN BENCHMARKED CAMERA SETUPS.

Metric	Method	Dual-band RGB2NIR	Multi-camera system (reference)					
			4 wavelengths	3 wavelengths		2 wavelengths		1 wavelength
			675-760-800-905	675-800-905	760-800-905	760-905	800-905	905
RMSE (bpm)	PBV	**9.50**	8.45	10.73	12.78	16.22	17.55	18.84
	DIS	**5.90**	5.19	6.14	7.11	9.33	10.82	12.72
SR-AUC	PBV	**0.70**	0.72	0.66	0.62	0.56	0.51	0.49
	DIS	**0.78**	0.78	0.76	0.73	0.68	0.63	0.59
Coverage (%)	PBV	**75.34**	76.53	71.09	66.44	59.86	55.65	53.11
	DIS	**83.53**	84.10	81.09	77.35	71.00	65.57	62.18

* Underline and **boldface** entry denote the best and second best camera setup per row.

C. Evaluation metrics

We use three metrics to evaluate the performance of pulse-rate extraction. The purpose of using different evaluation metrics is to provide different interpretations of the results (e.g. clinicians are typically more comfortable with RMSE, while automotive/fitness/well-being manufactures are more interested in the measurement coverage based on their applications/products).

• **Root-Mean-Square Error (RMSE)** We use RMSE to measure the difference between the reference PPG-rate trace and camera PPG-rate trace. RMSE represents the sample standard deviation of the absolute difference between reference and measurement, i.e. smaller RMSE suggests more accurate measurement.

• **Area Under Curve (AUC) of Success Rate** The Success Rate (SR) refers to the percentage of video frames where the absolute difference between the reference PPG-rate trace and camera PPG-rate trace is bound within a tolerance range (T). To enable the statistical analysis, we estimate a SR curve by varying $T \in [0, 10]$ (i.e. $T = 0$ means completely matching between the camera and reference, and $T = 10$ means allowing a 10 bpm difference), and use the Area Under Curve (AUC) as the quality indicator, i.e. larger AUC means more accurate extraction. We note that the AUC is normalized by 10 (the total area) and thus varies in the range of [0,1].

• **Coverage at ± 3 bpm** We use the Success Rate with setting $T = 3$ to denote the measurement coverage of a remote-PPG method. This is a straightforward interpretation giving the percentage of time that the camera PPG-rate has a difference less than or equal to 3 bpm w.r.t. the reference.

IV. RESULTS AND DISCUSSION

This section presents the benchmark results of seven camera solutions based on two remote-PPG extraction algorithms (PBV and DIS). Their performance is compared and discussed quantitatively and qualitatively. Note that for the table and figures, we place the proposed dual-band RGB2NIR on the left side and multi-camera solutions (with different wavelength combinations, ordered by wavelength) on the right side, which is our benchmark system (reference). Placing two systems separately gives a clear overview of their individual performance and cross-system comparison.

Table I shows average metric values obtained by PBV and DIS on seven camera solutions, from which we can see that the 4-wavelength combination of the multi-camera system is the best but only marginally better than the second best: dual-band RGB2NIR, i.e. when using DIS, their difference in terms of measurement coverage is smaller than 1%. The dual-band RGB2NIR shows better performance than the 3-wavelength combinations prototyped in [11]. The most fair and straightforward comparison is between the dual-band RGB2NIR and 2-wavelength combinations, as both are essentially the dual-wavelength system. Table I shows that the dual-band RGB2NIR significantly improves the 2-wavelength multi-camera setup, i.e. when using DIS, the RMSE is reduced by half, while the measurement coverage is increased by 15%. The reason for the improvement is presumably that the multi-camera setup has parallax between cameras that may causes additional color gradients (artifacts) when the video frames from two cameras are not perfectly registered. The parallax-induced color distortions will be magnified when the subject-to-camera distance is small or varying (due to body motion). Therefore, the multi-camera setup has more distortions to address than the dual-band RGB2NIR setup that has single optical path, where a single distortion (e.g. intensity variation) was assumed.

For a more detailed comparison, we show the box-plots of PBV and DIS per evaluation metric in Fig. 7. The ranking of seven camera solutions is as follows: the best is the 4-wavelength multi-camera setup; second by the dual-band RGB2NIR; followed by 3-wavelength, 2-wavelength and 1-wavelength combinations of the multi-camera setup. As a matter of fact, the performance of dual-band RGB2NIR is very close to that of the 4-wavelength setup in three evaluation metrics. We also stress that in the experimental setting, parallax has been reduced for the multi-camera setup (e.g. the subject-to-camera distance is 3 m), which is an optimized condition for the multi-camera system as compared to practical settings (e.g. automotive, incubator). This figure also shows that for the 3- or 2-wavelength setup, the wavelength selection matters, i.e. 675-800-905 nm is better than 760-800-905 nm; 760-905 nm

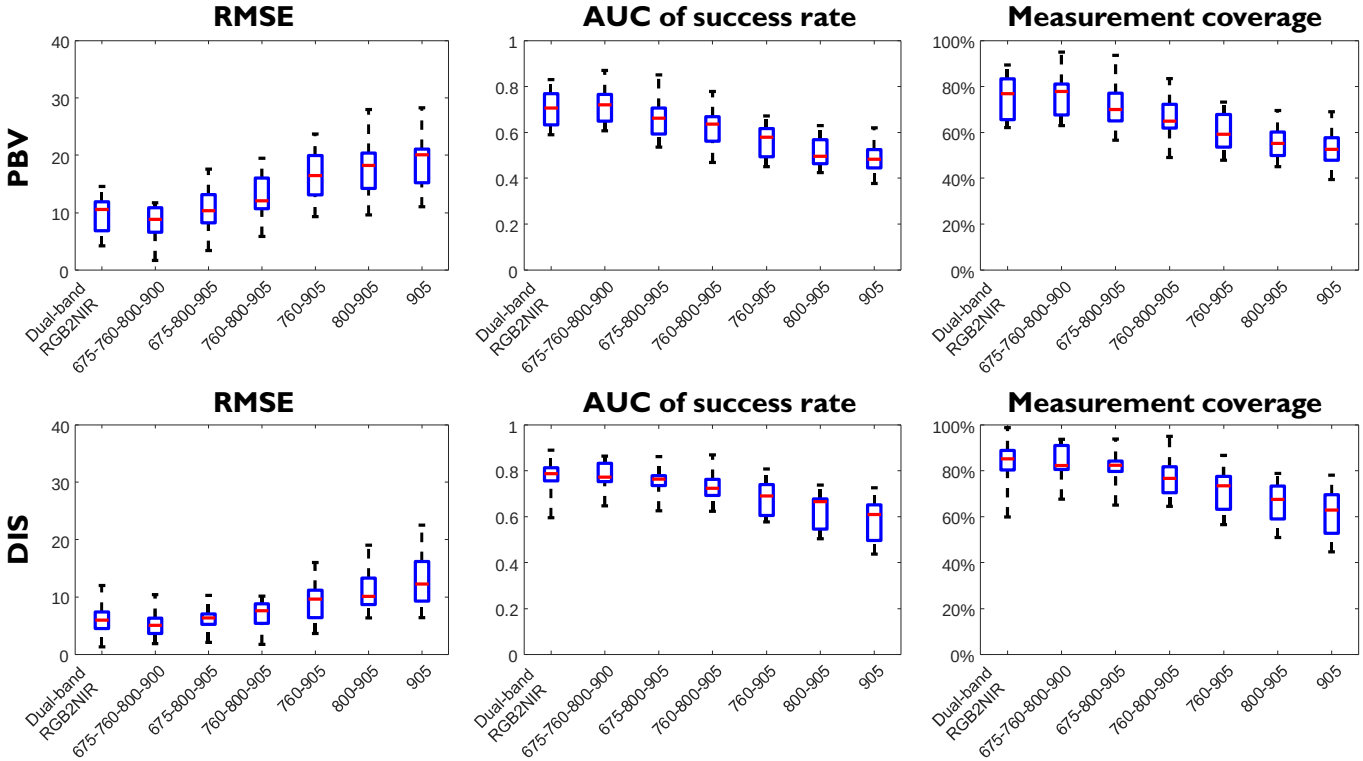

Fig. 7. Statistical comparison of the metric values obtained by PBV and DIS on the overall dataset recorded by seven camera setups. The median values are indicated by horizontal bars inside the boxes, the quartile range by boxes, the full range by whiskers.

is better than 800-905 nm. Though we cannot iterate over all wavelength combinations in the multi-camera setup (due to parallax), we express the following expectations: if not considering the problem of parallax, we expect 675-905 nm to be better than 760-905 nm, because its wavelength contrast (i.e. relative PPG amplitude contrast) is larger, which improves the differentiation with motion distortions. But we expect it to be worse than 675-800-905 nm that uses three wavelengths to eliminate distortions. The far-apart wavelengths increase the channel contrast and thus improve the pulsatile signature characteristics for PPG extraction, which is the same principle we used to specify the wavelengths for the dual-band RGB2NIR camera.

To further understand in which condition (e.g. stationary or motion) the dual-band RGB2NIR setup makes a clear improvement, we split the results into two parts: stationary and motion, according to the video recording protocol (e.g. stationary phase and motion phase), and plot their averaged metric values per condition in Fig. 8. It is found that the major variation among the camera setups is due to the motion phase. When the subject remains stationary, if there are no other signal sources (i.e. disturbances) apart from the PPG source, even the single-wavelength camera (905 nm) can measure a clean pulse. In the stationary phase, the major difference is between the 905 nm camera and other setups, though 905 nm alone also performs well, i.e. when using DIS, it has an averaged measurement coverage of 92.09%. In contrast, we see a much clearer performance difference between the seven camera setups during the motion phase, i.e. when using DIS, the dual-band RGB2NIR almost doubles the measurement coverage of the 2-wavelength multi-camera setup. Since the wavelength "850 nm" is used in our dual-band RGB2NIR, we add a discussion on the single-wavelength 850 nm and the benchmarked 905 nm: we expect 850 nm alone to be worse than 905 nm alone. This is because that (de-)oxygenated hemoglobin has lower spectral absorption variations at 850 nm than at 905 nm, meaning that the PPG amplitude at 850 nm is lower. Thus the PPG measurement at 850 nm is expected to be more susceptible to (motion) distortions.

The parallax in the multi-camera system does not play a major role when there is no motion. When the monitored subject has body motions, the parallax-induced color distortions become severe and deteriorate the measurement. This implies that (i) the proposed dual-band RGB2NIR setup is suited for both the stationary and motion scenarios, and its major improvement is in motion, which is thus considered to be a motion-robust hardware modification; (2) the multi-camera setup or even the single-camera setup works well in stationary but has performance limitations during motion. Though adding more wavelength channels (e.g. 4-wavelengths) may suppress the additional distortions due to parallax, it comes at the price of a bulky and expensive device by adding more sensors and optics.

In Fig. 9, we show the PPG spectra and corresponding PPG-rate traces of a single subject (as a qualitative example from 15 subjects) obtained by five camera solutions with the use of two core PPG-extraction algorithms (PBV and DIS), i.e. the best 2-wavelength and 3-wavelength setups are shown. It shows that the dual-band RGB2NIR particularly

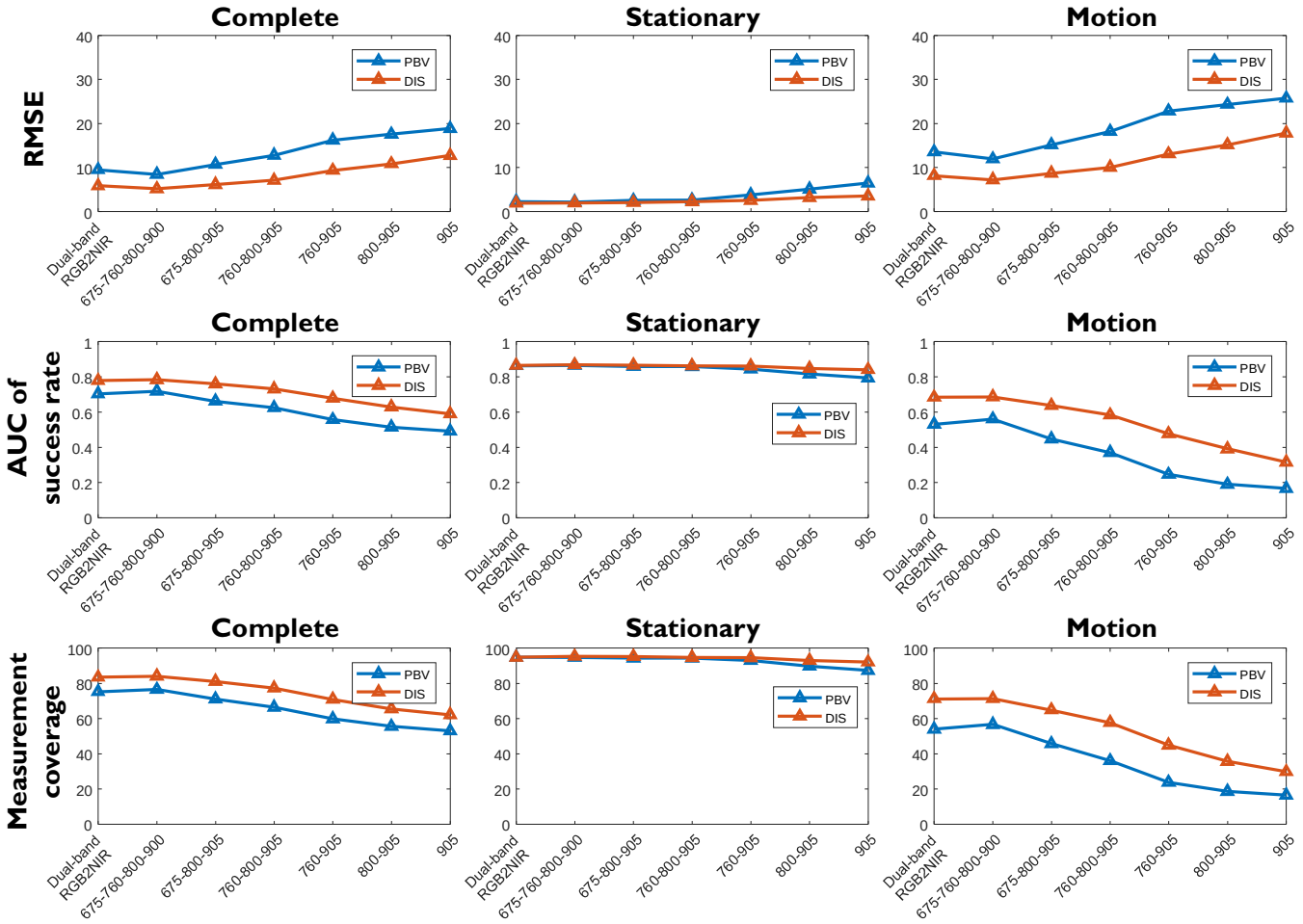

Fig. 8. Comparison of the average metric values obtained by PBV and DIS on different video segments (i.e. complete video, stationary part, motion part) recorded by seven benchmarked camera setups.

improves the multi-camera system ($<$ 4-wavelengths) during the motion phase of the recording, i.e. its PPG spectrum is cleaner and the derived PPG-rate is more consistent with the reference. It has a performance close to that of the 4-wavelength setup. We also see that the erroneous measurement is gradually reduced when using more wavelengths in the multi-camera setup. Although the benchmark is focused on the comparison of camera hardware setups, we mention that DIS is consistently better than PBV (see Fig. 9), which is in line with the conclusions of [11] that DIS using additional motion channels as input can better differentiate pulse and motion.

In view of the present results, we conclude that the proposed dual-band RGB2NIR, i.e. adding a narrow dual-band IR cut filter for the RGB camera, is a motion robust hardware modification for NIR, at least within the considered experimental setting. It increases the wavelength independency (i.e. pulsatile contrast) between the modified NIR-channels for better PPG extraction. The proposed solution is clearly better than the multi-camera setups with 1-3 wavelengths and even shows a competitive performance against the 4-wavelength setup in the condition with reduced parallax. The dual-band RGB2NIR solution significantly reduces the cost on hardware (e.g. sensors and optics) of the multi-camera solution. It also leads to a more compact, lightweight, and portable design that will facilitate the landing of vital signs camera technology in the healthcare domain and ease its implementation in various application scenarios (e.g. hospital care units, home care, automotive).

In the end, we also discuss the advantage of the dual-band RGB2NIR in terms of price. As compared to a N-wavelength multi-camera setup, the dual-band RGB2NIR setup will be N times cheaper in these three components: camera sensor, lens and optical filter. It has a price close the single-wavelength NIR camera but has a clearly improved performance. It is much cheaper than the 4-wavelength camera while achieving a similar performance. In addition, the dual-band RGB2NIR does not need the synchronization and image registration as required by the multi-camera setup between different cameras, thus its video streaming has a reduced CPU footprint. It has a more compact form factor that facilitates its integration in practical settings.

As for future studies, we intend to further validate the dual-band RGB2NIR setup in real-life applications with more practical challenges, such as driver monitoring in automotive, where the measured subject has severe body motions; the vehicle has vibrations; and ambient light has significant changes. We also intend to validate the setup in clinical trials with

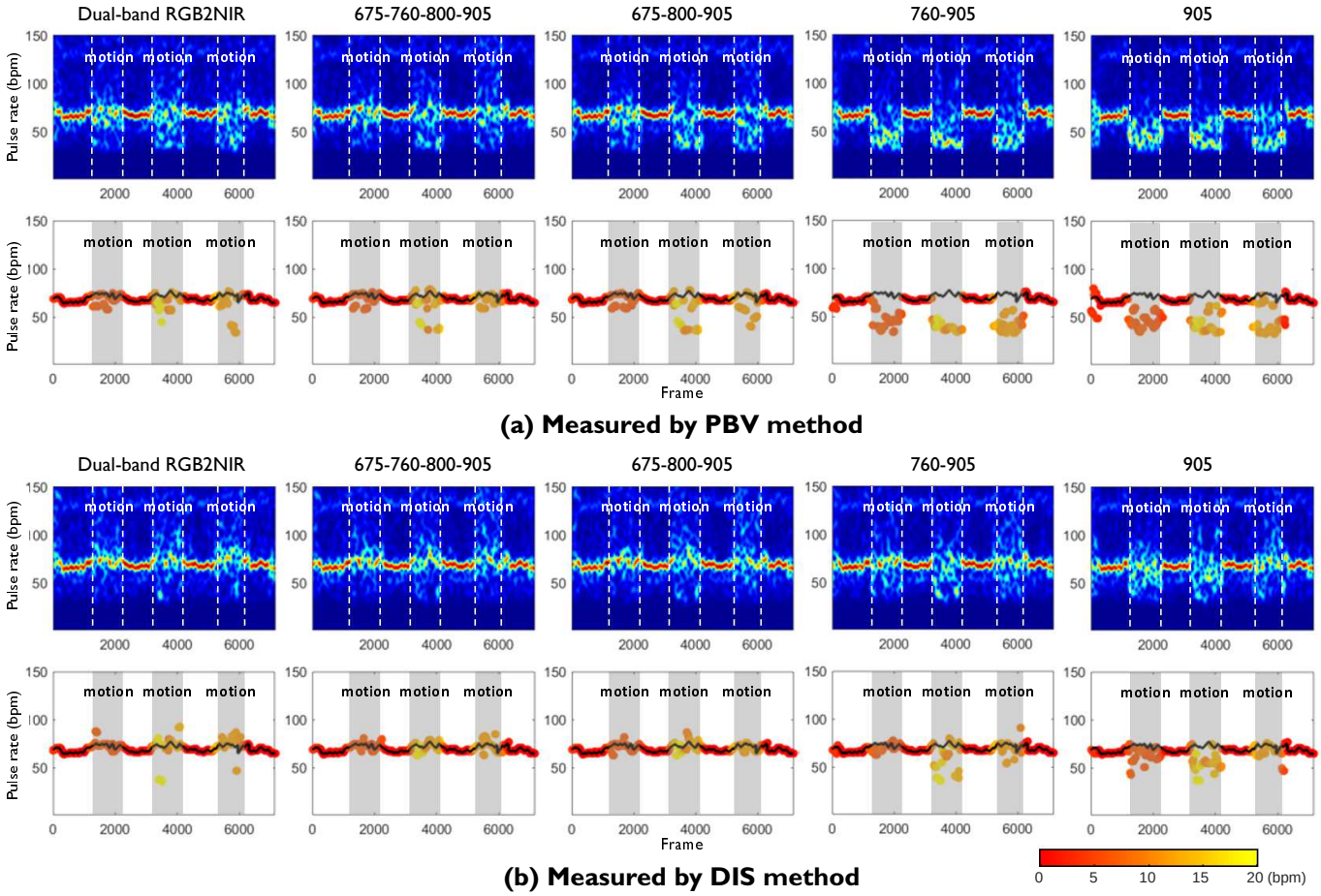

Fig. 9. The PPG spectra and corresponding PPG-rate traces of a test subject (exemplified from 15 subjects) measured by (a) PBV and (b) DIS on five camera setups (i.e. the best 2-wavelength and the best 3-wavelength setups are selected). The x-axis denotes the frame number (time axis) and the y-axis denotes the pulse rate (frequency axis). The reference PPG-rate is denoted by the solid black line. The motion periods are indicated in gray. The color coding of the estimated pulse rate (dots) in the PPG-rate traces is according to the absolute difference w.r.t. the reference as indicated in the lower-right color bar.

real patient data, such as in ICU and NICU that require 24/7 continuous monitoring including nights. We expect the dual-band RGB2NIR setup to show more advantages than the multi-camera setup in these real application scenarios, especially for those with small subject-to-camera distance such as in automotive or incubator.

V. Conclusion

We introduce a novel camera apparatus for remote-PPG monitoring in infrared. We propose to modify the existing RGB camera to a multi-wavelength NIR camera by replacing the infrared block filter with a narrow dual-band IR filter. The essence is to increase the wavelength independency (pulsatile contrast) between the modified NIR-channels for better pulse extraction. The proposed dual-band RGB2NIR modification is shown to be a motion-robust hardware solution as compared to the reference multi-camera system with narrow-band filters, within an experimental setup using an incandescent light source. It has competitive performance combined with a more favorable price point and compact form factor. It thereby facilitates wide-spread application of NIR remote-PPG in healthcare, and it is simple to implement.

Acknowledgment

The authors would like to thank Dr. Wim Verkruysse for assistance in modifying the RGB camera, Dr. Mark van Gastel for discussion on the topic, and the volunteers from Philips Research for their participation in the recordings.

Appendix A

We used a high-end multi-spectral Silios camera (type: CMS-S 650-930 nm) that has eight infrared bands (600-1000 nm) to measure the raw PPG signals from a human skin. An example of measurement is shown in Fig. 11. It was found to have strong cross-talk (i.e. overlap redundancy) between bands such that the AC/DC PPG signals do not differ much from each other in terms of amplitude, i.e. their relative PPG amplitudes are [0.32, 0.30, 0.30, 0.32, 0.34, 0.37, 0.37, 0.37, 0.31] (l_2-norm normalized), i.e. the minimum and maximum are 0.30 and 0.37. We further used the Principal Component Decomposition (PCA) to analyze the eigenvalues of the 8-band PPG signals. Their PCA eigenvalues are [1, 0.0042, 0.0025, 0.0006, 0.0004, 0.0003, 0.0002, 0.0001, 0.0000] (normalized w.r.t the first eigenvalue), i.e. the energy is dominated by the primary component, the rest components are rather small.

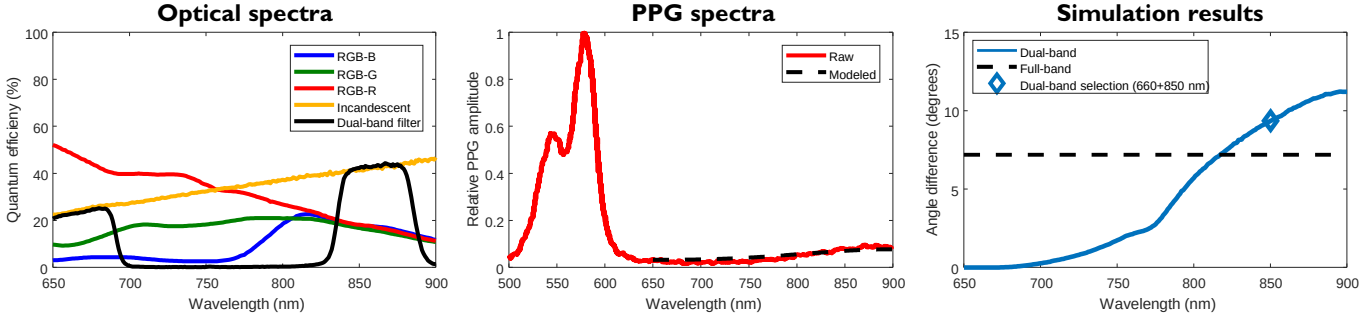

Fig. 10. The optical and PPG spectra used in the simulation, left and center panel respectively. Right panel: angle between PBV and $\mathbf{1}$ as a function of wavelength λ_2. The linearly modeled light source and light skin model were used and the first optical filter is fixed at $\lambda_1 = 660\,\text{nm}$. The dashed line in the right panel indicates the angle using a full-band NIR setup; the diamond indicates the combination of a 660-850 nm dual-band filter.

This suggests that eight infrared channels of the Silios camera provide very limited wavelength-independent information.

APPENDIX B

This appendix gives the details of the models used in the simulations. The models apply only to the wavelength range 650-900 nm, as this is the range of interest.

For the light source we used both an actual measurement of the incandescent lamp used in the experiments as well as a simple model having a linear increase in intensity over the target range:

$$I(\lambda) = I_0(\lambda - 450).$$

For the spectral sensitivities $S_i(\lambda)$ we took the measured camera responses as shown in Fig. 1.

For the skin $\rho_0(\lambda)$, we took two models representing the extreme skin tones based on the plots in [12]. For light skin we took a linear decrease from 1 to 0.9 over the range from 650 nm to 950 nm while for dark skin a linearly increase from 0.2 to 0.85 was used.

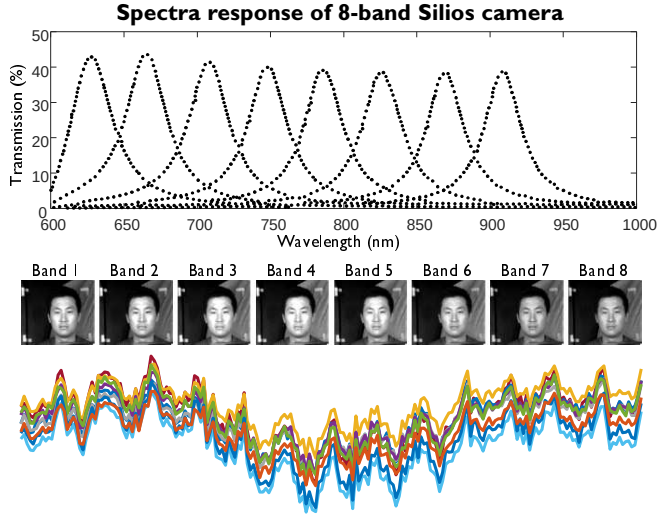

Fig. 11. The spectral response of the 8-band Silios camera, and the 8-band DC images and AC/DC signals measured from the subject facial skin area. Due to the strong cross-talk between the bands, the multi-spectral signals provide very limited wavelength-independent information.

Assuming the oxygenated blood, the remote-PPG characteristic in the NIR range is generally considered as a function having a minimum around 660 nm followed by an monotonic increase to a maximum near 900 nm [19], [23]. Correspondingly we took:

$$10\log_{10} r = r_0 - 1.9\cos((\lambda - 660)\pi/240),$$

we note that the simulations are not affected by the choice of r_0.

Fig. 10 shows the plot of the calculated angle as a function of the wavelength λ_2 (second optical filter) with the first optical filter set to 660 nm (i.e. low band of the NIR range) for the light skin model and the incandescent source modeled by the linear function. As expected, having the second wavelength close to the first, is a poor choice. According to these simulations, taking the wavelength $\lambda_2 > 810\,\text{nm}$ results in an increased angle difference relative to the full-band setup where the maximum occurs at 900 nm. In view of the fact that we did not take the camera sensor noise into account (i.e. lower Signal-to-Noise Ratio in the longer wavelength part), it was decided to set the second optical filter at $\lambda_2 = 850\,\text{nm}$.

REFERENCES

[1] W. Verkruysse *et al.*, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21 434–21 445, Dec. 2008.
[2] C. H. Antink *et al.*, "A broader look: Camera-based vital sign estimation across the spectrum," *Yearbook of Medical Informatics*, vol. 28, no. 1, pp. 102–114, 2019.
[3] W. Wang, "Robust and automatic remote photoplethysmography," Ph.D. dissertation, Eindhoven University of Technology, The Netherlands, 2017.
[4] L. Tarassenko *et al.*, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, p. 807, May 2014.
[5] S. Yeung *et al.*, "A computer vision system for deep learning-based detection of patient mobilization activities in the ICU," *NPJ Digital Medicine*, vol. 2, 2019.
[6] J.-C. Cobos-Torres *et al.*, "Non-contact, simple neonatal monitoring by photoplethysmography," *Sensors*, vol. 18, no. 12, p. 4362, 2018.
[7] S. Rasche *et al.*, "Remote photoplethysmographic assessment of the peripheral circulation in critical care patients recovering from cardiac surgery," *Shock*, vol. 52, no. 2, pp. 174–182, 2019.
[8] T. Vogels *et al.*, "Fully-automatic camera-based pulse-oximetry during sleep," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPM) Workshops*, Salt Lake City, UT, USA, June 2018, pp. 1349–1357.
[9] K. De Miguel *et al.*, "Home camera-based fall detection system for the elderly," *Sensors*, vol. 17, no. 12, p. 2864, 2017.
[10] S. Leonhardt *et al.*, "Unobtrusive vital sign monitoring in automotive environments - a review," *Sensors*, vol. 18, no. 9, p. 3080, 2018.

[11] W. Wang *et al.*, "Discriminative signatures for remote-PPG," *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2019. DOI = 10.1109/TBME.2019.2938564.
[12] Y. Kanzawa *et al.*, "Human skin detection by visible and near-infrared imaging," in *IAPR Conference on Machine Vision Applications*, vol. 12, Nara, Japan, June 2011, pp. 14–22.
[13] E. M. Nowara *et al.*, "SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, June 2018, pp. 1353–135 309.
[14] J. Hu *et al.*, "Illumination robust heart-rate extraction from single-wavelength infrared camera using spatial-channel expansion," in *Proc. IEEE Conf. Eng. Med. Biol. Soc. (EMBS)*, Berlin, Germany, July 2019.
[15] W. Wang *et al.*, "Single element remote-PPG," *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2018.
[16] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
[17] W. Wang *et al.*, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, July 2017.
[18] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiol. Meas.*, vol. 35, no. 9, pp. 1913–1922, Oct. 2014.
[19] M. van Gastel *et al.*, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.
[20] W. Wang *et al.*, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1974–1984, Sept. 2016.
[21] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, Oregon, USA, June 2013, pp. 532–539.
[22] W. Wang *et al.*, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015.
[23] A. V. Moco *et al.*, "New insights into the origin of remote PPG signals in visible light and infrared," *Scientific Reports*, vol. 8, no. 1, May 2018.