Discriminative Signatures for Remote-PPG

Wenjin Wang, Albertus C. den Brinker, and Gerard de Haan

***Abstract*—Near-infrared (NIR) remote photoplethysmography (PPG) promises attractive applications in darkness, as it involves unobtrusive, invisible light. However, since the PPG strength (AC/DC) is much lower in the NIR spectrum than in the RGB spectrum, robust vital signs monitoring is more challenging. In this paper, we propose a new PPG-extraction method, DIScriminative signature based extraction (DIS), to significantly improve the pulse-rate measurement in NIR. Our core idea is to use both the color signals containing blood absorption variations and additional disturbance signals as input for PPG extraction. By defining a discriminative signature, we use one-step least-squares regression (joint optimization) to retrieve the pulsatile component from color signals and suppress disturbance signals simultaneously. A large-scale lab experiment, recorded in NIR with heavy body motions, shows the significant improvement of DIS over the state-of-the-art method, whereas its principle is simple and generally applicable.**

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

I. Introduction

CAMERA-based remote photoplethysmography (remote-PPG) enables contactless measurement of the blood volume pulse by detecting the pulse-induced subtle color changes from the human skin surface [1]. It can be used for various healthcare applications (e.g. patient monitoring [2], neonate monitoring [3], sleep monitoring [4], elderly care [5], cardio-fitness training [6], driver monitoring in automotive [7], etc.) to measure different vital signs (e.g. pulse rate (variability), respiratory rate, blood oxygen saturation, pulse transit time, etc.) from human face and body.

Recent remote-PPG publications [8]–[10] reported excellent performance of pulse-rate monitoring in visible light using RGB cameras, and even demonstrated encouraging robustness in challenging fitness applications [6]. In contrast, the efforts and progress made in invisible light are still at a lower level [11], though it is also an important application scenario for contactless health monitoring, especially for clinical applications that require long-term continuous monitoring (24/7) including night, such as in Intensive Care Unit (ICU), Neonatal Intensive Care Unit (NICU) and Coronary Care Unit (CCU). The most recent studies in near-infrared (NIR) focus on relatively stable cases such as sleep monitoring [4], with a controlled environment (e.g. stable lighting condition) and a compliant subject that has little body motions. In general, remote-PPG in challenging NIR applications involving subject body motions is considered less mature/validated, such as patient monitoring in emergency department triage or driver monitoring in automotive where more motion disturbances are expected than during sleep. This paper addresses the above, in particular it sets out to *improve the robustness of remote pulse-rate monitoring in NIR in case of heavy body motion*. Though using the challenging NIR application as the showcase, the concept is also applicable to RGB.

Extracting the PPG signal in NIR is much more challenging than in RGB, mainly due to two physiological reasons: (i) the pulsatile strength (AC/DC) in NIR is in general much lower than in RGB except the R-wavelength, due to the low spectral absorption of (de)-oxygenated hemoglobin in the NIR range; (ii) the pulsatile contrast between different NIR wavelengths (> 750 nm that excludes red) is smaller, because the blood absorption spectrum in NIR is flatter than in RGB, i.e. it is even flatter for de-oxygenated blood (low SpO2). These two conditions lead to two unfavorable consequences: (i) the pulsatile component is easily disrupted by strong disturbances (e.g. heavy body motion); (ii) the pulsatile component is difficult to be separated from disturbances in the (DC-normalized) color space [8] as their color variation directions are similar, i.e. the main disturbance in NIR is the intensity variation (i.e. less specular variation as light can penetrate deeper into the skin) and its color variation direction is closer to the pulsatile direction in NIR than in RGB [9].

A recent sleep study in NIR [4] suggested that the blood volume pulse signature based method (PBV [9]) is a reliable approach for multi-wavelength NIR application, though its performance is worse in NIR than in RGB. The essence of PBV is using a blood volume pulse signature (i.e. normalized AC/DC-amplitude vector across multiple wavelengths) to retrieve the PPG signal from the color signals by one-step least-squares regression. In addition to the PPG-extraction method, different pre-processing and post-processing methods [12]–[14] have been introduced to improve the PPG quality in a separate step, such as using extra motion signals to further suppress the motion frequency in the extracted PPG signal [15].

Inspired by the PBV method [9] and existing de-noising strategies based on disturbance signals, we propose a new PPG-extraction method, called DIScriminative signature based extraction (DIS), that explicitly integrates pulse retrieval and disturbance suppression into a single process for joint optimization. Different from earlier works, we use both the color signals[1] and disturbance signals (e.g. motion) as input for PPG extraction. Correspondingly, we extend the blood volume pulse signature to include the disturbance channels next to the color channels, which is a discriminative signature. Next, we use one-step least-squares regression to retrieve the pulsatile

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

[1]The "color signals" in this paper refer to the wavelength signals extracted from the skin pixels that contain blood absorption variations. It can be RGB signals or NIR signals (pseudo-color information).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2938564, IEEE Transactions on Biomedical Engineering

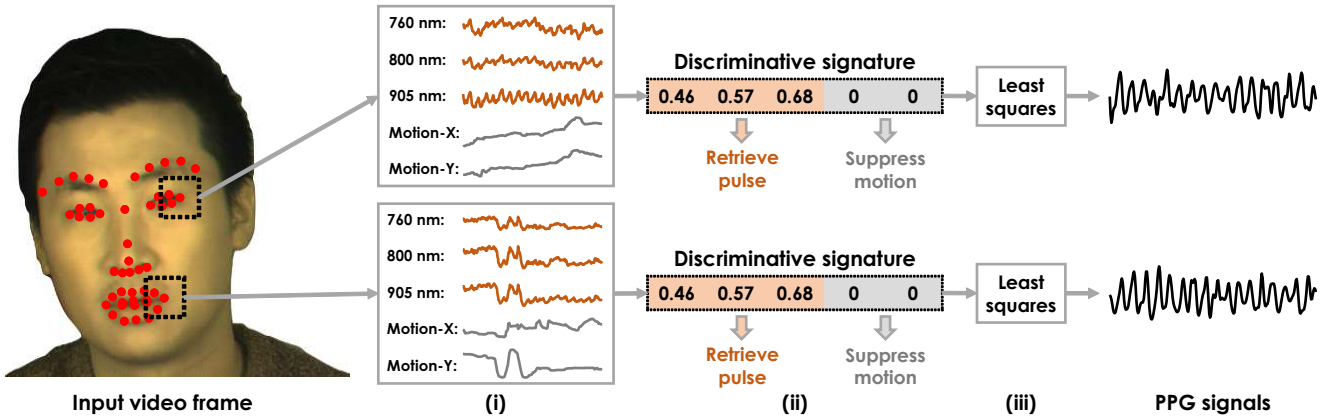

Fig. 1. The essence of the proposed DIS method can be interpreted in three steps: (i) generate both the color signals and disturbance signals (e.g. motion traces) from the (local) skin area; (ii) define a discriminative signature that promotes the pulsatile component from the color signals and suppresses disturbance components; and (iii) perform one-step least-squares regression to extract the pulse from the input signals (e.g. 5 channels in our case). There are multiple ways to define disturbance signals for the step (i), motion traces from facial landmarks is used for illustration purpose here.

component from the color signals, and at the same time, suppress the disturbance signals (noise). The flowchart of DIS is shown in Fig. 1. This idea is based on the observation that artifacts in color signals have similarity with input disturbance signals in terms of the signal waveform, as they have the same origin (e.g. skin motion).

The elegance of DIS is that we put two separate steps, PPG extraction and PPG de-noising, into a single step for joint optimization, with a closed-form solution in terms of least-squares regression, where extraction and de-noising can facilitate each other. From both the analytical and experimental perspective, we prove that one-step (joint) optimization is better than sub-steps (separate) optimizations (e.g. de-noising + extraction or extraction + de-noising). Moreover, we show two different approaches to create disturbance signals: local motion features and spatial statistics of the color channels. We create a large benchmark dataset, consisting of lab recordings made in NIR with heavy body motions, to validate the proposed method. The benchmark shows significant improvement of DIS over the state-of-the-art remote-PPG extraction method in NIR (i.e. PBV [9]). We stress that the benchmark was executed in a lab environment with healthy volunteers for the proof-of-concept validation of the core PPG-extraction algorithm, which is not yet the clinical validation of the full-fledged monitoring system on real patients with 24/7 recordings. What is reported in the following is a first step towards this goal. The clinical trials shall be organized after the technical proof, as the future work.

The remainder of this paper is structured as follows. In Section II, we describe the DIS method and its principles. In Section III, we specify our expectations of DIS. In Sections IV and V, we use a large benchmark dataset recorded in lab to verify DIS. Finally, in Section VII, 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, and $^\top$ denotes transposition.

A. Signal model

Consider a piece of human tissue illuminated by a light source (visible and/or infrared) and observed by a remote sensor system. The sensor is a camera with multiple channels particularly sensitive to specific wavelength ranges. According to the dichromatic skin reflection model [8], the sensor will pick up specular light directly reflected from the skin surface and diffuse light penetrating into the skin, being partially absorbed and scattered in the tissue. The light penetrating into the skin is influenced by the changing blood volume and thus its reflection carries blood volume pulse information. This information is what we want to measure.

The geometry of the light source, skin and camera affects the recorded light: the movement of the skin influences the collected signals. The main additional signal components [8] can be described as intensity variations (associated with movement of the skin towards/from camera/light source) and specular variations (which can change dramatically with small angular changes out of the recording plane).

We denote the multi-wavelength camera signals obtained at time t as $\mathbf{C}(t)$ with entries $C_i(t)$, $i \in \{1, I\}$ where I is the number of channels. As a general model we describe $\mathbf{C}(t)$ as:

$$\mathbf{C}(t) = \mathbf{v_s} \cdot I_0 + \mathbf{v_p} \cdot p(t) + \sum_{m=1}^{M} \mathbf{u_m} \cdot q_m(t), \qquad (1)$$

where I_0 is the average intensity, $p(t)$ is the pulse-signal source (i.e. the signal of interest); $q_m(t)$ are additional non-pulsatile signal sources; $\mathbf{v}_s$ is the color direction associated with the average skin color (DC); $\mathbf{v_p}$ is the color variation direction induced by blood pulsation; and $\mathbf{u_m}$ are the color variation directions associated with various (M) additional sources. The additional sources are typically movement related but may encompass other effects as well, such as intensity modulation of the light.

For an RGB camera there are three channels, but we will allow an arbitrary number of channels denoted as I. Since the PPG extraction needs to be independent of the absolute DC level (e.g. light intensity, skin color) to enable the use of prior/assumption on the color variation directions of sources,

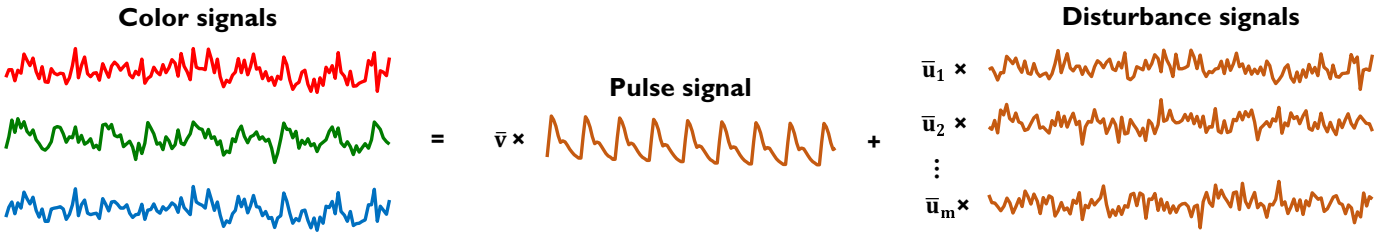

Fig. 2. Illustration of the model (3), in which the camera color signals are a weighted combination of the pulse signal and disturbance signals after DC normalization. The existing PPG-extraction methods exploiting (3) either use the assumption on the color vector of the pulse ($\bar{\mathbf{v}}$) [9], [16] or the assumption on the color vectors of disturbances ($\bar{\mathbf{u}}_\mathbf{m}$) [8], [17]. In this paper, we explore the third option that uses disturbance signals to facilitate the pulse extraction.

we first normalize each color channel by its average intensity and subtract 1 to center the AC/DC signals around 0. This results in zero-mean AC/DC signals. This step is in line with the DC normalization in existing remote-PPG works [8], [9], [17]. It can be achieved by normalizing $\mathbf{C}(t)$ with a diagonal matrix $\mathbf{R}$ such that $\mathbf{v_s} \cdot I_0 \cdot \mathbf{R} = \mathbf{1}$, thus creating zero-mean AC/DC signals $\tilde{\mathbf{C}}(t)$ as:

$$\tilde{\mathbf{C}}(t) = \mathbf{C}(t) \cdot \mathbf{R} - 1 = \mathbf{v_p} \cdot \mathbf{R} \cdot p(t) + \sum_{m=1}^{M} \mathbf{u_m} \cdot \mathbf{R} \cdot q_m(t). \quad (2)$$

To simplify the notation, we denote the mapped/normalized color variation directions as $\mathbf{v_p} \cdot \mathbf{R} = \bar{\mathbf{v}}$ and $\mathbf{u_m} \cdot \mathbf{R} = \bar{\mathbf{u}}_\mathbf{m}$, giving:

$$\tilde{\mathbf{C}}(t) = \bar{\mathbf{v}} \cdot p(t) + \sum_{m=1}^{M} \bar{\mathbf{u}}_\mathbf{m} \cdot q_m(t), \quad (3)$$

where $\bar{\mathbf{v}}$ and $\bar{\mathbf{u}}_\mathbf{m}$ are DC-normalized color variation directions of blood pulsation and disturbances, respectively, i.e. $\bar{\mathbf{v}}$ is also referred to as the blood volume pulse signature [9]. Fig. 2 illustrates the model of (3). To attain uniqueness of the constituent multiplicative components, normalization is needed. We arbitrarily choose ℓ^2 normalization of the mapped color vectors such that $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}^\top = 1$ and $\bar{\mathbf{u}}_m \cdot \bar{\mathbf{u}}_m^\top = 1$.

There are two categories of approaches to address the model of (3) for extracting the pulse signal $p(t)$.

1) PPG-knowledge based methods (PBV, SoftSig): We can use the knowledge of the blood volume pulse signature ($\bar{\mathbf{v}}$) to directly retrieve the pulse by least-squares projection. This typically requires accurate assumption on the prior $\bar{\mathbf{v}}$ [9]. We can also soften the dependency on the accuracy of $\bar{\mathbf{v}}$ by using multiple signatures and selecting the best estimate [16].

2) disturbance-knowledge based methods (CHROM, POS): Alternatively, we can use the knowledge of normalized color vectors of the disturbance signals ($\bar{\mathbf{u}}_\mathbf{m}$) to derive the pulse by projection and tuning [8], [17]. CHROM uses the specular variation direction as the main disturbance direction to design an orthogonal projection system/axes (only valid in the visible light) [17], while POS uses the intensity variation direction as the main disturbance to design such a system (valid in both the visible and invisible light) [8].

These PPG extraction methods require prior knowledge on the color direction of a particular signal component. Such knowledge is not available for the disturbance signals. However, the disturbance signals $q_m(t)$ can be measured separately in a quantitative way and used to facilitate the PPG extraction.

To this end, we consider the signal components in (3) in more detail. The assumption is a stationary environment in which the signals are observed and we consider the pulse and additional sources at a certain moment as stochastic signals. It is assumed that the pulse signal is independent of the additional disturbance signals: they are neither causally related nor ruled by the same driving force. Since these signals are zero-mean and independent, we have:

$$\mathcal{E}\{p(t) \cdot q_m(t)\} = 0, \quad (4)$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. There is a known phenomenon that may render (4) invalid: ballistocardiographic (BCG) motion, the subtle head oscillations caused by the blood pulsation from heart to head via the abdominal aorta and the carotid arteries [18]. BCG motion is typically much smaller than all other body motions (even unintentional body motions like cough), and therefore the associated color changes are typically negligible in non-stationary use cases. They are only observable when the subject tries to remain still. In that case, all signal components (both $p(t)$ and $q_m(t)$) will be in sync and any projection will create a clean signal where the fundamental frequency corresponds to the pulse rate, implying that for pulse-rate extraction BCG motion is unlikely to be an interfering component.

B. Extended signal set

We now turn to the core idea of this paper: the extension of the observation signal to include disturbance signals in the associated PPG extraction method. In the extended signal set, we do not only observe I color signals, but also a number (J) of disturbance signals, i.e. the disturbance signals are also measured from the video such as skin motion.

We define the extended signal set as $\mathbf{S}(t)$, which is a matrix containing I color signals $\mathbf{C}(t)$ and J disturbance signals $\mathbf{D}(t)$ that has entries $D_j(t)$, $j \in \{1, J\}$. $\mathbf{S}(t)$ is defined as:

$$S_i(t) = \begin{cases} C_i(t), & \text{if } i \in \{1, I\}, \\ D_{i-I}(t), & \text{if } i \in \{I+1, I+J\}. \end{cases} \quad (5)$$

Similar to the earlier process of (2), we normalize the DC of each signal in $\mathbf{S}(t)$ and center it at zero, obtaining $\tilde{\mathbf{S}}(t)$.

The signal model still has the same driving forces as before ($p(t)$ and $q_m(t)$), yet the color vectors have to be extended. In particular we take:

$$\tilde{\mathbf{S}}(t) = \bar{\mathbf{e}} \cdot p(t) + \sum_{m=1}^{M} \bar{\mathbf{f}}_\mathbf{m} \cdot q_m(t), \quad (6)$$

where $\bar{\mathbf{e}}$ and $\bar{\mathbf{f}}_\mathbf{m}$ are the color vectors in the extended $(I+J)$-dimensional space. We define disturbance signals as signals which are a function of one or more components $q_m(t)$ but not dependent on $p(t)$. This has the following consequences. The color vector $\bar{\mathbf{e}}$ equals $\bar{\mathbf{v}}$ extended with a number of J zero entries. We note that it still has unit norm: $\bar{\mathbf{e}}\cdot\bar{\mathbf{e}}^\top = 1$. It implies that the disturbance signals have to be zero-mean, which is ensured by our algorithm that uses the same normalization on the disturbance signals as for the color signals. As a consequence of (4), we have:

$$\mathcal{E}\{p(t)\cdot\tilde{\mathbf{S}}(t)^\top\} = \bar{\mathbf{e}}\cdot\sigma_p^2, \tag{7}$$

where $\sigma_p^2 = \mathcal{E}\{p(t)^2\}$ is the variance of $p(t)$. This property will be used by the following algorithm to extract $p(t)$.

C. Extraction of the pulse signal

Extraction of the pulse signal can now be done equally to the PBV method [9], i.e. constructing an optimal projection vector $\mathbf{w}$ that projects the DC-normalized observation signals $\tilde{\mathbf{S}}(t)$ onto an approximation $\hat{p}(t)$ of $p(t)$:

$$\hat{p}(t) = \mathbf{w}\cdot\tilde{\mathbf{S}}(t)^\top \propto p(t). \tag{8}$$

The optimality is defined in a least-squares sense by minimizing:

$$\min_w \sum_{t=1}^{N}(\mathbf{w}\cdot\tilde{\mathbf{S}}(t)^\top - p(t))^2. \tag{9}$$

In matrix notation, the solution for $\mathbf{w}$ can be written in the closed form [19] as:

$$\mathbf{w} = (\mathbf{p}\cdot\tilde{\mathbf{S}}^\top)\cdot(\tilde{\mathbf{S}}\cdot\tilde{\mathbf{S}}^\top)^{-1}. \tag{10}$$

In contrast to the ordinary linear regression case, $\mathbf{p}$ is unknown to us but the model aids us here. From (7) and the definitions of $\mathbf{p}$ and $\tilde{\mathbf{S}}$, it follows:

$$\mathcal{E}\{\mathbf{p}\cdot\tilde{\mathbf{S}}^\top\} = \bar{\mathbf{e}}\cdot\sigma_p^2\cdot N, \tag{11}$$

where N is the number of time samples (i.e. length of the signal). Substituting (11) and (10) into (8), we obtain:

$$\begin{aligned}\hat{\mathbf{p}} &= \mathbf{w}\cdot\tilde{\mathbf{S}} = (\mathbf{p}\cdot\tilde{\mathbf{S}}^\top)\cdot(\tilde{\mathbf{S}}\cdot\tilde{\mathbf{S}}^\top)^{-1}\cdot\tilde{\mathbf{S}}\\ &= \bar{\mathbf{e}}\cdot(\tilde{\mathbf{S}}\cdot\tilde{\mathbf{S}}^\top)^{-1}\cdot\tilde{\mathbf{S}}\cdot\sigma_p^2\cdot N\\ &\propto \bar{\mathbf{e}}\cdot(\tilde{\mathbf{S}}\cdot\tilde{\mathbf{S}}^\top)^{-1}\cdot\tilde{\mathbf{S}},\end{aligned} \tag{12}$$

which is a least-squares fit for the pulse signal with arbitrary gain.

Lastly, we elaborate on the definition of $\bar{\mathbf{e}}$, i.e. the discriminative signature that extends the blood volume pulse signature $\bar{\mathbf{v}}$ with zero entries corresponding to the disturbance channels. $\bar{\mathbf{v}}$ can directly be measured if the observed skin is disturbance-free: in that case the different wavelength channels contain clean PPG signals and the ratios of their PPG strengths can be used to determine $\bar{\mathbf{v}}$. Alternatively, $\bar{\mathbf{v}}$ can be calculated by integrating the PPG absorption spectrum, the camera filter response, and the light spectrum. Details of this calculation can be found in [9]. As the focus of this paper is in NIR, we take a NIR setup (see Fig. 3) to illustrate how to specify $\bar{\mathbf{e}}$, and this setup will be used later for benchmarking. The setup has

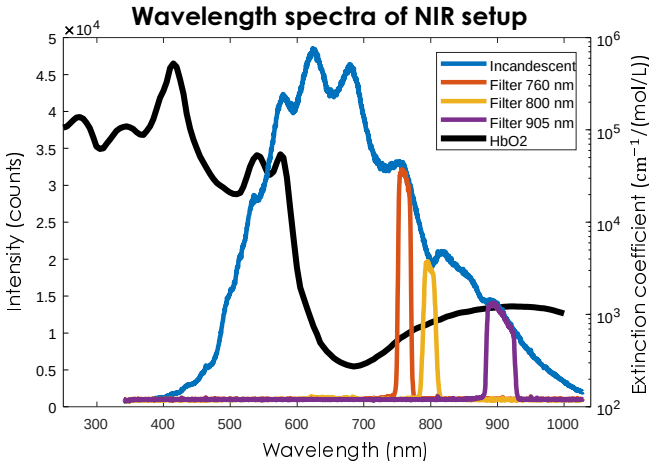

Fig. 3. The wavelength spectra of different components in a NIR setup, including the spectra of incandescent light, NIR filter responses, and oxygenated-hemoglobin (HbO2).

a 3-wavelength NIR camera[2] and a incandescent illumination source. Integrating the optical spectra shown in Fig. 3 results in $\bar{\mathbf{e}} = [0.46, 0.57, 0.68, 0, 0]$, where two motion channels are assumed as the disturbance source (i.e. vertical and horizontal motion). For RGB, $\bar{\mathbf{e}}$ can be constructed in a similar way. If a consumer-grade RGB camera with broad-/overlapping-band filter is used, the integration will use the whole filter response (not only its center/peak response).

There are two major differences between the PPG-extraction method proposed in this paper and PBV: (i) the input signals include both the color signals and disturbance signals, (ii) the signature is an extension of the blood volume pulse signature that explicitly deals with disturbances. The essence of our approach is using the waveform of the disturbance signals to guide the least-squares estimation of the projection vector, such that the projection can eliminate the disturbance waveform in the resulting signal. Therefore, we call it DIScriminitative signature based PPG extraction, in short DIS. The model and processing of DIS are shown in Fig. 1.

The core algorithm of DIS is shown in Algorithm 1, which can be implemented in a few lines of Matlab code (this example assumes 3 color signals and 2 disturbance signals as the input). We mention that a band-pass filtering step has been added to pre-process the AC/DC signals (e.g. $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$) to remove the clear out-band noise, which allows the least-

[2]The 3-wavelength NIR camera is constructed by multiple monochrome cameras with narrow band NIR filters centered at 760 nm (bandwidth 20 nm), 800 nm (bandwidth 20 nm) and 905 nm (bandwidth 33 nm).

Algorithm 1 Discriminative signature (DIS)

Input: The $3\times N$ color signals $\mathbf{C}$ and $2\times N$ disturbance signals $\mathbf{D}$, where N is the time dimension.

1: **Initialize:** $\bar{\mathbf{e}} = [0.46, 0.57, 0.68, 0, 0]$ (setup in this paper)
2: $\mathbf{S} = [\mathbf{C}; \mathbf{D}]$;
3: $\tilde{\mathbf{S}} = \mathbf{diag}(\mathbf{mean}(\mathbf{S}, 2))^{-1} * \mathbf{S} - 1$; $\rightarrow$ DC normalization
4: $\tilde{\mathbf{S}} = \mathbf{bandpass}(\tilde{\mathbf{S}}, [36, 240]\text{bpm})$; $\rightarrow$ Band-pass filtering
5: $\hat{\mathbf{p}} = \bar{\mathbf{e}} * \mathbf{pinv}(\tilde{\mathbf{S}} * \tilde{\mathbf{S}}^\top) * \tilde{\mathbf{S}}$; $\rightarrow$ Least-squares projection

Output: The $1\times N$ pulse signal $\hat{\mathbf{p}}$.

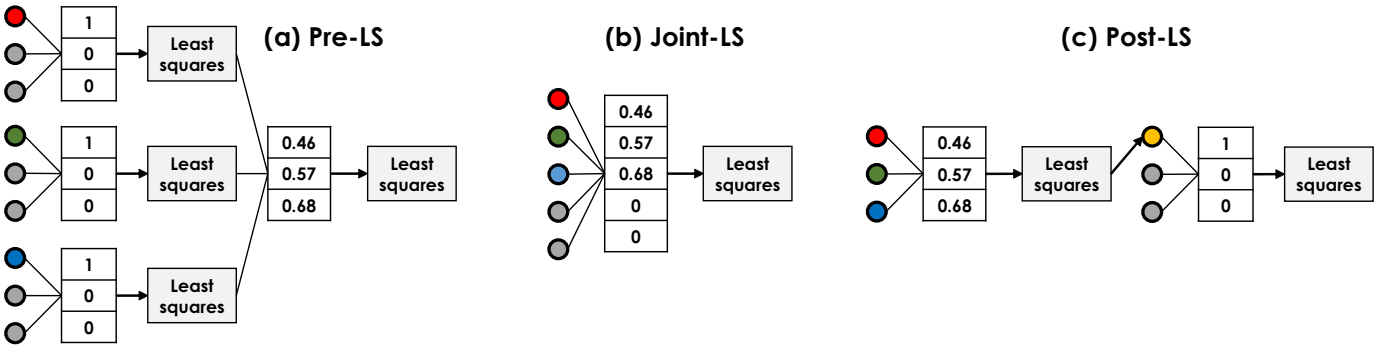

Fig. 4. Different optimization schemes using color signals (denoted by RGB circles) and disturbance signals (denoted by gray circles): (a) Pre-LS: first eliminate motion from each color signal and then combine motion-free color signals; (b) Joint-LS: combine color signals and eliminate motion simultaneously (our proposed method); and (c) Post-LS: first combine color signals and then eliminate motion from the combined (PPG) signal. Among these, (a) and (c) are two-step optimizations, (b) is one-step optimization. The shown signature [0.46, 0.57, 0.68] is based on the NIR setup of Fig. 3.

squares channel combination to be focused on addressing the in-band disturbances. We stress that the same band-pass pre-processing will be used for other benchmarked algorithms in the following section for fair comparison.

III. Assumptions and Expectations

The above framework sketches a discriminative signature based PPG extraction that extends PBV using a number of additional uncorrelated (disturbance) sources, thereby raising a couple of qualitative and quantitative questions: *why would it bring improvements?* and *how significant are the improvements?* The quantitative aspect is saved for later experiments, let's first consider why this approach can bring a benefit.

A. Joint optimization or sub-steps optimization?

As a first consideration, we mention that noise reduction techniques have already been deployed for vital signs extraction, as pre-processing of the input color signals or post-processing of the output pulse signal, where especially the motion signals were used [14]. The motion signals are assumed to influence the color signals as additive noise sources, while representative of physical disturbances in the real world (e.g. skin motion) [8]. This is the fundamental assumption of all existing remote-PPG algorithms that use linear channel combination for source de-mixing, including Blind Source Separation (BSS). From a least-squares optimization point of view, it is always better to do a joint optimization instead of splitting the optimization into two separate steps: only in the first case one fully exploits all knowledge on the covariance of various input signals. Splitting is usually considered as greedy form of optimization to reduce processing power.

To be more specific, given the same objective function for optimization (e.g. ℓ^2-norm minimization), joint optimization can find the global convex in the joint multi-dimensional space/channels of color and disturbance. In contrast, sub-steps optimization first finds one convex in a partial space (e.g. color space), and then, based on this convex, approaches another convex in the remaining space (e.g. disturbance space), or the other way around, depending on the order of optimization (e.g. pre-/post-processing). There is no guarantee that it can always find the global convex. It can also be understood as: the color signals and disturbance signals in a joint covariance matrix are fully correlated/connected, whereas in the covariance matrix of sub-steps optimization are partially correlated. Our following experiments will compare joint least-squares optimization (Joint-LS, same as DIS) and sub-steps optimization (e.g. Pre-LS + PBV and PBV + Post-LS), as illustrated in Fig. 4.

B. What is the benefit of joint optimization?

The second consideration is probably more relevant from a practical point of view. To start with, we note that the pulsatile component in the original color signals $\mathbf{C}$ is minute: the DC-normalization of (2) is essential, not free from errors and yields, in practical situations, the color variation signal $\tilde{\mathbf{C}}$ where the desired component $\mathbf{p}$ is smaller than disturbances like intensity variations [8]. In contrast, a motion signal can be easily and accurately obtained without doing much processing (e.g. using spatial coordinates of a face tracker). Having such a well-determined signal which is known to be orthogonal to the desired signal, makes the effect of (small) mis-matches between the assumed signature $\bar{\mathbf{v}}$ and the actual occurring ones much smaller. This is what the authors consider the main benefit of DIS: the uncorrelated disturbance signals act as beacons to steer away from undesired projection solutions, thereby limiting the influence of using a not fully matching/accurate blood volume pulse signature to that of the actual use case.

The last notion generates a new question namely how to balance the number of color and disturbance signals, I and J. It is obvious that we need at least one color signal, although then the sketched difference with pre-/post-processing is lost (i.e. similar to the single-channel de-noising). For dual-wavelengths PPG, as is common in contact-PPG solutions using red and infrared wavelengths (e.g. finger pulse oximetry), it is already a real option. To gain insights into this issue, our experiments will validate various options for the number of wavelength channels, such as 1 color channel + 1 disturbance channel, 1 color channel + 2 disturbance channels, 2 color channels + 2 disturbance channels, 3 color channels + 2 disturbance channels, etc.

C. How to define a disturbance source?

As for the disturbance signals, we use the motion signals derived from the spatial location of facial landmarks (e.g. (x, y)

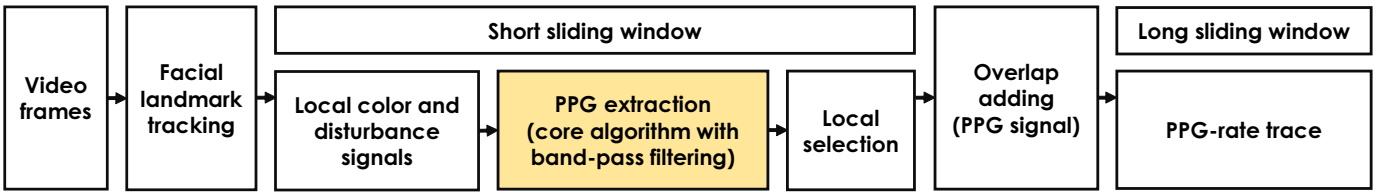

Fig. 5. The architecture of the vital signs monitoring system used for pulse-rate extraction. The input are the video frames and the output is the pulse-rate trace. The step of PPG extraction (yellow box) is the key component benchmarked in this paper, where different core remote-PPG algorithms are compared. The other steps are fixed across the benchmarked methods.

landmark coordinates in Fig. 1). In a 3-wavelength camera setup, this means that we have 3 color channels and 2 motion channels (e.g. horizontal motion and vertical motion). Another question is: are there, next to motion signals, other disturbance signals adhering to the given constraints and readily available? Inspired by [20], we can also use the standard deviation of spatial pixel values to generate disturbance signals. It shows complementary effect w.r.t. the mean of spatial pixel values (i.e. the way of generating $\mathbf{C}$) [20], i.e. if the spatial mean color signal shows pulse, the spatial standard deviation color signal will show non-pulse. This means that a 3-channels $\mathbf{D}$ can be created, leading to a 6-channels $\mathbf{S}$ that holds first- and second-order statistics of the spatial color characteristics.

Our experiments will benchmark two different approaches for creating disturbance signals, one is based on the facial landmarks and the other on the spatial standard deviation of skin pixels. We note that not limited to these two approaches, there are further alternatives to generate disturbance signals, such as using the light intensity changes.

IV. Experimental setup

This section presents the experimental setup for the benchmark. First, we introduce the benchmark dataset. Next, we discuss the compared remote-PPG algorithms and their settings. Finally, we present the evaluation metrics.

A. Benchmark dataset

This paper targets one of the most challenging application scenarios for remote-PPG: NIR heavy motion. Thus we create a benchmark dataset using a NIR camera setup built by the following components. The experiments are executed in the lab with healthy test subjects for the proof-of-concept validation of the PPG-extraction algorithms, i.e. it is not a clinical trial with a full-fledged monitoring system.

• **Camera** The NIR setup uses three monochrome cameras to sample different NIR wavelengths for a multi-spectral measurement. The camera type is: Global shutter Manta G-283 of Allied Vision Technologies GmbH (Sony ICX674ALG, CCD sensor), with 968×728 pixels, 8 bit depth, and 15 fps). Three separate monochrome cameras use different optical filters specified in Fig. 3. The camera lens is Schneider-Kreuznach Tele-Xenar 1:4/150. The cameras are placed around 4 m in front of the subject to reduce the parallax[3] between individual cameras. With the used focal length, it results in approximately 20-30% skin area in a video frame. All auto-adjustment functions (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 an *uncompressed* format at constant frame rate.

• **Light source** The illumination sources are two incandescent light fixtures consisting of 9 lamps each. Each fixture is powered at $220\,\mathrm{V} \times 1.2\,\mathrm{A} = 264\,\mathrm{W}$, providing sufficient energy for the NIR sensing. The spectrum of the used incandescent lamp is shown in Fig. 3. The spectral responses of the NIR filters, with the applied light spectrum, are also shown in this figure.

• **Ground-truth** The ground-truth is the raw PPG signal recorded by a finger-based transmissive pulse oximetry (Model CMS50E from ContecMedical), which is synchronized with the video acquisition using time stamps. The PPG reference is electrically decoupled from the video recording system. The reference 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. Corrupted segments in the reference PPG signal (e.g. due to body motions) are manually annotated and removed, as the ground-truth must be correct. The amount of excluded PPG-reference data (due to artifacts) is very little (see Figs. 7-8): it is less than 1% of the total measurement. We mention that the reason of choosing finger-PPG sensor as the reference instead of ECG is that we would like to benchmark the camera-PPG with contact-PPG as they measure the same physiological origin - blood absorption variations. Also, the purpose of camera-PPG is to replace contact-PPG, but not ECG (i.e. ECG can measure other cardiac features/activities that PPG cannot measure). The reason for using finger-PPG instead of earlobe-PPG is that the subject was asked to perform severe head motions that may corrupt the earlobe-PPG signal. In contrast, finger-PPG is much less affected by head motion.

• **Subjects** A total of 26 subjects (aged between 25 and 55, 19 male and 7 female), with different skin tones categorized from type I to type V according to the Fitzpatrick scale (15 subjects in type I-II, 5 subjects in type III, 6 subjects in type IV-V), participate in recordings. Skin type VI is not present

[3]Parallax is the displacement in the apparent position of an object viewed along different optical paths. To reduce the parallax, image planes from different monochrome cameras are first aligned manually using a real-time overlay feedback. Next, image planes are registered to the central camera plane (reference) using a similarity transformation model consisting of translation, rotation and scaling. The model is estimated from the initial frames based on epipolar geometry between cameras.

among the participants, but this is not considered as a limiting factor for the validity of conclusions for the following reasons. Coverage of the Fitzpatrick scale is important for visible light conditions but since NIR light can penetrate deeper into the skin (reaching arterioles) and the melanin absorption (including eumelanin and pheomelanin) is much lower and flatter in NIR than in RGB [21] (i.e. flatter means less wavelength dependency), skin-type differences are expected to have much less pronounced effects on PPG extraction in NIR.

This study has been approved by the Internal Committee Biomedical Experiments of Philips Research, and informed consent has been obtained from each subject. During the recording, the subject was asked to (i) move freely with a mixture of different types of head motions (e.g. rotation, translation and scaling) and facial expressions (e.g. smiling and talking) continuously. There is no instruction for the subject to perform a single type of motion like periodic head rotation, as these are not particularly relevant in practice. The continuous motion phase constitutes around 80% time per recording; (ii) stay relaxed without significant motion at the end of the recording (20% time per recording). Each video recording is around 8-10 minutes according to the protocol.

Snapshots of video recordings are exemplified in Figs. 7-8, i.e. the images are created by plugging the three NIR channels into an RGB rendering system.

B. Compared methods

We benchmark 6 core remote-PPG algorithms that can be used in the NIR condition: (i) PCA [22], (ii) ICA [23], (iii) POS (with adapted projection-axes for NIR) [8], (iv) PBV [9], (v) DIS-M (facial landmark as the motion source), (vi) DIS-C (spatial color standard deviation as the noise source). Note that both DIS-M and DIS-C belong to the DIS-family.

All compared core remote-PPG algorithms are plugged into the same (basic) vital signs monitoring architecture for pulse-rate trace extraction (see Fig. 5). The processing architecture includes a facial landmark tracker [24], a short sliding window (with 1 frame sliding step and 64 frames[4] window length) for extracting the local color and disturbance signals[5], a core PPG extraction algorithm with a band-pass filter[6] to extract local PPG signals, a Signal-to-Noise-Ratio (SNR) based metric for local PPG selection (i.e. the candidate with the maximum SNR is selected as the final output), and an overlap-adding procedure (similar to [8], [17]) to generate a long pulse signal by overlap-adding the pulse intervals measured from the short sliding window. Based on the extracted pulse signal, we use a long sliding window with 1 sample sliding step and 256 samples window length (around 17 s for 15 fps camera) to generate the pulse-rate trace. The pulse rate in the long sliding window is calculated in the frequency domain by using the frequency index of the maximum spectral peak of the pulse signal within the window. The frequency-based pulse-rate estimation has been used for both the camera and contact sensors to obtain PPG-rate trace for comparison.

All methods 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.

[4]The reason for setting the sliding window length to 64 is that the processing includes the SNR-based local PPG selection, where SNR is calculated based on the frequency spectrum. Since FFT is used to compute the frequency spectrum (without zero-padding), we define the signal length to be the power of 2 (e.g. $2^6 = 64$), in line with [8], [9], [17]. Given a 15 or 20 fps video camera, this window length allows quick adaption of disturbance suppression when combining color channels for PPG signal extraction [6].

[5]The local skin signals are generated by: (i) using the facial landmark tracker to detect and track the facial landmarks; (ii) using a local patch (with 30×30 pixels) centered at each landmark to extract the color signals (i.e. spatially averaged pixel values concatenated in time) and disturbance signals (i.e. landmark coordinates concatenated in time).

[6]A band-pass filter is used to pre-process the AC/DC signals for reducing clear out-band disturbances. It is the same filter used for processing the reference PPG, i.e. a 4-th order Butterworth band-pass filter with cutoff frequencies [0.6, 3] Hz operated in a zero-phase forward and reverse filtering mode.

TABLE I
AVERAGE EVALUATION METRIC VALUES OBTAINED BY 6 REMOTE-PPG METHODS. BOLDFACE ENTRIES DENOTE THE BEST RESULT IN EACH ROW.

Metric	PCA	ICA	POS	PBV	DIS-M	DIS-C
RMSE (bpm)	24.6	24.4	20.5	11.8	5.9	**5.5**
SR-AUC	0.30	0.31	0.40	0.65	0.77	**0.79**
Coverage (%)	30.0	31.3	41.5	69.0	82.1	**84.7**

TABLE II
AVERAGE EVALUATION METRIC VALUES OBTAINED BY PBV, DIS-M AND DIS-C ON DIFFERENT WAVELENGTH COMBINATIONS.

Metric	Wavelength (nm)	PBV	DIS-M	DIS-C
RMSE (bpm)	905	21.8	**14.8**	20.0
	[800, 905]	16.9	**9.4**	11.1
	[760, 905]	16.3	**8.9**	9.9
	[760, 800, 905]	11.8	5.9	**5.5**
SR-AUC	905	0.36	**0.52**	0.41
	[800, 905]	0.52	**0.68**	0.66
	[760, 905]	0.56	**0.71**	0.70
	[760, 800, 905]	0.65	0.77	**0.79**
Coverage (%)	905	37.3	**52.8**	42.8
	[800, 905]	54.1	**72.1**	69.2
	[760, 905]	59.1	**75.0**	74.2
	[760, 800, 905]	69.0	82.1	**84.7**

TABLE III
AVERAGE EVALUATION METRIC VALUES OBTAINED BY DIFFERENT OPTIMIZATION STRATEGIES USING DIFFERENT ARTIFACT SOURCES.

Metric	Noise	None	Pre-LS	Joint-LS	Post-LS
RMSE (bpm)	M	11.8	9.0	**5.9**	6.9
	C	11.8	8.5	**5.5**	7.5
SR-AUC	M	0.65	0.67	**0.77**	0.75
	C	0.65	0.70	**0.79**	0.75
Coverage (%)	M	69.0	69.9	**82.1**	79.6
	C	69.0	74.4	**84.7**	80.8

C. Evaluation metrics

We use three metrics to evaluate the performance of pulse rate extraction. All metrics are used for evaluating each remote-PPG method per video.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2938564, IEEE Transactions on Biomedical Engineering

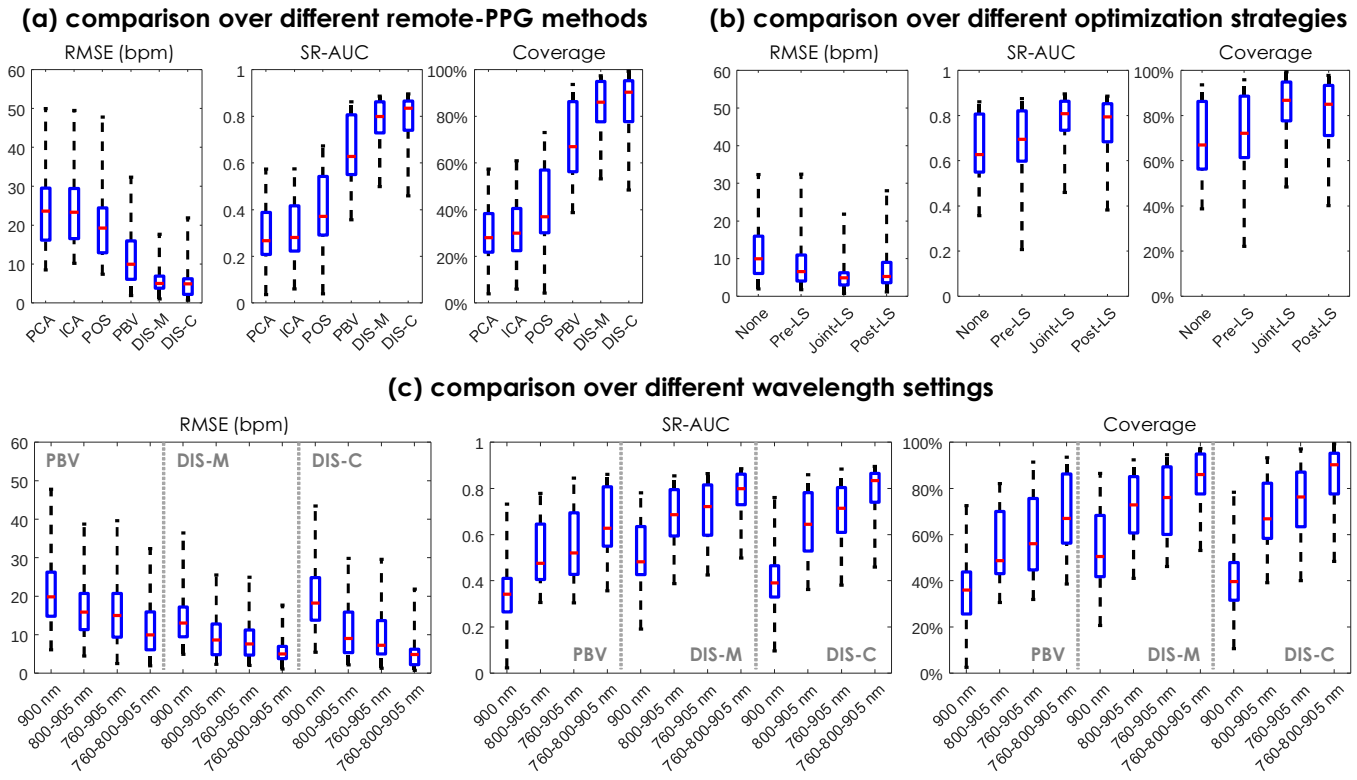

Fig. 6. Statistical comparison of the metric values obtained by benchmarked remote-PPG methods (or their different settings) in three protocols. The median values are indicated by horizontal bars inside the boxes, the quartile range by boxes, the full range by whiskers.

• **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 extraction.

• **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 suggests more accurate extraction. Note that the AUC is normalized by 10 (the total area) and thus varies in [0,1].

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

• **ANOVA** We apply the Analysis of Variance (ANOVA) to the aforementioned metric outputs to analyze the significance of the difference between the compared methods. In ANOVA, the p-value is used as the indicator and a common threshold 0.05 is specified to determine whether the difference is significant, i.e. if $p < 0.05$, the difference is considered to be significant.

V. Results and Discussion

This section presents the experimental results of benchmarked remote-PPG algorithms and different options for DIS (DIS-family: DIS-M and DIS-C) in NIR with heavy motions. Their performance is compared and discussed quantitatively and qualitatively in separate subsections.

A. Quantitative analysis

Table I shows average metric values obtained by 6 benchmarked remote-PPG algorithms, from which we can see that the DIS-family clearly outperforms other algorithms. The most direct comparison is between PBV and DIS (M and C). The motion robustness of PBV is considerably improved by simply extending the blood volume pulse signature of PBV to a discriminative signature using disturbance channels, i.e. the RMSE is lower while success rate and measurement coverage are higher. Comparing DIS-M and DIS-C, we find that DIS-C, using spatial color standard deviation as the artifact source, is slightly better. The reason could be that the spatial-color standard deviation has higher sensitivity in capturing the disturbance-induced local color changes. Also, in the extreme case without any motion (i.e. subject is perfectly stationary), the facial landmark based motion signals (M) can be zero, but not for the spatial color standard deviation based signals (C) given the camera sensor noise.

Table II shows average metric values obtained by PBV and DIS-family on setups with different wavelength combinations. The DIS-family is consistently better than PBV in all wavelength settings. For the single-wavelength PPG extraction (at

TABLE IV
ANOVA TEST (p-VALUE) OF SELECTED COMPARISONS. $p < 0.05$ SUGGESTS SIGNIFICANT DIFFERENCE BETWEEN COMPARED METHODS.

Statistics origin	Camera setup	Compared methods	ANOVA (p-value)		
			RMSE	SR-AUC	Coverage
Table III	1-wavelength	PBV, DIS-M, DIS-C	0.0277	0.0032	0.0050
Table III	2-wavelength	PBV, DIS-M, DIS-C	2.3×10^{-6}	2.5×10^{-8}	1.4×10^{-7}
Table III	3-wavelength	PBV, DIS-M, DIS-C	0.0002	0.0002	0.0005
Table II	3-wavelength	Pre-LS, Joint-LS, Post–LS	0.0227	0.0014	0.0011

905 nm), PBV is not really meaningful as its blood volume pulse signature is 1, whereas DIS-M can still act as a process of least-squares motion reduction by using the discriminative signature [1, 0, 0], i.e. no wavelength combination but still motion reduction. This implies that DIS could be very useful for the applications of wrist-band/fitness-tracker that has the single G-wavelength sensor or cheap single-wavelength NIR solution with a monochrome camera. We also see that DIS-M is better than DIS-C for the single-wavelength use case. This is due to the fact that DIS-M uses two motion channels (with signature [1, 0, 0]), which can better suppress motion artifacts than DIS-C that uses only one spatial color standard deviation channel (with signature [1, 0]). For the 2-wavelengths setup, all compared algorithms (including PBV and DIS-family) show better results on the combination of [760, 905] nm than the combination of [800, 905] nm. This is due to the larger relative (AC/DC) pulsatility contrast between 760 nm and 905 nm than that between 800 nm and 905 nm, which results in a blood volume pulse signature that can better be differentiated from the intensity variation direction [1, 1]. Comparing different wavelength combinations, it is clear that increasing the number of wavelength consistently improves the performance figures, i.e. 3-wavelengths is so far the best choice in available options.

Table III shows average metric values obtained by three optimization approaches using two different disturbance sources. Baseline is the PBV without optimization (denoted as None). Pre-LS and Post-LS are two-steps optimization including PBV-based pulse extraction and least-squares noise reduction, i.e. they differ in the order of these two steps. The results in this table suggest that one-step optimization of DIS (Joint-LS) is better than two-steps optimization of Pre-LS and Post-LS, which is in line with our expectation. But Pre-LS and Post-LS still improve the baseline, which means that adding an additional noise reduction step is somewhat helpful, though it is less effective than joint optimization. We also notice that Post-LS is better than Pre-LS. The reason could be that the blood volume pulse signature has been modified after independent pre-processing per wavelength channel, making the follow-up PBV-based extraction less accurate. According to our earlier analysis, the blood volume pulse signature direction remains the same after pre-processing only if the mixture of artifacts in all wavelength channels are equal, which is extremely unlikely. Although we can use the same least-squares noise reduction vector to pre-process all color channels, the artifact in each channel cannot be fully removed if their mixture is different. Post-LS does not have this problem, but it cannot influence the PPG-extraction. If PPG extraction already fails (i.e. the estimated PPG rate is off), post-processing cannot rescue it.

Based on Fig. 6, we probe the groups of comparisons where the difference between them does not seem immediately significant (i.e. statistical distribution is not clearly separated). Then we use ANOVA to clarify the significance of difference. Take Fig. 6 (a) for example, PCA, ICA and POS are obviously worse than PBV, DIS-M and DIS-C, i.e. the median of the first three methods are out of the quartile-range of the last three methods. Thus our ANOVA-test is focused on the comparison between PBV, DIS-M and DIS-C.

Table IV shows the ANOVA results (p-values) of targeted comparisons based on Fig. 6. We conclude that (i) DIS-family is significantly better than PBV for all wavelength setups in all metric evaluations; (ii) the most significant improvement achieved by DIS is in the 2-wavelengths setup. The reason could be that all methods are relatively poor for the 1-wavelength setup, as the pulsatile signature is not functional for a single-wavelength channel. For the 3-wavelengths setup, PBV can profit from the use of pulsatile signature for noise reduction, which shrinks its difference w.r.t. DIS. The 2-wavelengths setup is somewhat in between, where the pulsatile signature is not fully fledged but starts being functional; (iii) the joint-optimization of DIS (Joint-LS) is significantly better than the sub-steps optimization like combining pulse extraction with pre-/post-processing (Pre-LS and Post-LS).

B. Qualitative analysis

Figs. 7-8 show individual pulse-rate traces obtained by 6 benchmarked remote-PPG algorithms over the complete dataset. Since the NIR videos are recorded with heavy motions, there are many occasions that the facial landmark tracker fails, i.e. it cannot find a face in an extreme rotation angle (e.g. almost 90 degrees face rotation) or cannot find correct landmark matches in specific views (see the snapshots in the first column of Figs. 7-8). For fair comparison, we use the tracker quality (see the second column of Figs. 7-8) to annotate the tracker failures (in red) and exclude the pulse-rate intervals corresponding to those moments from benchmark, as they do not allow PPG extraction. The comparison is performed on the remaining pulse-rate intervals (in blue) where the tracker works properly. For the pruned measurements, the DIS-family is clearly better than the other compared methods, showing much more consistent alignment with the reference pulse rate.

More specifically, we consider the measurement on a subject with a coverage (allowing 3 bpm off) higher than 70% to

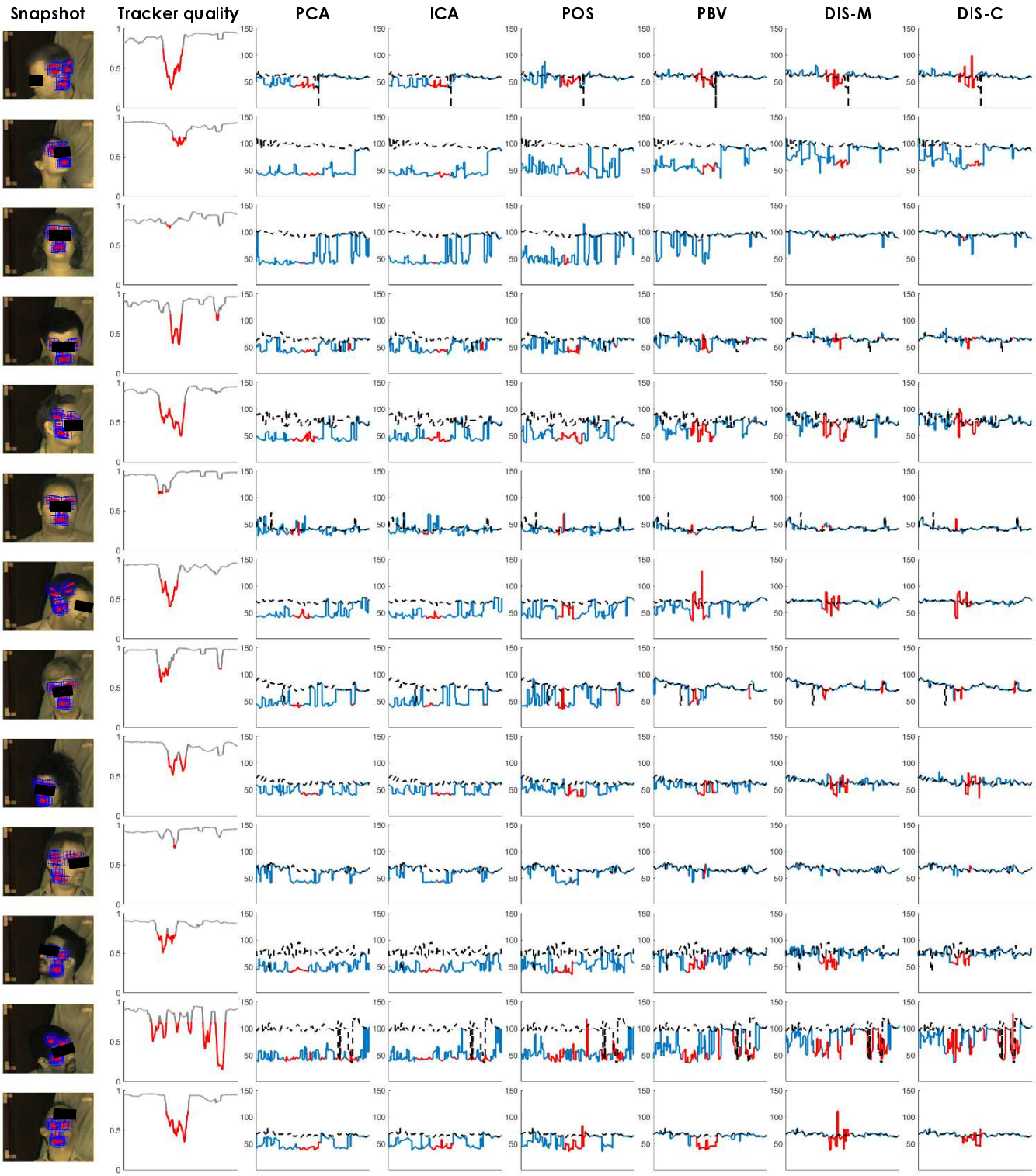

Fig. 7. The pulse-rate traces (measured in bpm for around 8-10 minutes long sequences) obtained by PCA, ICA, POS, PBV, DIS-M and DIS-C on the NIR recordings with heavy motions. The first column shows a snapshot of the recording and used facial landmark patches. The second column shows the tracker quality (from 0 to 1) over time, i.e. values toward 0 indicate tracker failure (annotated in red). The last six columns show the pulse-rate traces where the dashed-black and solid-blue traces denote the pulse rates measured by PPG reference and camera, respectively.

be successful. For BSS-based methods (PCA and ICA), no subject can achieve this, expressed as 0/26 subjects. POS is slightly better than BSS-based methods (2/26 subjects) and PBV improves this further (12/26 subjects). DIS-M and DIS-C show the largest measurement coverage among the test subjects, giving 21/26 subjects and 22/26 subjects, respectively.

To verify whether the performance of the proposed algorithm depends on the pulse rate (from 40 to 110 bpm of the test subjects), we show the scatter plots of the median pulse rates obtained by the contact-PPG and camera-PPG over the

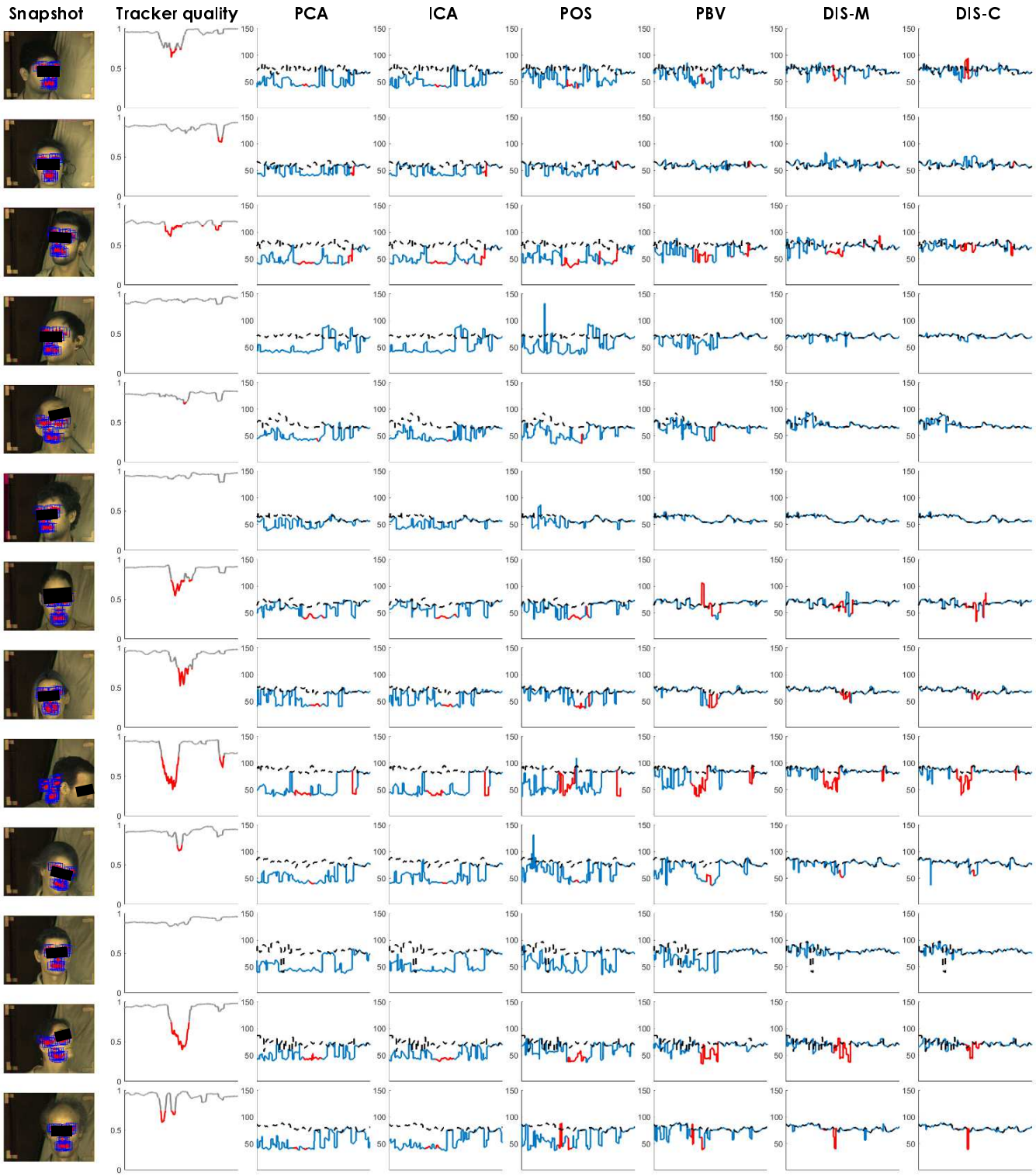

Fig. 8. The pulse-rate traces (measured in bpm for around 8-10 minutes long sequences) obtained by PCA, ICA, POS, PBV, DIS-M and DIS-C on the NIR recordings with heavy motions. The plotting conventions are the same as Fig. 7.

test subjects in Fig. 9. It is clear that PBV and DIS (M and C) have better correlation with the reference than PCA, ICA and POS. PCA, ICA and POS seem to underestimate the true pulse rate, the erroneous measurement of which are mainly due to the disturbances introduced by head motions at the lower frequency range as compared to the pulse rates (see Figs. 7-8). Furthermore, we can see that DIS (M and C) has better correlation with the contact-PPG than PBV.

In view of the results presented in this paper, we intend to further validate DIS in other challenging environments requiring NIR sensing 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 DIS in clinical trials with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2938564, IEEE Transactions on Biomedical Engineering

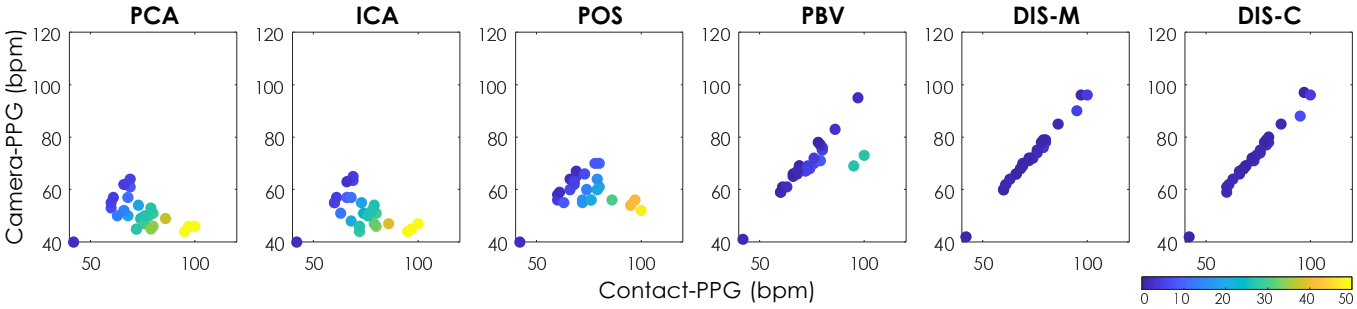

Fig. 9. The scatter plots of subjects' pulse rate measured by the contact-PPG and camera-PPG (i.e. using 6 different PPG-extraction algorithms), which indicates the correlation between two different measurement. Each scatter point represents a subject's median pulse rate measured throughout the recording.

real patient data such as emergency department triage and general ward. In addition, we shall investigate DIS in RGB applications in the future, as we expect it to bring improvement in visible light as well, i.e. the principle of using disturbance signals (e.g. motion) to improve PPG extraction is independent of the choice of camera wavelengths. We will also explore different options of creating disturbance signals for DIS and their hybrids (i.e. fuse multiple disturbance sources) to achieve further optimizations.

VI. Conclusion

In this paper, we introduce a novel discriminative signature based approach (DIS) for camera-PPG extraction, which significantly improves the robustness of pulse-rate measurement in NIR with heavy motions. The newly proposed DIS method takes both the color signals (blood absorption variation related) and disturbance signals (motion or illumination variation related) as input for PPG extraction. By defining a discriminative signature (extending the blood volume pulse signature with zero entries corresponding to the disturbance channels), we use one-step least-squares regression to conduct pulse extraction (from color signals) and noise reduction (in disturbance signals) simultaneously. A large-scale lab recordings in NIR with heavy body motions demonstrates the robustness of our prototype DIS and its significant improvement over prior art. Provided the face that can be detected and tracked correctly, our proposed PPG-extraction algorithm shows an improvement.

Acknowledgment

The authors would like to thank Aline Serteyn at Philips Research for creating the benchmark dataset in near infrared and Ralph van Dinther at Philips Research for discussions on the topic.

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] 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.
[3] J.-C. Cobos-Torres *et al.*, "Non-contact, simple neonatal monitoring by photoplethysmography," *Sensors*, vol. 18, no. 12, p. 4362, 2018.
[4] 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.
[5] K. De Miguel *et al.*, "Home camera-based fall detection system for the elderly," *Sensors*, vol. 17, no. 12, p. 2864, 2017.
[6] W. Wang *et al.*, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, no. 6, p. 1023, 2017.
[7] S. Leonhardt *et al.*, "Unobtrusive vital sign monitoring in automotive environments - a review," *Sensors*, vol. 18, no. 9, p. 3080, 2018.
[8] W. Wang *et al.*, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, July 2017.
[9] 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.
[10] 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.
[11] M. van Gastel *et al.*, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.
[12] W. Wang *et al.*, "Color-distortion filtering for remote photoplethysmography," in *12th IEEE Conf. Automatic Face Gesture Recognit.*, May 2017, pp. 71–78.
[13] ——, "Amplitude-selective filtering for remote-PPG," *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1965–1980, Mar. 2017.
[14] G. Cennini, J. Arguel, K. Akşit, and A. van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Express*, vol. 18, no. 5, pp. 4867–4875, Mar. 2010.
[15] C. van Dinther *et al.*, "Motion artifact reduction using multi-channel PPG signals," 2017, US Patent Application 2017/0071546 A1.
[16] W. Wang *et al.*, "Single element remote-PPG," *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2018.
[17] 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.
[18] G. Balakrishnan *et al.*, "Detecting pulse from head motions in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, Oregon, USA, June 2013, pp. 3430–3437.
[19] J. Friedman *et al.*, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.
[20] W. Wang *et al.*, "Full video pulse extraction," *Biomed. Opt. Express*, vol. 9, no. 8, pp. 3898–3914, Aug. 2018.
[21] G. Zonios *et al.*, "Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection," *J. Biomed. Opt.*, vol. 13, no. 1, p. 014017, 2008.
[22] M. Lewandowska *et al.*, "Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inform. Syst. (FedCSIS)*, Szczecin, Poland, Sept. 2011, pp. 405–410.
[23] M.-Z. Poh *et al.*, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
[24] 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.