

Amplitude-selective filtering for remote-PPG

WENJIN WANG,^{1,*} ALBERTUS C. DEN BRINKER,² SANDER STUIJK,¹
AND GERARD DE HAAN^{1,2}

¹*Electronic Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

²*Philips Innovation Group, Philips Research, Eindhoven, The Netherlands*

*w.wang@tue.nl

Abstract: Biometric signatures of remote photoplethysmography (rPPG), including the pulse-induced characteristic color absorptions and pulse frequency range, have been used to design robust algorithms for extracting the pulse-signal from a video. In this paper, we look into a new biometric signature, i.e., the relative pulsatile amplitude, and use it to design a very effective yet computationally low-cost filtering method for rPPG, namely “amplitude-selective filtering” (ASF). Based on the observation that the human relative pulsatile amplitude varies in a specific lower range as a function of RGB channels, our basic idea is using the spectral amplitude of, e.g., the R-channel, to select the RGB frequency components inside the assumed pulsatile amplitude-range for pulse extraction. Similar to band-pass filtering (BPF), the proposed ASF can be applied to a broad range of rPPG algorithms to pre-process the RGB-signals before extracting the pulse. The benchmark in challenging fitness use-cases shows that applying ASF (ASF+BPF) as a pre-processing step brings significant and consistent improvements to all multi-channel pulse extraction methods. It improves different (multi-wavelength) rPPG algorithms to the extent where quality differences between the individual approaches almost disappear. The novelty of the proposed method is its simplicity and effectiveness in providing a solution for the extremely challenging application of rPPG to a fitness setting. The proposed method is easy to understand, simple to implement, and low-cost in running. It is the first time that the physiological property of pulsatile amplitude is used as a biometric signature for generic signal filtering.

© 2017 Optical Society of America

OCIS codes: (280.0280) Remote sensing and sensors; (170.3880) Medical and biological imaging.

References and links

1. W. Verkruyse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” *Opt. Express* **16**(26), 21434–21445 (2008).
2. M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, “Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity,” in *Proceedings of Federated Conference on Computer Science and Information Systems (IEEE, 2011)*, pp. 405–410.
3. M. Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Trans. Biomed. Eng.* **58**(1), 7–11 (2011).
4. W. Wang, S. Stuijk, and G. de Haan, “A novel algorithm for remote photoplethysmography: Spatial subspace rotation,” *IEEE Trans. Biomed. Eng.* **63**(9), 1974–1984 (2016).
5. G. de Haan and V. Jeanne, “Robust pulse rate from chrominance-based rPPG,” *IEEE Trans. Biomed. Eng.* **60**(10), 2878–2886 (2013).
6. G. de Haan and A. van Leest, “Improved motion robustness of remote-PPG by using the blood volume pulse signature,” *Physiol. Meas.* **35**(9), 1913–1922 (2014).
7. W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote-PPG,” *IEEE Trans. Biomed. Eng.* **PP**(99), 1 (2016). (posted 13 September 2016, in press).
8. W. Verkruyse, M. Bartula, E. Bresch, M. Rocque, M. Meftah, and I. Kirenko, “Calibration of contactless pulse oximetry,” *Anesth. Analg.* **124**(1), 136–145 (2017).
9. W. Wang, B. Balmaekers, and G. de Haan, “Quality metric for camera-based pulse rate monitoring in fitness exercise,” in *Proceedings of IEEE International Conference on Image Processing (IEEE, 2016)*, pp. 2430–2434.
10. C. Julien, “The enigma of Mayer waves: Facts and models,” *Cardiovasc. Res.* **70**(1), 12–21 (2006).
11. A. A. Kamshilin, I. S. Sidorov, L. Babayan, M. A. Volynsky, R. Giniatullin, and O. V. Mamontov, “Accurate measurement of the pulse wave delay with imaging photoplethysmography,” *Biomed. Opt. Express* **7**(12), 5138–5147 (2016).
12. F. Bousefsaf, C. Maaoui, and A. Pruski, “Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate,” *Biomed. Sig. Proc. Control* **8**(6), 568–574 (2013).

13. X. Li, J. Chen, G. Zhao, and M. Pietikäinen, "Remote heart rate measurement from face videos under realistic situations," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2014), pp. 4264–4271.
14. M. Hülsbusch, *An Image-based Functional Method for Opto-electronic Detection of Skin Perfusion* (Ph.D. dissertation, 2008).
15. G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras," *J. Biomed. Opt.* **20**(4), 048002 (2015).

1. Introduction

Remote photoplethysmography (rPPG) enables contactless monitoring of human cardiac activities by measuring the pulse-induced subtle color variations on the human skin surface through a regular RGB camera [1]. This measurement is based on the fact that the pulsatile blood propagating in the human cardiovascular system changes the blood volume in skin tissue. The oxygenated blood circulation leads to fluctuations in the amount of hemoglobin molecules and proteins thereby causing a fluctuation in the optical absorption across the light spectrum. An RGB camera can be used to identify the phase of the blood circulation based on minute color changes in skin reflections.

Recently, several multi-channel pulse-extraction methods have been proposed. These include: (i) BSS-based approaches (PCA [2] and ICA [3]), which use different criteria to de-mix temporal RGB traces into uncorrelated or independent signal sources to retrieve the pulse; (ii) a data-driven approach (2SR [4]), which measures the temporal hue change from the spatial subspace rotation of skin-pixels as the pulse; and (iii) model-based approaches (CHROM [5], PBV [6] and POS [7]), which exploit characteristic properties of skin reflections (e.g., typical color absorption variations due to blood volume changes in living skin-tissues) and different assumptions on the distortions in the color channels to design a projection function from which the pulse-signal is extracted. A thorough review of these algorithms can be found in [7]. Among them, the model-based approaches demonstrate superior robustness in dealing with practical challenges such as the skin-tone variations, body-motions and illumination conditions. This is due to the deployment of the physiological and optical properties of skin reflections as (rPPG-related) priors to facilitate the pulse extraction. In contrast, BSS-based approaches that do not use such priors need more data to get the high-quality statistics for solving the source de-mixing problem. In general, there are two major biometric signatures being used by current rPPG algorithms: (i) *characteristic color absorptions* (i.e., pulse-induced color variation directions in a multi-spectrum camera), which has been exploited by model-based rPPG [5–7] to differentiate between the pulse-induced color changes and noise-induced color changes; (ii) *characteristic pulse-frequency range*, which has been used in (almost) all rPPG algorithms to eliminate the frequency noise outside the assumed frequency-band (e.g., [40, 240] beat per minute (bpm)), using a Band-Pass Filter (BPF).

Based on the recent findings reported by [8], we recognize that the relative amplitude (AC/DC) of the human pulsatile component varies in a specific lower range as a function of RGB channels, i.e., [0.0001, 0.002] AC/DC for the R-channel based on our experiments. The AC/DC is defined as the variation amplitude of the target signal after the DC normalization. Take a color-signal measured by the camera as an example, its AC/DC can be measured by first dividing the color-signal by its temporal mean (DC) and then measuring the frequency amplitude (AC) of the target component (i.e., pulse) in the DC-normalized signal. In this paper it is investigated if rPPG can be improved by incorporating this prior knowledge. Our basic idea is using the spectral amplitude of, e.g., the R-channel, to select the frequency components in RGB channels inside the assumed "*characteristic pulsatile amplitude-range*" for pulse extraction, while the frequency components outside this range are pruned as noise. We shall refer to this novel approach as "Amplitude Selective Filtering" (ASF), and illustrate its principles in Fig. 1. Similar to BPF, ASF can also be applied to a broad range of rPPG algorithms as a pre-processing or post-processing step to improve the pulse extraction. Here we particularly use it for *pre-processing*, as it can repair/correct

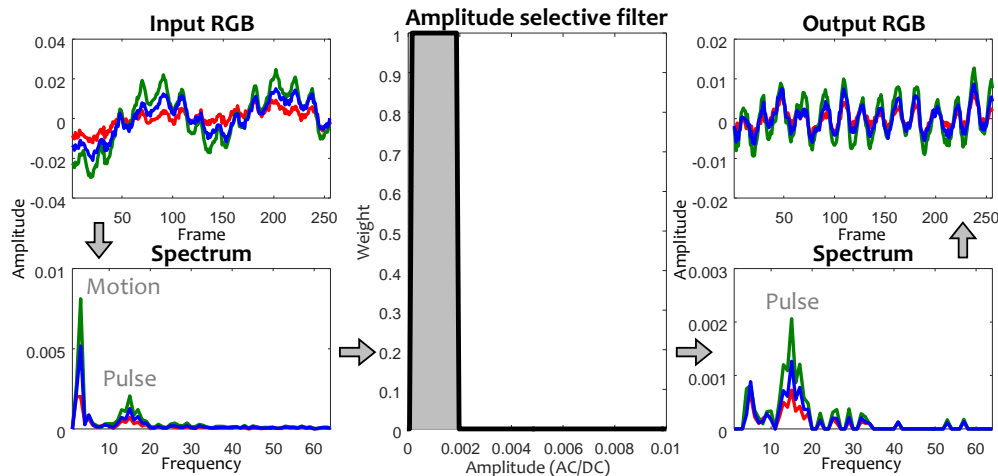


Fig. 1. Illustration of the amplitude-selective filtering. The essence of this filter is in the step of selecting the RGB frequency components within the pulsatile amplitude-range for pulse extraction, thus removing large components due to motion.

the color distortions in RGB-signals at an earlier stage, such that the cleaned AC-components will benefit the pulse extraction in rPPG algorithms. Here we clarify the difference between *pre-processing* and *post-processing* in rPPG: (i) pre-processing cleans the input (raw RGB-signals) and thus can influence the pulse extraction; (ii) post-processing cleans the output (extracted pulse-signal), which cannot influence the pulse extraction. A benchmark on challenging fitness recordings shows that applying ASF (ASF+BPF) as a pre-processing step brings significant and consistent improvements to the existing multi-channel pulse extraction methods. The strength of ASF is evident not only from the fact that it improves all benchmarked (multi-wavelength) rPPG algorithms but in particular because it drives them to a similar quality-level.

The novelty of this work is that we introduce a simple yet powerful pre-filtering method (i.e., ASF) that significantly improves the performance of rPPG methods, particularly in the challenging use-case of fitness. It is the first time that the physiological property of pulsatile amplitude is exploited as a biometric signature for generic signal filtering. The proposed method is easy to understand, simple to implement, and low-cost in running, i.e., the challenging motion problem in fitness can be addressed by a few lines of Matlab code. Most importantly, the improvement introduced by ASF is general to all multi-wavelength rPPG methods, as demonstrated by seven existing core rPPG algorithms in the benchmark. As such, it can not only be used as an add-on function by the existing rPPG methods to increase their robustness, but also by rPPG methods that will be developed in the future. Similar to the widely used BPF, ASF can be standardized as a generic filtering step in vital signs monitoring systems/frameworks to benefit the community at large. Though the concept of ASF and its implementation is simple, it has so far not been considered nor evaluated as a means to shape the frequency response of the commonly used Band-Pass Filter depending on the energy/amplitude per frequency bin. We will show that our ASF provides an effective solution for the previously considered very hard problem of measuring heart-rate during vigorous exercise in a fitness setting. We use a very simple step to solve a challenging problem that existing core rPPG algorithm cannot deal with.

The remainder of this paper is structured as follows. In Section II, we define the problem. In Section III, we analyze the considered problem in detail and describe the amplitude-selective filtering method. In Section IV and V, we benchmark the proposed filter and discuss its performance. Finally in Section VI, we draw our conclusions.

2. Problem definition

Unless stated otherwise, vectors and matrices are denoted as boldface characters throughout this paper. Considering the RGB-signals measured by a camera within a time interval as \mathbf{C} (i.e., RGB channels are ordered in rows), it is composed of different source-signals observed from the environment (e.g., pulse and motion). Each channel of \mathbf{C} can be physically expressed as:

$$\mathbf{C}_i = \mathbf{S}_{i,1} + \mathbf{S}_{i,2} + \dots + \mathbf{S}_{i,M} = \sum_{n=1}^M \mathbf{S}_{i,n}, \quad (1)$$

where \mathbf{C}_i denotes the i -th channel of \mathbf{C} ; $\mathbf{S}_{i,n}$ denotes the n -th source-signal contributing to \mathbf{C}_i ; M is the total number of sources. The component of interest, i.e., pulse, is one of the sources among $\mathbf{S}_{i,n}$, which we assume to be $\mathbf{S}_{i,1}$. The goal of filtering is to derive a $\hat{\mathbf{C}}_i$ that approximates $\mathbf{S}_{i,1}$. Thus our first step is to separate \mathbf{C}_i into different components, where pulse is assigned to, preferably, a single component. Since pulse is a periodic signal, we can capture this property by decomposing \mathbf{C}_i into different temporal frequency components:

$$\mathbf{C}_i = \mathbf{F}_{i,1} + \mathbf{F}_{i,2} + \dots + \mathbf{F}_{i,N} = \sum_{n=1}^N \mathbf{F}_{i,n}, \quad (2)$$

where $\mathbf{F}_{i,n}$ denotes the n -th frequency component of \mathbf{C}_i , which can be obtained by the frequency decomposition; N is the total number of frequency components, with, in practice, $N > M$. The targeted source-signal, $\mathbf{S}_{i,1}$, is expressed as a single component or very limited number of components in $\mathbf{F}_{i,n}$. The main task of filtering is to select the pulse-related components from $\mathbf{F}_{i,n}$.

Intuitively, some $\mathbf{F}_{i,n}$ in (2) cannot be related to the pulse-signal, as the (periodic) pulsatile component does not spread in the entire frequency spectrum. Hence, a band-pass filter is typically used to select the pulse-related $\mathbf{F}_{i,n}$ with specifications based on the assumption that the human pulse-rate can only vary in a range, e.g., [40, 240] bpm. This procedure can be expressed as weighting \mathbf{F}_i using a binary mask:

$$\hat{\mathbf{C}}_i = \sum_{n=1}^N w_{i,n} \cdot \mathbf{F}_{i,n} \text{ with } w_{i,n} = \begin{cases} 1, & \text{if } n \in [b_{min}, b_{max}], \\ 0, & \text{elsewhere,} \end{cases} \quad (3)$$

where $w_{i,n}$ denotes the combining weight for $\mathbf{F}_{i,n}$; $[b_{min}, b_{max}]$ denotes the assumed pulse frequency-band. In essence, BPF uses the *frequency-index* (n) of $\mathbf{F}_{i,n}$ to determine their combining weights. As a consequence, it cannot deal with the case when the noise-frequency enters the assumed frequency-band, which typically occurs in fitness applications where the periodic body motions usually occur in the pulse frequency-band [9].

Looking at this problem from a different angle, we recognize that it is also possible to use the *amplitude* of $\mathbf{F}_{i,n}$ to determine their combining weights. The rationale is: the human relative pulsatile amplitude (AC/DC) also varies in a specific range as a function of RGB channels [8]. Therefore, we can define a (narrow) pulsatile amplitude-band to only select the $\mathbf{F}_{i,n}$ with the in-band amplitude for pulse extraction. This is equivalent to defining an amplitude-selective filter. We hypothesize that such an approach could be highly attractive for the fitness use-case, as the periodic motion distortions are typically stronger than the pulse-signal itself, certainly in the R-channel that contains rather low pulse-energy due to the low blood absorption in red. Moreover, we expect its advantage to increase for more vigorous exercise as compared to the band-pass filter, since the significant body motions may enter the pulse frequency-band, but are increasingly less likely to enter the low pulsatile amplitude-band.

Thus the problem at hand is the design of an amplitude-selective filter to improve the robustness of existing rPPG algorithms. In the next section, we shall analyze this problem in detail and present our solution.

3. Method

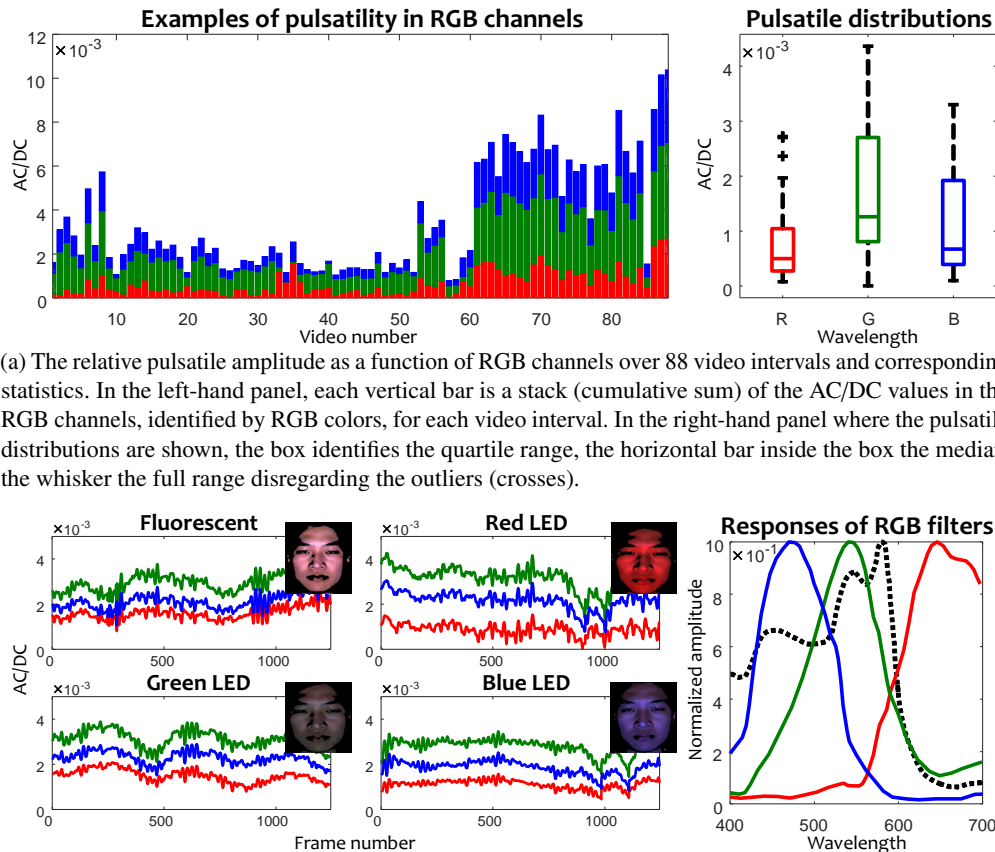
In this section, we first analyze the criteria for defining an amplitude-selective filter, and then describe the proposed algorithm.

3.1. Analysis

The study of [8] shows that the relative pulsatile amplitude, measured in red (675 nm) and infrared (842 nm) wavelengths, varies around $(1 \pm 0.5) \times 10^{-3}$ AC/DC dependent on different skin temperatures (from 7°C to 23°C). It is in fact a quite narrow range as compared to the motion amplitudes (particularly in fitness), which triggers us to restrict the frequency components admitted to the pulse-extraction processing to the ones falling inside the characteristic pulsatile amplitude-range. This can be translated into the procedure of using the spectral amplitude (AC-level) of $\mathbf{F}_{i,n}$ to determine the $w_{i,n}$ in (3), such that the $\mathbf{F}_{i,n}$ outside the assumed amplitude-range can be suppressed. However, it remains questionable whether the quantitative data on the relative pulsatile amplitude-range provided by [8] can be directly used in our approach, since [8] used monochrome cameras with selected narrow-band block filters to measure the pulsatility, which are different from the optical filters in a regular RGB camera. In order to find the proper quantitative data for such a setup, we re-do this experiment by measuring the relative pulsatile amplitudes (AC/DC) in RGB channels using a regular multi-wavelength RGB camera (Global shutter RGB CCD camera USB UI-2230SE-C from IDS, with 768×576 pixels, 8 bit depth, and 20 frames per second (fps)).

Targeting fitness applications, our experiment considers adult subjects in a moderate room temperature (i.e., 20°C - 25°C). We used 22 video sequences recorded on 15 stationary subjects with different skin-tones and under different illumination conditions. The recruited 15 subjects have three skin-types based on the Fitzpatrick scale: 5 Western European subjects (skin-type III), 5 Eastern Asian subjects (skin-type III), and 5 Sub-Sahara Africa/Southern Asian subjects (skin-type IV-V). Four lighting conditions are used: Fluorescent, Red LED, Green LED and Blue LED. The reference PPG-signal is recorded by a finger-contact transmissive pulse oximetry (Model CMS50E from ContecMedical) and synchronized with the video frames. Since the pulsatility of a stationary subject is also time-varying due to the Mayer-waves (i.e., oscillations of arterial pressure occurring spontaneously in conscious subjects) [10], we can split one video recording into multiple non-overlap/independent intervals (e.g., 4 intervals) to extend the measurement. Eventually, we have 88 video intervals in total. By using the contact-PPG reference signal recorded in parallel, we can precisely locate the pulse frequency components in RGB channels and obtain their relative amplitudes (AC/DC). Fig. 2(a) shows the relative pulsatile amplitude as a function of RGB channels over 88 video intervals and their corresponding statistics. We observe that (i) the G-channel has the highest pulsatility, followed by the B and R channels; (ii) the R-channel has the lowest pulsatile variation, followed by the B and G channels, i.e., the variation of AC/DC is proportional to its average level; and (iii) the difference between RGB channels is in fact not large. The reason for (iii) can be explained by the overlap of the transmission spectra of the optical filters in the Bayer pattern of the used RGB camera, where the R-channel also sees the information in the G-channel (see Fig. 2(b)). The wavelength-overlap could be an advantage for extreme lighting spectra (e.g., non-white illumination), where all three channels can more or less sense the pulsatility across channels, i.e., if the R-channel is completely dark that cannot sense any information (e.g., in blue-LED illumination), the functions of both the filtering and pulse extraction will break down.

To design an effective amplitude-selective filter, we have two hypotheses: (i) the pulsatile components have small relative amplitudes (AC/DC) that are bounded in a specific lower range; (ii) the (fitness) motion distortions have larger AC/DC that allow them to be distinguished from the pulsatile amplitude. To this end, we choose to use the R-channel to select the pulsatile components in RGB channels. The reasons are following: (i) the R-channel has the lowest average



(a) The relative pulsatile amplitude as a function of RGB channels over 88 video intervals and corresponding statistics. In the left-hand panel, each vertical bar is a stack (cumulative sum) of the AC/DC values in the RGB channels, identified by RGB colors, for each video interval. In the right-hand panel where the pulsatile distributions are shown, the box identifies the quartile range, the horizontal bar inside the box the median, the whisker the full range disregarding the outliers (crosses).

(b) The continuous relative pulsatile amplitude traces obtained from RGB channels in videos recorded under different (extreme) lighting conditions (e.g., colored LED illumination). The right-hand panel shows the responses of RGB filters of the used camera (red/green/blue curves) and the absolute PPG-amplitude (black dashed curve).

Fig. 2. (a) The statistically measured relative pulsatile amplitude as a function of RGB channels. (b) The continuous relative pulsatile traces exemplified in extreme lighting conditions and the optical filter responses of the used RGB camera.

AC/DC, which is easier to be differentiated from that of the large motions; (ii) the AC/DC of the R-channel has the smallest range (e.g., [0.0001, 0.002] AC/DC), the mean/median of which is more bounded and less uncertain than that of the G and B channels. Essentially, we only need to define the maximum amplitude bound for the R-channel, while the minimum amplitude bound is non-critical, because the influence of small noise variations (with an amplitude even smaller than the pulse) on the estimated pulse-signal is negligible. Based on our experiments, we define the maximum amplitude bound for the R-channel as 0.002, i.e., the component with an amplitude larger than this threshold will be suppressed. Such a filter is expected to be particularly effective for eliminating large motion distortions in challenging use-cases like fitness, and less effective for simple use-cases where distortions have small AC/DC that is similar to pulse. Although less effective, the proposed filter can never harm the pulse extraction in the rPPG module, as it does not introduce new distortions/artifacts.

3.2. Algorithm

Given the raw RGB-signals \mathbf{C} , our first step is to eliminate the dependency of \mathbf{C} on the average skin reflection color (DC-level). This can be done by the temporal normalization (AC/DC-1) [5]:

$$\tilde{\mathbf{C}}_i = \frac{\mathbf{C}_i}{\mu(\mathbf{C}_i)} - 1, \quad (4)$$

where $\tilde{\mathbf{C}}_i$ denotes the zero-mean color variation signal in the i -th channel, the DC of which is normalized and removed; $\mu(\cdot)$ denotes the averaging operator. Note that an alternative to (4) is to take the logarithm of \mathbf{C}_i and remove the mean, which for small variations as the PPG-signal has practically the same effect [7].

Here we mention that \mathbf{C}_i is a 1D color-signal averaged of the skin-pixel values from a whole face, where the spatial distribution of the pulsatile signals is eliminated. We understand that in the domain of imaging-PPG (iPPG) [11], the PPG-waveforms from different locations of a human body are different, but using the spatially averaged pixel values for pulse extraction is still a valid option in our rPPG task. The reasons are threefold: (i) we consistently use the complete face area as a single spot for pulse measurement. Thus the spatially averaged color-signal assembles a stable and consistent PPG-waveform averaged from the whole face in time. This is similar to the iPPG where the measurement also depends on the resolution of the used local pixel-/patch- sensors, i.e., the RoI sizing 7×7 (or 20×20) pixels may already combine different PPG-waveforms. But this does not constitute a problem as long as the RoI is consistently sampled from the same location of the skin. In a similar vein, we use the whole face as a single RoI for measurement, where the only difference is the RoI resolution; (ii) we use a 20 fps camera for video recording. With such a low frame-rate, the pulse transit time or the pulse wave delay between pixels on face can be neglected. Also, the camera in our setup is placed around 2 meters in front of the exercising subject. With the used focal length, the percentage of the face area in a video frame (640×480 pixels) is approximately 15-20%, which is much smaller than that in a typical iPPG imager. With such a resolution, the signal differences between pixels on a face are trivial; and (iii) in fitness applications, our ultimate goal is to estimate a rough heart-rate (HR) trace (i.e., a single parameter signal instead of imaging) to optimize the effectiveness of a workout in real-time. We are not looking for the detailed cardiac features in the PPG-waveform shape to guess the arterial stiffness or cardiovascular age during exercise. IPPG is far more challenging, as it requires accurate pixel-to-pixel registration, which, with the state-of-the-art algorithms, is not yet feasible for the fitness task with significant body motions. Note that using the “spatial pixel averaging” to estimate the physiological signal (e.g., pulse) is a common step in all existing rPPG works, which is not a contribution of this paper

To analyze the AC-components, we transform $\tilde{\mathbf{C}}_i$ into the frequency-domain using the Fast Fourier Transform (FFT):

$$\mathbf{F}_i = \frac{\mathbf{FFT}(\tilde{\mathbf{C}}_i)}{L}, \quad (5)$$

where \mathbf{F}_i denotes the frequency spectrum of the i -th channel; $\mathbf{FFT}(\cdot)$ denotes the FFT operator; L is the signal length. Note that \mathbf{F}_i needs to be scaled by dividing L to eliminate the energy variance due to different signal lengths (e.g., number of frames). Based on earlier reasoning, we choose to use the R-channel to derive the combining weights for selecting the AC-components in RGB channels. The weighting vector \mathbf{W} , consisting of different combining weights \mathbf{W}_n , is derived by:

$$\mathbf{W}_n = \begin{cases} 1, & \text{if } \mathbf{abs}(\mathbf{F}_{1,n}) < a_{max}, \\ \frac{\Delta}{\mathbf{abs}(\mathbf{F}_{1,n})}, & \text{elsewhere,} \end{cases} \quad (6)$$

where \mathbf{W}_n denotes the weight for the corresponding $\mathbf{F}_{1,n}$; $\mathbf{F}_{1,n}$ denotes the n -th component of \mathbf{F}_1 (i.e., the spectrum of the R-channel, the first channel in \mathbf{F}); $\mathbf{abs}(\cdot)$ takes the absolute

value (i.e., spectrum amplitude); a_{max} is the maximum amplitude threshold used for selecting the AC-components, which is set to 0.002 based on our earlier analysis; Δ is a small number that prevents the zero-weight, which is specified as 0.0001 based on the lower bound of the pulsatile range (in the R-channel) found by our experiments, i.e., the unselected components are suppressed to a level lower than the relative pulsatile amplitude.

We note that in (6), the weight assigned to the unselected AC-component is spectrum-dependent small value instead of 0. This is to avoid the situation that less than 3 AC-components are selected for pulse extraction, which renders some rPPG algorithms invalid, i.e., the filtered RGB-signals must contain at least 3 AC-components for BSS-based [2, 3] and PBV [6] algorithms, otherwise it results in a (near-) singular covariance matrix that cannot be solved. Next, we use \mathbf{W} to weight each channel of \mathbf{F} :

$$\hat{\mathbf{F}}_i = \mathbf{W} \odot \mathbf{F}_i, \quad (7)$$

where \odot denotes the element-wise product; $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_L]$. Consequently, the weighted spectrum $\hat{\mathbf{F}}_i$ is transformed back into the time-domain using the Inverse Fast Fourier Transform (IFFT):

$$\hat{\mathbf{C}}_i = \mu(\mathbf{C}_i) \cdot (\mathbf{IFFT}(\hat{\mathbf{F}}_i) + 1), \quad (8)$$

where $\mathbf{IFFT}(\cdot)$ denotes the IFFT operator. Note that DC of the color is re-instated to the filtered signals to keep the original meaning of RGB-channels, as some rPPG algorithms cannot work with DC-free signals, such as the HUE algorithm [15]. $\hat{\mathbf{C}}$ is the final output of the filtering, which can be used as the input of rPPG algorithms for pulse-signal extraction.

The complete algorithm of Amplitude-Selective Filtering (ASF) is shown in Algorithm 1, which is very easy to replicate and allows all kinds of refinements, i.e., the implementation only requires a few lines of Matlab code. The ASF-algorithm is kept as simple and clean as possible to highlight the essence of our idea. Further dedicated algorithmic optimization on ASF, i.e., adapting the maximum amplitude threshold (a_{max}) to the pulsatility of the measured subject in a specific video to further restrict the component selection, is not considered in this paper but shall be left to future work.

Algorithm 1: Amplitude-Selective Filtering

Input: Raw RGB-signals \mathbf{C} with dimension $3 \times L$

Initialization: $a_{max} = 0.002$, $\Delta = 0.0001$ (default)

- 1 $\tilde{\mathbf{C}} = \text{diag}(\text{mean}(\mathbf{C}, 2))^{-1} * \mathbf{C} - 1$;
- 2 $\mathbf{F} = \text{fft}(\tilde{\mathbf{C}}, [], 2)/L$;
- 3 $\mathbf{W} = \Delta ./ \text{abs}(\mathbf{F}(1, :)); \leftarrow \mathbf{F}(1, :)$ is the R-channel
- 4 $\mathbf{W}(\text{abs}(\mathbf{F}(1, :)) < a_{max}) = 1$;
- 5 $\hat{\mathbf{F}} = \mathbf{F} * \text{repmat}(\mathbf{W}, [3, 1])$;
- 6 $\hat{\mathbf{C}} = \text{diag}(\text{mean}(\mathbf{C}, 2)) * (\text{ifft}(\hat{\mathbf{F}}, [], 2) + 1)$;

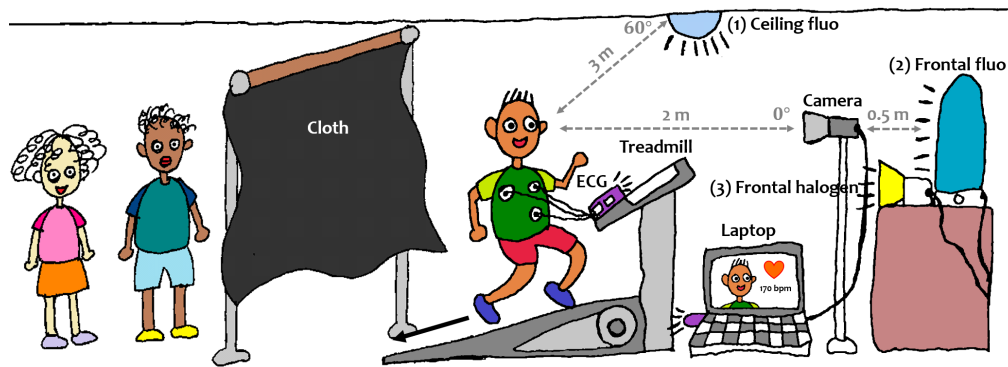
Output: Filtered RGB-signals $\hat{\mathbf{C}}$

4. Experimental setup

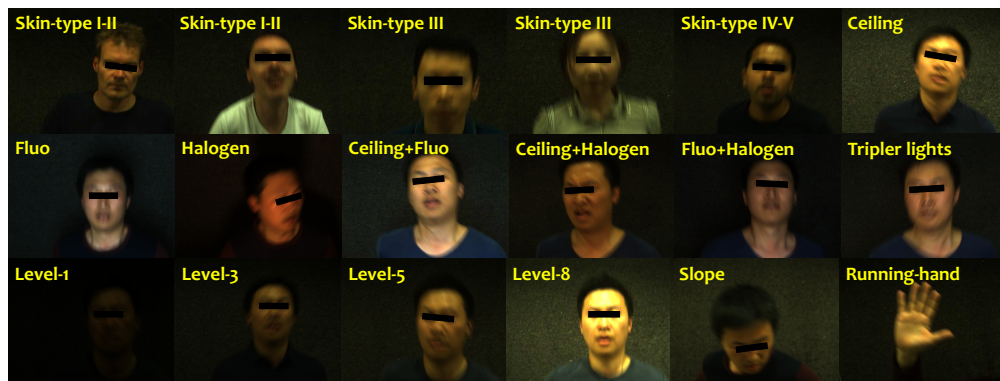
This section introduces the experimental setup for the benchmarking. First, a challenging fitness video dataset is created. Next, two evaluation metrics are presented. Finally, three filtering methods are compared as a function of pre-processing for eight existing rPPG algorithms.

4.1. Benchmark dataset

The purpose of our benchmark is to verify the effectiveness of the proposed ASF as the pre-processing step in rPPG algorithms, in particular in dealing with the motion challenges in fitness



(a) The recording setup used for creating the benchmark video dataset. It includes three different light sources and involves subjects with different skin-tones. The ceiling light is a fluorescent lamp emitting light at a 60° angle, while the frontal fluorescent light and frontal halogen light provide a frontal illumination on the face.



(b) Snapshots of some benchmark videos, which show different challenges included in our recordings such as skin-tone variations, different light sources or luminance intensity levels, other body-parts, etc.

Fig. 3. Experimental setup and video snapshots.

applications. To this end, we create a benchmark dataset containing 23 videos (with 161,051 frames) recorded from different subjects running on a treadmill. Note that these benchmark videos are completely different from the 22 stationary videos recorded for investigating the relative pulsatile amplitude in our earlier experiment. The videos are recorded by a regular RGB camera (Global shutter RGB CCD camera USB UI-2230SE-C from IDS, with 640×480 pixels, 8 bit depth, and 20 fps) at a constant frame rate in an uncompressed bitmap format. The ground-truth/reference is the contact-based ECG-signal sampled by the NeXus device (The wireless physiological monitoring and feedback device. The type of the device is NeXus-10 MKII) and synchronized with the video frames. This study has been approved by the Internal Committee Biomedical Experiments of Philips Research, and informed consent has been obtained from each subject.

Figure 3(a) illustrates the experimental setup. Unless mentioned otherwise, each video is recorded using the following default settings: the camera is placed at about 2 meters in front of the subject running on the treadmill, which results in approximately 20,000 skin-pixels given the used optics. The default subject is a male adult with a skin-type III according to the Fitzpatrick scale, and his face region is recorded for the pulse extraction. The subject is illuminated by the office ceiling light (i.e., fluorescent lamp) with an illumination direction oblique to the skin-normal, which is a common lighting condition in the fitness environment. During the recording, the subject varies the running speed between low-intensity (3 km/h) and high-intensity (12 km/h)

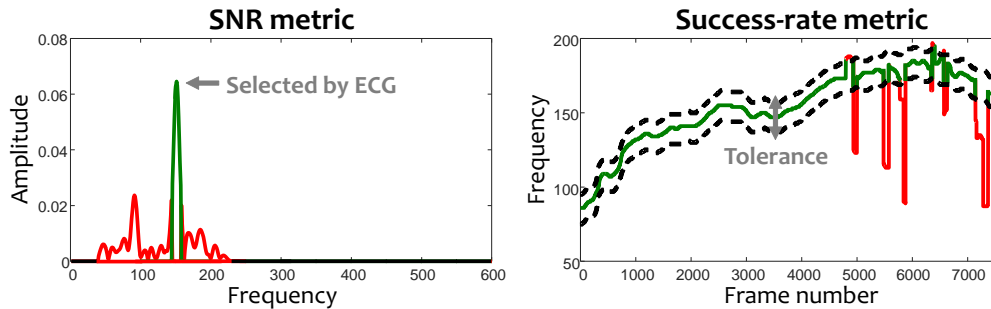


Fig. 4. Illustration of the two quality metrics (SNR and success-rate) used for evaluating the rPPG performance. In the SNR metric, the frequency components of pulse (green) and noise (red) are defined by the ECG-reference. In the success-rate metric, the inlier estimates (green) and outlier estimates (red) are defined by a tolerance (dashed black line) w.r.t. the ECG-reference.

within 5-8 minutes, depending on his endurance. The background is a skin-contrasting cloth to facilitate the skin-segmentation, which we regard as an independent research challenge outside the scope of this paper.

To thoroughly investigate the functionality of ASF, we include various realistic challenges in the recordings by changing the default experimental settings. These challenges include: different skin-types, light sources (i.e., fluorescent and halogen lamps), luminance intensity levels (i.e., from dark to bright), and other body-parts (i.e., running hand). Since ASF is designed for reducing large motion distortions in general but not for a specific challenge like skin-tone, we only perform an overall analysis/comparison on the entire dataset instead of the categorized individual challenges. Fig. 3(b) exemplifies the snapshots of some benchmark videos. Since a skin-contrasting background is used in the recording setup, we apply a simple thresholding method in YCrCb space [12] to segment the skin-region across the video and save the temporal RGB traces of spatially averaged skin-pixels for processing (i.e., pulse extraction). In this way, we ensure that the experiment relies on the minimal non-rPPG techniques, to highlight the effect/essence of the proposed method and facilitate the replication of the experiment.

4.2. Evaluation metric

To evaluate the quality of estimated rPPG-signals, we used two metrics: SNR and success rate. The SNR and success rate measure the cleanness and correctness of the output signal, respectively.

- **SNR** The Signal-to-Noise-Ratio (SNR) metric used by [5] is adopted. The SNR is derived by the ratio between the energy around the fundamental pulse frequency and remaining components within [40, 240] bpm in the frequency spectrum, where the fundamental pulse frequency is precisely located by the reference ECG-signal recorded in parallel (see Fig. 4). Since the pulse frequency of an exercising subject is time-varying, we use a sliding window to measure the SNR of the extracted pulse-signal in a short time-interval, and average the SNR measured from different time-intervals as the output metric value. More specifically, the length of the sliding window used for measuring SNR is 256 frames (6.4 s in 20 fps camera), with the sliding step 1 frame. As mentioned earlier, ASF shall be assessed on the entire dataset, the final quality indicator for each method is the globally averaged SNR over 23 video sequences.

- **Success-rate** We also measure the success-rate of the rPPG-signals, where the “success-rate” refers to the percentage of video frames where the absolute difference between the measured rPPG-frequency and reference ECG-frequency is bounded within a tolerance range (T). The rPPG-/ECG- frequency is the index of the maximum frequency peak of the rPPG-/ECG- spectrum (see Fig. 4). To enable the statistical analysis, we estimate a success-rate curve by varying $T \in [0, 10]$ (e.g., $T = 3$ means allowing 3 bpm difference), and use the Area Under Curve (AUC)

Table 1. The globally averaged SNR (dB) over 23 benchmark videos of each rPPG algorithm by using different filters in the pre-processing. The bold entry denotes the best result obtained by each rPPG algorithm using the corresponding filter.

Pre-processing	G	G-R	HUE	PCA	ICA	CHROM	PBV	POS
None (baseline)	-15.84	-8.39	-6.90	-11.00	-8.96	-4.62	-4.22	-4.12
BPF	-15.16	-7.81	-5.32	-8.38	-6.04	-3.38	-1.84	-2.57
ASF	-9.16	-3.80	-1.78	-1.67	-1.01	-1.23	-0.40	-0.23
ASF+BPF	-8.42	-3.20	-1.20	-0.78	-0.32	-0.70	0.37	0.51

Table 2. The AUC of success-rate obtained each rPPG algorithm over 23 benchmark videos by using different filters in the pre-processing. The bold entry denotes the best result obtained by each rPPG algorithm using the corresponding filter.

Pre-processing	G	G-R	HUE	PCA	ICA	CHROM	PBV	POS
None (baseline)	0.00	0.03	0.04	0.07	0.08	0.07	0.11	0.06
BPF	0.05	0.23	0.37	0.32	0.35	0.57	0.69	0.62
ASF	0.15	0.52	0.63	0.64	0.67	0.67	0.72	0.72
ASF+BPF	0.30	0.59	0.71	0.72	0.72	0.79	0.83	0.83

as the output quality indicator (i.e., larger AUC means better performance). Note that the AUC is normalized by 10, the total area. Similar to the SNR, the success-rate of an rPPG algorithm is measured across all video frames in the entire dataset.

We mention that the use of ECG in this work is limited to the experimental setup. It is not essential to the application of rPPG, i.e., ECG is used as a reference to verify whether an rPPG measurement is correct or not. This is similar to all prior works/studies that need a ground-truth for benchmarking. In the off-the-shelf fitness applications, the rPPG can be used independently, without the assistance of ECG.

4.3. Compared methods

We compared three filtering methods, i.e., Band-Pass Filter (BPF), Amplitude-Selective Filter (ASF), and ASF + BPF, as the pre-processing step in eight existing rPPG algorithms, i.e., G [1, 13], G-R [14], HUE [15], PCA [2], ICA [3], CHROM [5], PBV [6], and POS [7]. The baseline for each rPPG algorithm is the bare version (None) without pre-processing, i.e., only the core algorithm for pulse extraction is addressed. Note that the recently developed 2SR method [4] is not used in the benchmark as it does not use the temporal RGB- signals as the input for pulse extraction. This is to make a fair comparison between the different pre-processing methods by using exactly the same RGB-signals as the input, i.e., 2SR uses the spatial covariance matrix and is thus not compatible with RGB-signals, although its essence is similar to HUE [15]. Both the filters and core rPPG algorithms have been implemented in MATLAB and run on a laptop with an Intel Core i7 processor (2.70 GHz) and 8 GB RAM. The implementation of ASF strictly follows Algorithm 1.

We stress that our benchmark focuses on comparing different filters, but not on comparing different core rPPG algorithms. Thus only the parameters of filters are changed, while the parameters of core rPPG algorithms are fixed according to the original papers. Four parameters are defined for the benchmarked filters: the temporal window length (L) for pre-processing the RGB-signals, the frequency-band ($[b_{min}, b_{max}]$) for BPF, the maximum amplitude threshold (a_{max}) and the small offset (Δ) for ASF. The default parameter settings are: $L = 128$ frames, $[b_{min}, b_{max}] = [6, 24]$, $a_{max} = 0.002$, and $\Delta = 0.0001$. The parameters related to ASF will be varied for investigating their sensitivity.

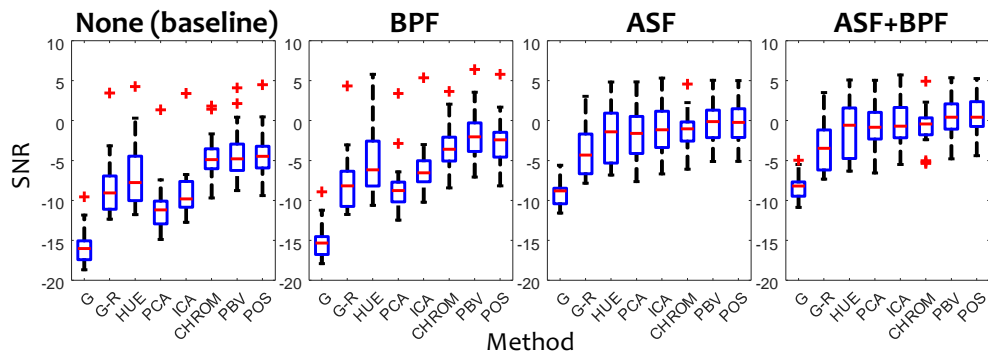


Fig. 5. The SNR comparison between eight rPPG algorithms as a function of pre-processing. Different panels show the results obtained by using different filters (e.g., None, BPF, ASF and ASF+BPF) in the pre-processing, where the median values are indicated by red bars, the quartile range by blue boxes, the full range by whiskers, disregarding the outliers (red crosses).

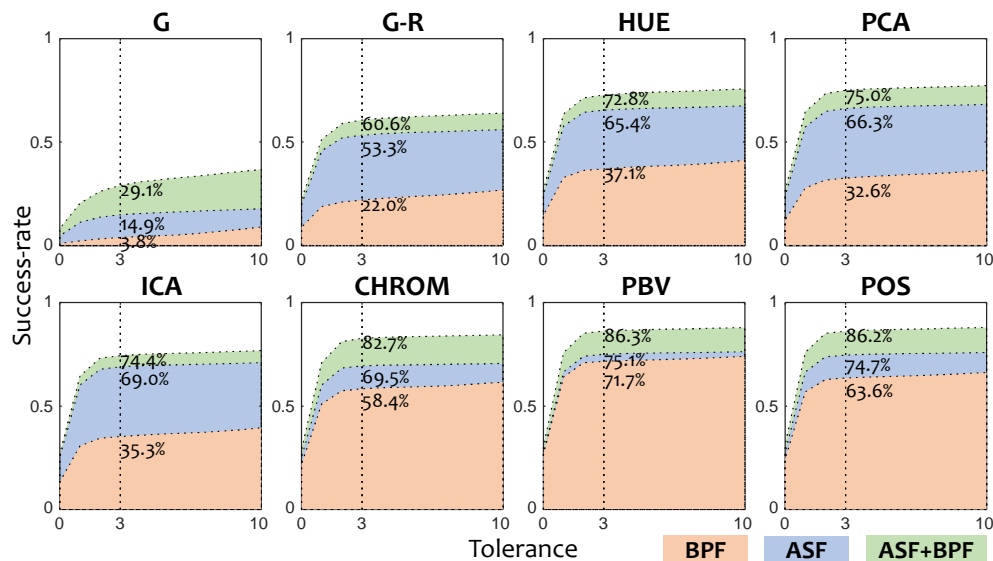


Fig. 6. The success-rate curves (and corresponding AUC) obtained by eight rPPG algorithms over 23 benchmark videos by using different filters in the pre-processing. Each panel shows the contribution of three filters (i.e., BPF, ASF and ASF+BPF) to a particular rPPG algorithm, where different colors denote the AUC for different filters and the percentage numbers exemplify their success-rate at $T = 3$, i.e., allowing 3 bpm difference with the ECG-reference.

5. Results and discussion

This section presents the benchmarking results. We first discuss the overall performance of different filters on the entire dataset, and then discuss the parameter sensitivity of ASF.

5.1. Overall discussion

Table I-II summarizes the globally averaged SNR and the AUC of success-rate of eight rPPG algorithms (obtained over 23 benchmark videos) for different filters in the pre-processing, based on the default parameter settings specified in Section 4.3. It shows that (i) all benchmarked filters, when being used as the pre-processing step, improve the baseline result of all rPPG algorithms to

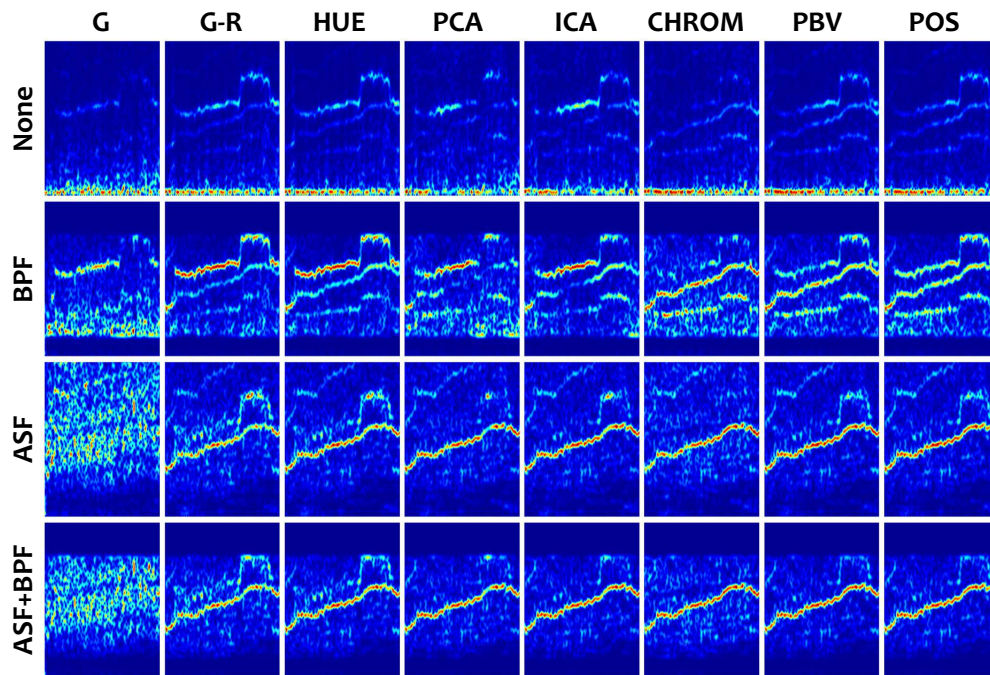


Fig. 7. Spectrograms obtained by eight rPPG algorithms on a fitness video by using different filters in the pre-processing. From top to bottom: baseline without pre-processing (None), Band-Pass Filter (BPF), Amplitude-Selective Filter (ASF), and the combination of ASF and BPF (ASF+BPF).

different extents; (ii) ASF yields the most substantial improvement from the baseline, where the SNR difference between rPPG algorithms largely disappears except for the simplest approaches that either only use G or G-R; and (iii) ASF+BPF, i.e., the combination of two filters, achieves the best performance, although its improvement on top of ASF is not as large as that from the baseline to ASF. The distribution of the SNR over 23 videos in Fig. 5 confirms our observation that ASF brings the largest improvement. Fig. 6 shows success-rate curves of individual rPPG algorithms by using different filters. We observe that (i) all the improvements gained in SNR are reflected by the success-rate of pulse-rate estimation, i.e., higher SNR implies higher success-rate; (ii) ASF is particularly beneficial for non-model based rPPG (e.g., G-R, HUE, PCA and ICA), for which it achieves a success-rate that is almost twice higher than that of BPF. The improved results of BSS-based approaches suggest the success of noise suppression, as the cleaned AC-components in RGB-signals lead to the correct pulse extraction, including the source de-mixing and pulsatile component selection.

Figure 7 exemplifies the spectrograms of eight rPPG algorithms by using different filters in the pre-processing. It shows that (i) BPF mainly eliminates the large low-frequency distortions in the baseline (possibly due to the motion drift) that are obviously outside the assumed pulse frequency-band (e.g., [6, 24]), but cannot deal with the in-band distortions (e.g., horizontal and vertical body motions). We mention that the “motion drift” could be due to the fact that the subject can hardly stay in the same position on a treadmill when he/she is running, but drifts to different positions during the long-term exercise. The motion drift is a (long-term) low-frequency distortion as compared to the instant horizontal and vertical body motions due to running. The unsuppressed motion components in RGB-signals are especially harmful for BSS-based approaches, i.e., the large periodic motion variations may drive the signal de-mixing and also confuse the component selection, having the motion-source retrieved in the end (see the spectrograms of PCA and ICA

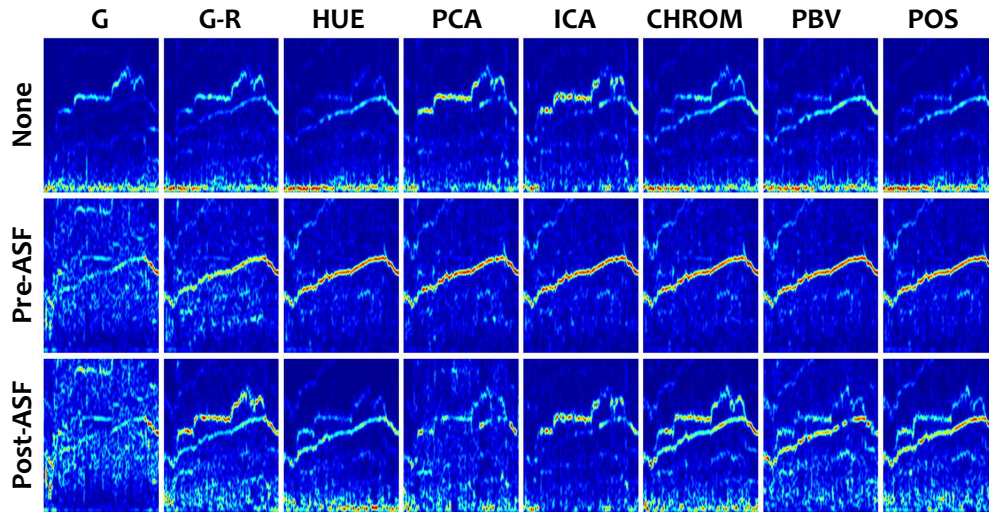


Fig. 8. Spectrograms obtained by eight rPPG algorithms on a fitness video by using ASF in the pre-processing (Pre-ASF) or the post-processing (Post-ASF). From top to bottom: baseline without pre-/post- processing (None), Pre-ASF, and Post-ASF.

using BP). Even model-based approaches cannot remove these in-band distortions completely (see two clear motion-frequency traces remaining in the spectrograms of PBV and POS using BP); (ii) ASF significantly suppresses the distortions in the spectrograms of all rPPG algorithms except G, especially it improves the non-model based approaches. By its nature, ASF eliminates the large motion distortions across the entire frequency spectrum. This is highly attractive for fitness applications, where the vigorous body motions may occur in any frequency range and may well be in the pulse frequency-band during the exercise, but can hardly enter the (lower) pulsatile amplitude-range. If body motions enter the pulsatile amplitude range, they are obviously comparable in amplitude to the pulse and therefore less challenging for the core rPPG algorithm to handle. Nevertheless, small periodic distortions may still be problematic for the component selection in BSS-based approaches; and (iii) ASF+BPF gives the cleanest spectrogram for each algorithm. Adding BPF on top of ASF to further restrict the out-band noise will, anyway, improve the results, but we conclude that ASF dominates the improvement and BPF only marginally adds to that.

Figure 8 demonstrates that pre-processing of the color-channels with ASF is to be preferred over post-processing of the pulse-signal, as the pre-processing cleans the RGB-signals at an earlier stage and improves the pulse extraction. The main reasons for the observed benefits in the ASF pre-processing are the following: (i) ASF eliminates large motion-induced color variation directions (e.g., specular changes), thus correcting the color projection direction estimated by BSS-based/model-based approaches for pulse extraction; (ii) ASF removes large motion components, especially improving the component selection in BSS-based approaches, i.e., the motion components in fitness are usually periodic. Using ASF in the post-processing can more or less clean the spectrogram as compared to the baseline results, as it removes the large distortions residing in the pulse-signal. However, putting ASF in the post-processing is much less effective, which is in line with our earlier expectation.

The benchmark shows that by adding only the ASF pre-processing step, the task of choosing a core rPPG algorithm for pulse extraction becomes less critical. This is different from the conclusions drawn in earlier studies [4–7] that selecting a proper rPPG algorithm for a specific task is highly important, because different algorithms show very different performances, i.e., model-based rPPG (e.g., CHROM, PBV and POS) is much more robust than non-model based

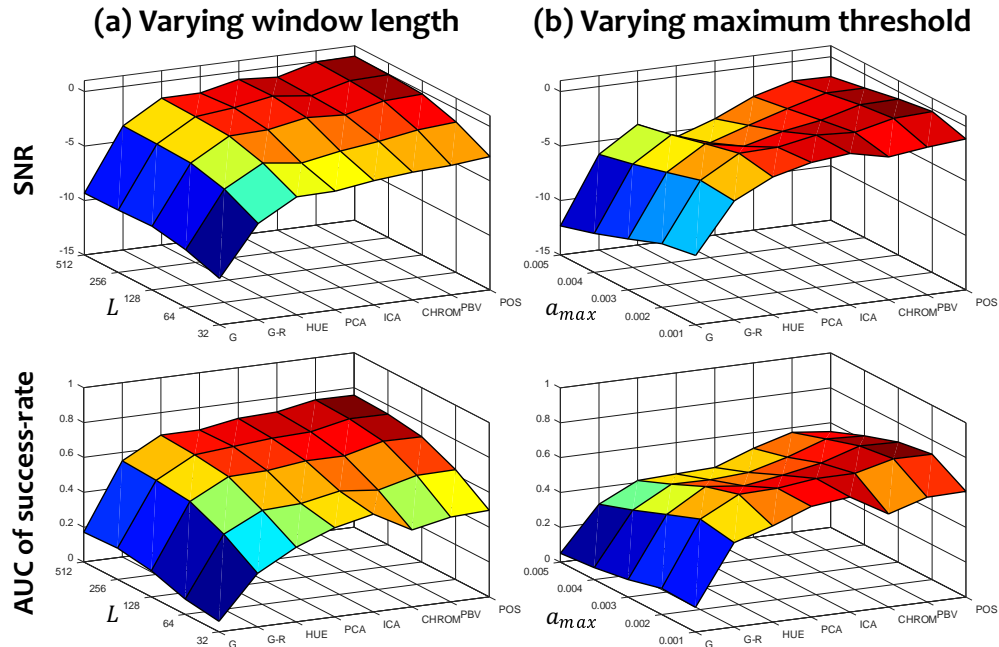


Fig. 9. The 3D mesh shows different SNR (first row) and AUC of success-rate (second row) when varying the parameters of ASF for eight rPPG algorithms. The two varied parameters are: the sliding window length L (first column) and the maximum amplitude threshold a_{max} (second column). When changing the investigated parameter, the other one remains constant. The red/blue color represents the high/low values for SNR and AUC of success-rate.

rPPG (e.g., PCA and ICA) in fitness. This paper shows that the performance differences between the various rPPG algorithms are minimal when using ASF as a pre-processing tool. But we also note that the single channel method (G) is hardly improved when combining it with ASF, which suggests that exploiting the multi-channel information of an RGB camera (i.e., channel combination) to cancel distortions is still essential for creating a robust rPPG.

We stress that the spectrograms in Fig. 7 and Fig. 8 are plotted from the raw rPPG-signals given by different methods, without the assistance of ECG, i.e., ECG is only used to check whether the rPPG spectrograms are correct or not.

5.2. Parameter discussion

To investigate the parameter sensitivity of ASF, we vary two parameters, i.e., the sliding window length (L) and the maximum pulsatile-amplitude threshold (a_{max}), in the default settings and re-run the overall benchmark for each rPPG algorithm. The Δ in ASF, only serving to prevent zero-entries in the frequency spectrum, is not varied as it is not expected to be critical to the filtering performance when it has been set to a small value. Note that BPF is not considered in this experiment, as we focus on validating the independent performance of ASF.

Figure 9 shows different SNR and AUC of success-rate for eight rPPG algorithms when setting (a) L to [32, 64, 128, 256, 512] and (b) a_{max} to [0.001, 0.002, 0.003, 0.004, 0.005], separately. From Fig. 9(a), we observe that for each rPPG algorithm, a longer window length improves both the SNR and success-rate. We expect this is because a longer time-signal has higher frequency resolution that improves the separability of pulse and motion frequencies. The results obtained by the default setting $L = 128$ in the overall benchmark are not optimal. Obviously, the gains in performance from using a long window come at the price of an increased latency. From Fig. 9(b), we observe that (i) changing a_{max} causes variations in both the SNR and success-rate, although

the effect is not significant; (ii) increasing a_{max} leads to quality drops for all rPPG algorithms, as the larger maximum threshold may include the motion components as well. However, a large a_{max} can never make the results worse than the baseline version without filtering. The filter will become less selective, but will not introduce additional distortions/artifacts; (iii) decreasing a_{max} may cause serious problems, since it creates a more selective (narrow-band) filter that may also suppress the pulse-induced signal components.

As a final remark, we emphasize that the proposed ASF is a generic filtering method that is compatible with a broad range of rPPG algorithms using RGB-signals as the input. The proposed ASF is a principle-new method that is simple/intuitive to understand, easy to implement, computationally low-cost, and very effective in dealing with significant noise distortions in a measurement, which also shows a large potential to be extended/optimized in future.

6. Conclusion

In this paper, we exploit a new biometric signature, i.e., the relative pulsatile amplitude, to design a very effective yet computationally low-cost filtering method for improving the robustness of rPPG. Based on the observation that the human relative pulsatile amplitude varies in a specific lower range as a function of RGB channels, we use the spectral amplitude of, e.g., the R-channel, to select the frequency components in RGB channels within the assumed pulsatile amplitude-range for pulse extraction. We named this method “Amplitude-Selective Filtering” (ASF), which uses the amplitude to eliminate noise distortions, instead of the more common frequency criterion used in the Band-Pass Filtering (BPF). The proposed ASF can be used as a pre-processing step in general rPPG algorithms to improve their robustness. Our benchmark containing challenging fitness videos shows that using ASF (ASF+BPF) in the pre-processing brings significant and consistent improvements. It improves different multi-channel pulse extraction methods to the extent where quality differences between individual approaches almost disappear. The novelty of the proposed method is using the simple amplitude-based pre-filtering to achieve large improvements for different rPPG methods in challenging fitness applications. The proposed method is easy to understand, simple to implement, and low-cost in running. It is the first time that the physiological property of pulsatile amplitude is used as a biometric signature for generic signal filtering.

Funding

The Philips Research and Eindhoven University of Technology financially supported the research (project number: 10017352 Vital Signs Monitoring).

Acknowledgments

The authors would like to thank Dr. Wim Verkruijsse at Philips Research for the discussion, Mr. Ger Kersten at Philips Research for creating the video recording system, and also the volunteers from Philips Research and Eindhoven University of Technology for their efforts in creating the benchmark dataset.