

Color-Distortion Filtering for Remote Photoplethysmography

Wenjin Wang¹, Albertus C. den Brinker², Sander Stuijk¹, 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

Abstract—This paper introduces a powerful filtering method that exploits the physiological and optical properties of skin reflections to improve the performance of remote photoplethysmography (rPPG). Based on the fact that the pulsatile and non-pulsatile (e.g., intensity and specular changes) components have different reflection-spectra in a multi-wavelength camera, we propose to use their different characteristic color changes as a soft criterion to filter the RGB-signals in the frequency domain, such that the AC-components containing clear color distortions can be suppressed before the actual pulse extraction. This leads to a novel “Color-Distortion Filter” (CDF) that can be used as a common pre-processing step for arbitrary rPPG algorithms to increase their robustness. The benchmark in challenging fitness recordings shows that CDF brings significant and consistent improvements to all benchmarked rPPG algorithms, and drives all multi-channel approaches to a similar high quality-level.

I. INTRODUCTION

Remote photoplethysmography (rPPG) enables contactless monitoring of human cardiac activity by measuring the pulse-induced subtle color changes in the light reflected from the human skin using a regular RGB camera [1]–[3]. Recently, several robust rPPG algorithms have been proposed. These include Blind Source Separation (BSS) based approaches (PCA [5] and ICA [6]), a data-driven approach (2SR [9]), and model-based approaches (CHROM [7], PBV [8] and POS [10]). A thorough in-depth review of these algorithms can be found in [10]. All these methods separate pulse and motion-induced distortions using a linear combination of color channels, but rely on different assumptions to determine the combining weights. The model-based approaches demonstrate superior robustness in dealing with various practical challenges due to leveraging the physiological and optical properties of the skin.

All methods are fundamentally limited, as the linear combination of three (n) color channels can maximally eliminate (or be independent of) two (n-1) AC-distortions. Inspired by filtering techniques such as the Band-Pass Filter (BPF), we recognize that it is possible to deal with different spectral components independently in the frequency domain. In a similar vein, we propose a new filtering method that exploits the essence of model-based rPPG algorithms, i.e., different reflection spectra of PPG and optical distortions in a multi-wavelength camera, to eliminate the spectral components that have no pulsatile contributions but which originate from clear color distortions. In particular, the spectral components that are clearly oriented to the presumed color variation directions of motion-induced intensity and specular changes are attenuated. We expect that using such a filter as a

pre-processing step will add the benefits of the model-based rPPG algorithms to other alternative algorithms, and also lessen the impact of their mathematical limitation by reducing the number of distortions upfront.

To design such a filter, we use the knowledge [10] that the PPG-signal varies the reflection-spectra in a different color direction w.r.t. the motion-signal. Our basic idea is to suppress the AC-components of the input RGB-signals that suffer from clear color distortions irrelevant to the pulse. To this end, we first project the RGB frequency components to a direction orthogonal to the expected color variation directions of distortions (e.g., intensity and specular changes), and then use the ratio of the energy projected in this direction and the total energy as a measure to weight the RGB frequency components. This should suppress the components due to motion relative to the components of the pulse. We call this method the “Color-Distortion Filter” (CDF), which can be used as a common pre-processing step for arbitrary rPPG algorithms to clean the RGB-signals before extracting the pulse. A benchmark on challenging fitness recordings¹ shows that CDF brings substantial and consistent improvement to all benchmarked rPPG algorithms, and drives all multi-channel approaches to a high quality-level where differences between the individual approaches almost disappear.

The remainder of this paper is structured as follows. In Section II, we describe the Color-Distortion Filter with step-by-step reasoning. In Section III and IV, we benchmark the proposed filter and discuss its performance. Finally in Section V, we draw our conclusions.

II. METHOD

The fundamental assumption used for designing our Color-Distortion Filter, grounded on the findings of [7], [8], [10], is that *the cardiac- and motion- induced signal-components have different relative strengths in the RGB camera-outputs, expressed as different characteristic AC/DC color variations*. We use this assumption as the criterion to filter the RGB-signals in the frequency domain (see the flowchart in Fig. 1). In the following text, we shall describe the proposed method in detail. Unless stated otherwise, vectors and matrices are denoted as boldface characters throughout this paper.

We denote the raw temporal RGB-signals measured by a video camera pointing at living-skin as \mathbf{C} , where \mathbf{C} is a $3 \times L$ matrix with RGB-channels sorted in rows and L is

¹“Fitness recordings” refer to the videos recorded from a subject running on a treadmill in an indoor fitness environment.

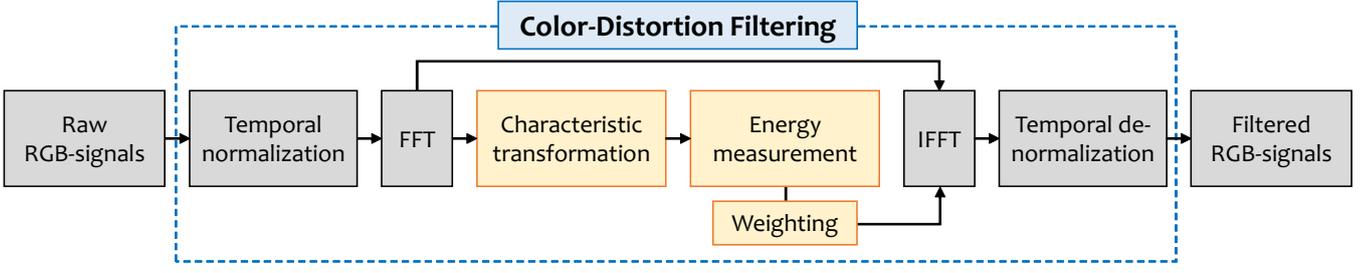


Fig. 1: Flowchart of the proposed Color-Distortion Filter. The essential step (yellow boxes) is using the different characteristic color changes of the pulsatile and non-pulsatile components as a criterion (through the characteristic transformation) to suppress the RGB frequency components containing clear color distortions.

the temporal length of the recording. In line with the model-based rPPG algorithms [7], [8], [10], we first eliminate the dependency of \mathbf{C} on the average skin reflection color (DC-level). This can be achieved by temporal normalization:

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

where $\tilde{\mathbf{C}}_i$ denotes the zero-mean color variation signal in the i -th channel and $\mu(\cdot)$ denotes the averaging operator that calculates the mean of the signal. To enable independent analysis of multiple AC-components in $\tilde{\mathbf{C}}_i$, we transform it into the frequency-domain using the Fast Fourier Transform (FFT):

$$\mathbf{F}_i = \text{FFT}(\tilde{\mathbf{C}}_i), \quad (2)$$

where \mathbf{F}_i denotes the frequency spectrum of the i -th channel and $\text{FFT}(\cdot)$ denotes the FFT operator. The RGB frequency spectra are stacked in rows of \mathbf{F} to form a $3 \times L$ matrix. We use \mathbf{V}_b to denote the column vector at a certain frequency bin b of \mathbf{F} , representing the b -th AC-component across RGB-channels. For example, a Band-Pass Filter (BPF) can be used to remove the \mathbf{V}_b outside of the human pulse-rate band (e.g., [40, 240] beats per minute (bpm)), exploiting the frequency-index b . In our approach, we use the color variation direction of \mathbf{V}_b , in the temporally normalized RGB space, to determine whether it is dominated by a pulse-signal or distortions, i.e., the G-channel has the largest pulsatile variations, followed by the B- and R- channels. However, this cannot be easily turned into a robust numerical filtering

algorithm in cases where the pulse-specific color changes are invisible in the pulsatile components, this is especially true in challenging fitness use-cases where large motion distortions dominate \mathbf{F} (see Fig. 3 (a)).

Since [10] shows that in case of large distortions a projection to the Plane Orthogonal to the Skin-tone (POS) is helpful for creating robust rPPG solutions, we adopt the same strategy here by defining a 3D orthonormal space (i.e., a space = a plane + an orthogonal axis), where the plane defined by two axes contains presumable color distortions (i.e., intensity and specular) and the remaining axis is orthogonal to this distortion plane. We transform \mathbf{V}_b to the newly defined space to analyze its behavior. This step can be expressed as:

$$\mathbf{S}_b = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ -1/\sqrt{6} & 2/\sqrt{6} & -1/\sqrt{6} \end{pmatrix} \cdot \mathbf{V}_b, \quad (3)$$

where \mathbf{S}_b denotes the transformation of \mathbf{V}_b in the new space (with unit length). The definition of this space follows the same reasoning as [7], [10]: (i) the three projection-axes must be orthogonal to each other, such that the transformed-signals are linearly independent, and (ii) they are related to the characteristic color changes of different sources, i.e., the first and second axes correspond to the color variation directions of the intensity and specular changes [10]. Thus the third axis orthogonal to the plane defined by the first two axes is independent of presumed color distortions, while, on the other hand, this direction will contain pulse content if

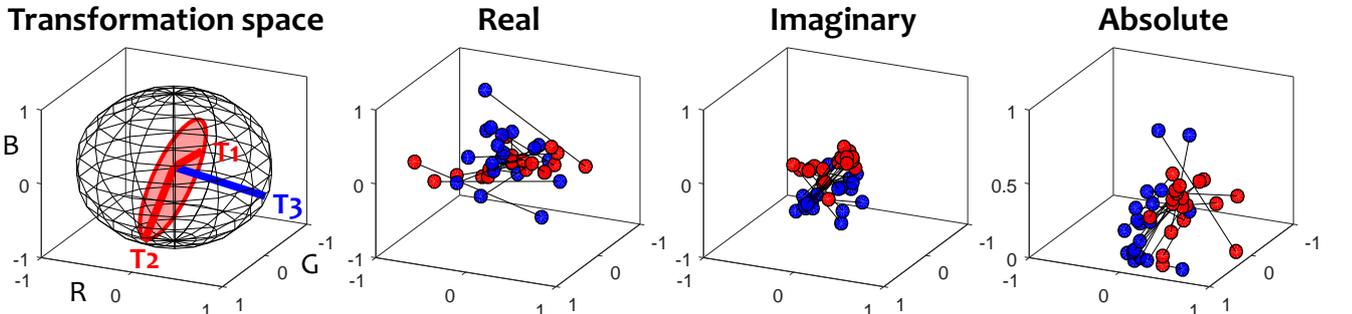


Fig. 2: In the 3D transformation, the T1- and T2- axes together define a plane (red) containing color distortions (i.e., intensity and specular), and the T3-axis is orthogonal to this distortion plane. The transformation of the spectral components from RGB space (red points) to the new space (blue points) is performed on the complex values, i.e., including not only the spectral amplitudes but also the phase information.

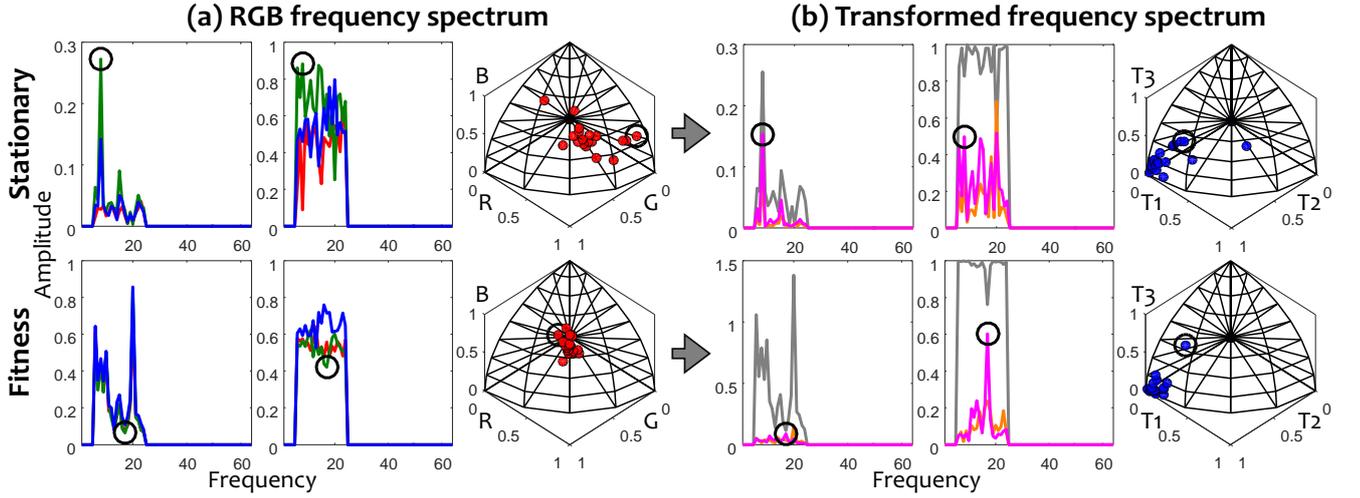


Fig. 3: The frequency spectra and the color directions of AC-components (L2-norm normalized) in the (a) temporally normalized RGB space and (b) transformed space (i.e., gray, orange and magenta denote the first, second and third transformed-signals, respectively). The black circle annotates the pulsatile component identified by the reference PPG/ECG signal (i.e., ground-truth). The frequency spectra in (a) and (b) are from the same stationary and fitness cases, but visualized in different color spaces. The pulse-specific color variation direction in the RGB frequency spectra (e.g., a relatively large G) is only visible in the stationary case, but not in the fitness case where motion-induced large color variations dominate the spectra (i.e., all AC-components are oriented in the intensity variation direction of $[1, 1, 1]^T$, meaning that the amplitudes of all three channels are equally strong). In contrast, the largest amplitude of the T3-signal (magenta) identifies the correct pulsatile components for both cases.

the pulse-signal holds this frequency.

Fig. 2 illustrates the defined characteristic color space and also a transformation of a real example. We stress that the transformation of (3) is performed on the complex values of \mathbf{V}_b , so the phase information is exploited as well. We expect that the transformed-signal on the third axis, i.e., orthogonal to the distortion plane, contains less distortions, and thus its pulsatile features shall be more visible and more prominent than those in the original RGB space (see Fig. 3 (b)). Note that the transformation axes (or combining coefficients) in (3) are rough estimates, which do not need to be very accurate as our proposed filter is merely a pre-processing step. The actual projection direction used for pulse retrieval is left to the core rPPG algorithm following this pre-processing step.

As shown in Fig. 1, we next use the energy contributions of the transformed components in (3) as a measure to quantify how much a \mathbf{V}_b suffers from large color distortions, i.e., if a \mathbf{V}_b has larger energy on the third axis w.r.t. the total energy on the three axes, it means that this \mathbf{V}_b has less energy on the distortion plane. Practically speaking, we only need to know the third axis, as the total energy can directly be estimated from \mathbf{V}_b . So we simplify (3) as:

$$S_b = (-1/\sqrt{6} \quad 2/\sqrt{6} \quad -1/\sqrt{6}) \cdot \mathbf{V}_b, \quad (4)$$

and define the relative contribution of S_b in the total energy of \mathbf{V}_b as:

$$W_b = \begin{cases} \frac{S_b \odot \text{conj}(S_b)}{\sum \mathbf{V}_b \odot \text{conj}(\mathbf{V}_b)}, & \text{if } b \in \mathbf{B} = [b_1, b_2], \\ 0, & \text{elsewhere,} \end{cases} \quad (5)$$

where \odot denotes the element-wise multiplication; $\text{conj}(\cdot)$

denotes the complex conjugate; and $\mathbf{B} = [b_1, b_2]$ denotes the assumed human pulse-rate band (e.g., $[40, 240]$ bpm). \mathbf{B} reduces the out-band distortions (i.e., clear non-pulsatile components) and, more importantly, other physiological signals that have the same characteristic color changes as the pulse, such as the respiration and Mayer-wave (i.e., oscillations of arterial pressure occurring spontaneously in conscious subjects) [16]. Note that the purpose of using the power of the spectrum in (5) to derive the weights is for increasing the separability between pulsatile and non-pulsatile components in the following weighting step.

Clearly, larger W_b in (5) means that the corresponding \mathbf{V}_b has larger pulsatile contributions w.r.t. motion distortions. To suppress the RGB frequency components containing clear color distortions, we directly use W_b to weight \mathbf{V}_b as:

$$\hat{\mathbf{F}}_i = \mathbf{W} \odot \mathbf{F}_i, \quad (6)$$

where $\hat{\mathbf{F}}_i$ denotes the weighted frequency spectrum of the i -th channel and $\mathbf{W} = [W_1, W_2, \dots, W_L]$, i.e., its elements vary within $[0, 1]$ based on their definition. Afterwards, $\hat{\mathbf{F}}_i$ is transformed back into the time-domain using the Inverse Fast Fourier Transform (IFFT) and de-normalized as:

$$\hat{\mathbf{C}}_i = \mu(C_i) \cdot (\text{real}(\text{IFFT}(\hat{\mathbf{F}}_i)) + 1), \quad (7)$$

where $\text{IFFT}(\cdot)$ denotes the IFFT operator; $\text{real}(\cdot)$ takes the real part of a complex number. Note that DC of the color is re-instated to the filtered signals to keep the original meaning of RGB-channels (i.e., the temporal de-normalization step in Fig. 1), as required by some rPPG algorithms such as the HUE-based method [4]. Since the weighting of (6) retains the original physical meaning of the RGB-channels, the filtered

outcome \hat{C} can directly be used as the input to arbitrary rPPG systems using the RGB-signals for pulse extraction, without the need for modifying the core rPPG algorithms.

The complete algorithm of the proposed Color-Distortion Filter (CDF) is shown in Algorithm 1. We kept the CDF-algorithm as clean and simple as possible to highlight the essence of our idea and to facilitate the replication, i.e., the implementation only requires a few lines of Matlab code.

Algorithm 1 Color-Distortion Filter

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

- 1: **Initialize:** $B = [b_1, b_2]$ ([40, 240] bpm adapted to L)
- 2: $\hat{C} = \text{diag}(\text{mean}(C, 2))^{-1} * C - 1$
- 3: $F = \text{fft}(C, [], 2)$
- 4: $S = [-1, 2, -1]/\sqrt{6} * F$
- 5: $W = (S * \text{conj}(S))./\text{sum}(F * \text{conj}(F))$
- 6: $W(:, 1 : B(1) - 1) = 0; W(:, B(2) + 1 : \text{end}) = 0$
- 7: $\hat{F} = F * \text{repmat}(W, [3, 1]);$
- 8: $\hat{C} = \text{diag}(\text{mean}(C, 2)) * (\text{real}(\text{ifft}(\hat{F}, [], 2)) + 1)$

Output: Filtered RGB-signals \hat{C}

III. 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, two filtering methods are compared when being used as a pre-processing step for eight existing rPPG algorithms.

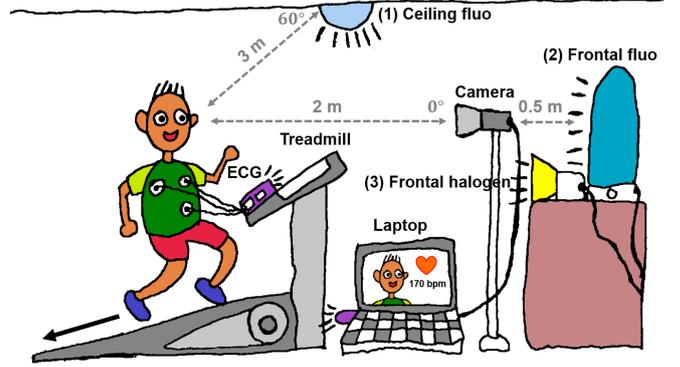
A. Benchmark dataset

The purpose of our benchmark is to verify the effectiveness of the proposed CDF as a pre-processing step in rPPG algorithms, in particular in dealing with the motion challenges in fitness applications. To this end, we create a benchmark dataset containing 23 videos (with 161,051 frames) recorded from different subjects running on a treadmill. A total of 6 healthy subjects (5 males and 1 female, age from 24 to 35) participate in our recordings. The videos are recorded by a regular RGB camera² at a constant frame rate in an uncompressed bitmap format. The ground-truth/reference is the contact-based ECG-signal sampled by the NeXus device³ and synchronized with the video acquisition. This study has been approved by the Internal Committee Biomedical Experiments of Philips Research, and informed consent has been obtained from each subject.

Fig. 4 (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 focal length. The default subject is a male adult with a skin-type III according to the Fitzpatrick scale, and his face is recorded

²Global shutter RGB CCD camera USB UI-2230SE-C from IDS, with 640×480 pixels, 8 bit depth, and 20 fps.

³The wireless physiological monitoring and feedback device. The type of the device is NeXus-10 MKII, with the sampling rate of 128 samples per second.



(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, and other body-parts.

Fig. 4: Experimental setup and video snapshots.

for 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) 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. On top of the default settings, we include various realistic challenges in the recordings, such as 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).

Fig. 4 (b) exemplifies 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 [17] 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

TABLE I: The globally averaged SNR (dB) obtained by eight rPPG algorithms over 23 benchmark videos, using either BPF or CDF in the pre-processing with different sliding window lengths.

Method	BPF(32)	CDF(32)	BPF(64)	CDF(64)	BPF(128)	CDF(128)	BPF(256)	CDF(256)	BPF(512)	CDF(512)
G	-13.38	-8.96	-14.44	-5.90	-15.16	-1.52	-15.61	1.24	-15.78	1.91
G-R	-6.56	-2.84	-7.33	-0.32	-7.81	2.70	-8.20	4.79	-8.32	5.34
HUE	-4.13	-1.03	-4.86	0.83	-5.32	3.49	-5.70	5.37	-5.83	5.82
PCA	-6.35	-1.38	-8.22	0.27	-8.38	3.35	-7.80	5.43	-9.02	5.55
ICA	-5.60	-1.14	-6.08	0.93	-6.04	4.11	-5.88	6.05	-6.52	5.96
CHROM	-2.11	-0.97	-2.90	0.30	-3.38	2.78	-3.83	4.53	-3.99	5.09
PBV	-1.07	-1.11	-1.52	0.20	-1.84	2.36	-2.30	3.87	-2.49	4.12
POS	-1.39	-0.22	-2.10	1.24	-2.57	3.75	-3.05	5.57	-3.25	5.93

* The **bold** entry denotes the best filtering method for each rPPG algorithm per L. The **blue** and **red** entries denote the best result of BPF and CDF obtained over all rPPG algorithms and all L. The same holds for Table II.

TABLE II: The AUC of success-rate obtained by eight rPPG algorithms over 23 benchmark videos, using either BPF or CDF in the pre-processing with different sliding window lengths.

Method	BPF(32)	CDF(32)	BPF(64)	CDF(64)	BPF(128)	CDF(128)	BPF(256)	CDF(256)	BPF(512)	CDF(512)
G	0.08	0.20	0.06	0.36	0.05	0.59	0.04	0.72	0.04	0.73
G-R	0.28	0.50	0.24	0.65	0.23	0.76	0.21	0.82	0.20	0.84
HUE	0.42	0.62	0.39	0.72	0.37	0.79	0.35	0.84	0.34	0.85
PCA	0.39	0.68	0.30	0.73	0.32	0.79	0.36	0.82	0.28	0.83
ICA	0.43	0.71	0.36	0.76	0.35	0.81	0.36	0.84	0.33	0.84
CHROM	0.64	0.72	0.60	0.76	0.57	0.81	0.54	0.83	0.53	0.84
PBV	0.74	0.70	0.71	0.75	0.69	0.80	0.66	0.84	0.64	0.84
POS	0.67	0.72	0.64	0.78	0.62	0.82	0.59	0.86	0.57	0.86

experiment. Note that vigorous body motion due to running is the dominant factor in each video, as compared to other factors like the luminance spectra and skin-tone. Also, the proposed CDF is aimed at reducing large color distortions in general but not designed for a specific challenge. Therefore, we only present an overall analysis and comparison on the entire dataset for statistical conclusions.

B. Evaluation metric

The quality of extracted rPPG-signals are measured by two metrics: SNR and success-rate, which evaluate the cleanness and correctness of the output signal, respectively. The success-rate is the quality indicator for those who are only interested in the heart-rate measurement, while the SNR provides more detailed information like “how clean the pulse is”, which is useful for those who want to measure additional information such as the instantaneous heart-rate variability.

- **SNR** The Signal-to-Noise-Ratio (SNR) metric used by [7] 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. 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 obtained 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.

- **Success-rate** The “success-rate” refers to the percentage

of video frames where the absolute difference between the rPPG-rate and reference ECG-rate is bounded within a tolerance range (T). The rPPG-/ECG- rate is measured in the frequency domain (e.g., the index of the maximum frequency peak). To enable statistical analysis, we estimate a success-rate curve by varying $T \in [0, 10]$ (i.e., $T = 3$ means allowing 3 bpm difference), and use the Area Under Curve (AUC) as the output quality indicator (i.e., larger AUC means more accurate measurement). 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.

C. Compared methods

We compare two filters: Band-Pass Filter (BPF) and Color-Distortion Filter (CDF), as a pre-processing step in eight existing core rPPG algorithms⁴, i.e., G [1], G-R [15], HUE [4], PCA [5], ICA [6], CHROM [7], PBV [8], and POS [10]. Both the filters and core 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 CDF strictly follows Algorithm 1. Note that the comparison with the raw rPPG-outputs is not included. The reason is that the rPPG methods, suffering from significant fitness motions, do not show meaningful heart-rate information without filtering, especially with the long sliding window length. For example,

⁴The recently developed 2SR method [9] is not used in the benchmark, as it does not use the temporal RGB-signals as the input but the spatial RGB correlation matrix. The projection-axis in CDF needs to be modified when combining it with 2SR. For fair comparison, we keep the input of CDF unchanged during the comparison.

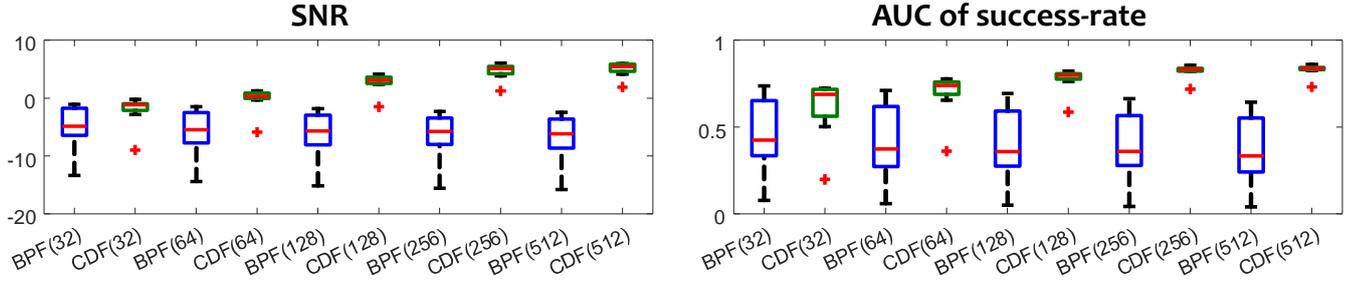


Fig. 5: The statistical comparison between BPF and CDF over eight rPPG algorithms. In each panel, the median values are indicated by red bars, the quartile range by boxes (blue for BPF and green for CDF), the full range by whiskers, disregarding the outliers (red crosses), i.e., all outliers concern the combination of CDF and G method.

the SNR for POS (without BPF) stays below -4.0 dB, and the AUC of success-rate stays below 0.4.

We stress that our benchmark focuses on comparing two different filters (i.e., BPF and CDF), but not different core rPPG algorithms. Thus only the parameters of filters are varied, while the parameters of core rPPG algorithms are fixed according to the original papers. The temporal window length L is the only parameter that needs to be defined for BPF and CDF (i.e., the pulse-rate band \mathbf{B} is automatically adapted to L). We define 5 groups of parameters to investigate the sensitivity and robustness of both filters, which are (i) $L = 32$ (1.6 s), $\mathbf{B} = [3, 6]$; (ii) $L = 64$ (3.2 s), $\mathbf{B} = [4, 12]$; (iii) $L = 128$ (6.4 s), $\mathbf{B} = [6, 24]$; (iv) $L = 256$ (12.8 s), $\mathbf{B} = [10, 50]$; and (v) $L = 512$ (25.6 s), $\mathbf{B} = [18, 100]$. Note that the unit of \mathbf{B} is not beats per minute. It is adapted to the window length L and corresponds to $[40, 240]$ bpm of a temporal sliding window.

IV. RESULTS AND DISCUSSION

Table I-II summarize the globally averaged SNR and the AUC of success-rate obtained by eight rPPG algorithms (over 23 benchmark videos) when using BPF or CDF in the pre-processing, with different sliding window lengths L (and corresponding \mathbf{B}) defined in Section III.C. Both tables clearly show that CDF outperforms BPF in all benchmarked rPPG algorithms in all conditions, except for the case of PBV with $L = 32$. In terms of the SNR, the best combination for BPF is using it in conjunction with PBV at $L = 32$, i.e., **BPF+PBV (32)** gives -1.07 dB; and (ii) the best combination

for CDF is using it in conjunction with ICA at $L = 256$, i.e., **CDF+ICA (256)** gives 6.05 dB. In terms of the success-rate, (i) the best combination for BPF is still **BPF+PBV (32)**, which gives 0.74; and (ii) the best combination for CDF is using it in conjunction with POS at $L = 256$ or 512, i.e., **CDF+POS (256 or 512)** gives 0.86.

Fig. 5 shows the statistical comparison between BPF and CDF across all benchmarked rPPG algorithms with different L . For each L , CDF has a higher median value and a lower quantile range than BPF, meaning it is statistically more accurate and stable. BPF does not have major performance differences across different L , although it is slightly better at the shorter L . The reason is that the longer L includes more distortions (due to motion) in the pulse-rate band that cannot be removed by BPF. In comparison, CDF has a clearly improved performance when increasing L , as evidenced by the increased median value and decreased quantile range. This seems reasonable, since the longer L provides an increased frequency-resolution, which allows a more accurate discrimination between pulse and distortion components. A short L , on the other hand, increases the risk that pulse and distortion components fall into the same frequency-bin and can no longer be separated by CDF, i.e., it has only 4 frequency-bins to suppress distortions at $L=32$, but 41 frequency-bins at $L=256$. However, the gains in performance from using a longer L come at the price of an increased latency. Based on Fig. 5, we conclude that the best setup for BPF (i.e., $L=32$) is still statistically worse than the worst setup for CDF (i.e., $L=32$). We note that the outliers (red

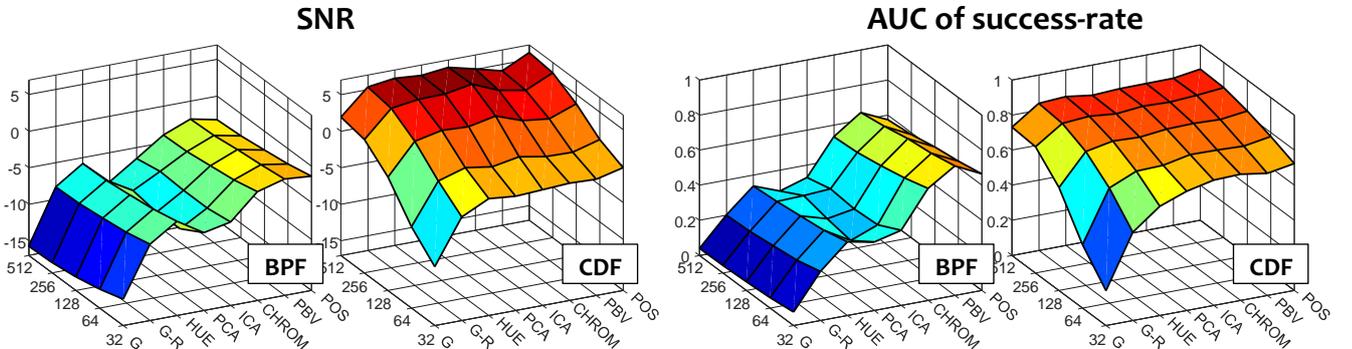
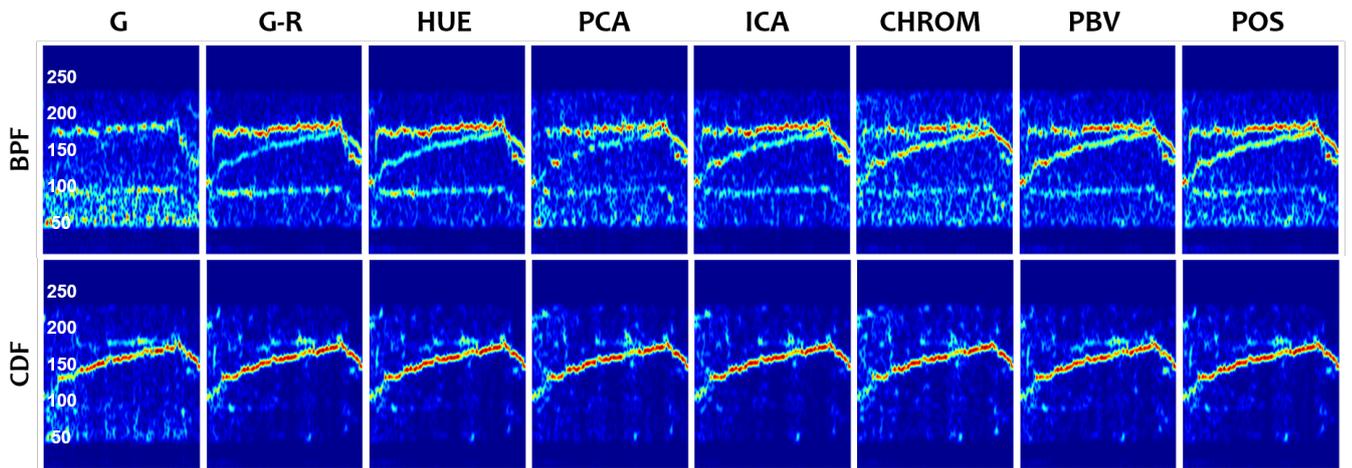
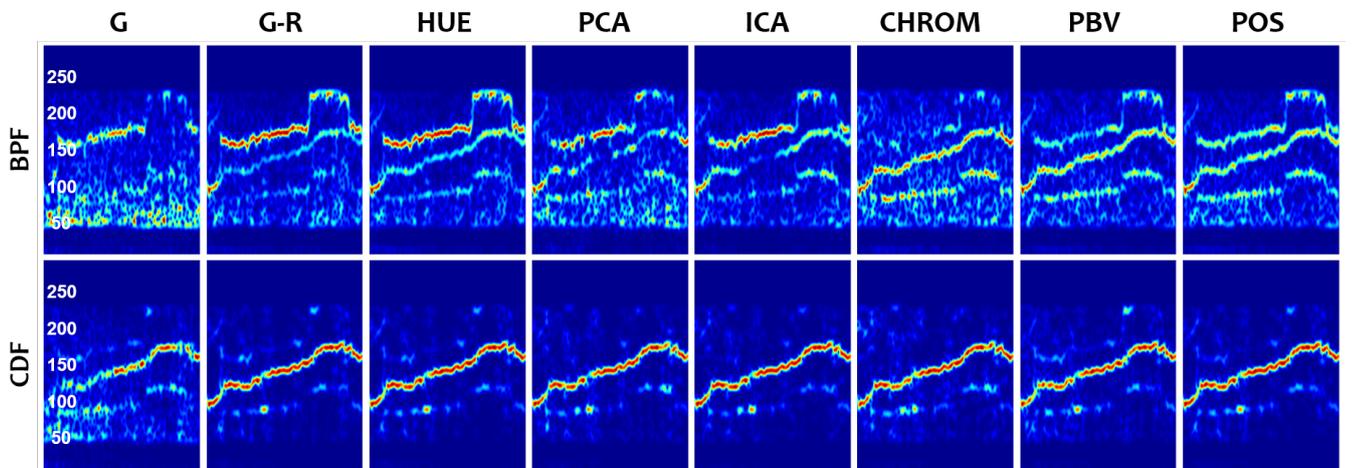


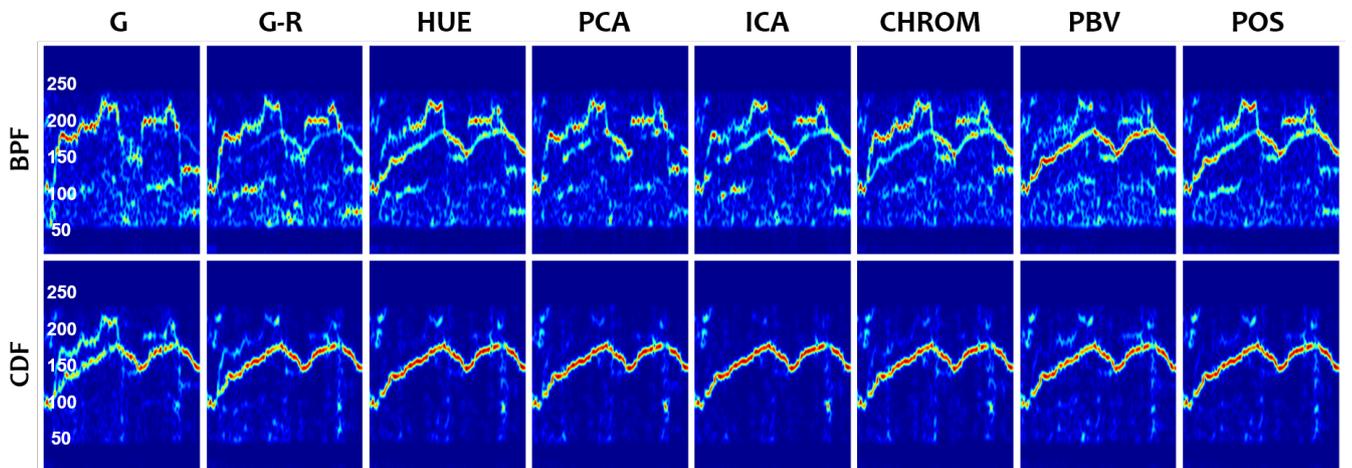
Fig. 6: The 3D meshes show SNR (left) and AUC of success-rate (right) obtained by BPF and CDF (with different sliding window lengths) in eight rPPG algorithms.



(a) Exemplified fitness video 1, where the subject ran at a constant speed (i.e., the speed is decreased before stopping) during the recording.



(b) Exemplified fitness video 2, where the subject particularly increased and decreased the running speed one time during the recording.



(c) Exemplified fitness video 3, where the subject particularly increased and decreased the running speed two times during the recording.

Fig. 7: Spectrograms obtained by eight rPPG algorithms from a fitness video, using either BPF or CDF in the pre-processing. The x-axis denotes the time (s) and the y-axis denotes the frequency (bpm).

crosses in Fig. 5) in the CDF condition stem from the G method. It shows that at the short window length $L = 32$, the single channel approach performs significantly worse than any of the multi-channel approaches.

In Fig. 6, the results of Fig. 5 are split out per core rPPG algorithm. This allows us to see which of the algorithms are responsible for the larger interquartile ranges. Looking at the BPF condition, we see that the performance of G and BSS

(PCA and ICA) methods are consistently poorer than that of the model-based methods (HUE, CHROM, PBV and POS) in the considered tough measurement conditions of fitness. This is in line with findings in [10] although the BPF was not used. This means that by using the BPF, a (presumably small) performance gain may be obtained by its addition but does not fundamentally change the results. In contrast, the deployment of the color variation direction information in CDF changes the whole picture as it brings all multi-channel approaches to a similar quality-level (with maybe the exception of the G-R at the short window length).

The fact that the BSS-based methods gain most in performance by using CDF and the model-based methods less is also understandable. There are two essential ingredients in CDF: (i) usage of the color variation direction information, and (ii) usage of the frequency-dimension to extend the degrees-of-freedom for distortion suppression. For the BSS-based methods, both ingredients of CDF lead to a performance improvement. For the model-based methods, the first ingredient was already used and thus only the second ingredient contributes to the increased performance. Having a closer look at Fig. 6, we find that it is particularly the PBV that profits least from CDF. The reason is that PBV needs sufficient amount of distortions in RGB-signals for a stable covariance-matrix inversion [8]. If the color channels are relatively clean, the unstable inversion will lead to a decreased SNR, which has been mentioned earlier in [10] to explain the lower SNR of PBV in (nearly) static subjects as compared to that of CHROM and POS. Counter-intuitively, PBV with CDF can be improved by adding synthetic noise to the cleaned RGB-signals to stabilize the matrix inversion, as reasoned in [10].

The spectrograms shown in Fig. 7 qualitatively compare BPF and CDF per rPPG algorithm when using them in challenging fitness recordings ($L=256$ in the example shown). We can see that CDF gives much cleaner spectrograms. BPF cannot eliminate the strong and periodic motion frequencies in the assumed pulse-rate band, i.e., the two clear motion-frequency traces remaining in the spectrograms are the horizontal and vertical body motions due to running. In particular, we find that the single channel G method or double channel G-R method are much improved by CDF (at $L=256$), although one should be aware that these methods profit from the availability of three color channels in CDF.

Finally, we conclude that the benchmarking shows that the quality differences between the various rPPG algorithms are minimal when using CDF as a pre-processing tool. By only adding the simple step of CDF, the task of choosing a multi-channel rPPG algorithm for pulse extraction is much less critical. Therefore, no matter what core rPPG algorithms will be used or developed in future, we recommend to use CDF as a pre-processing to clean the input RGB-signals prior to pulse extraction⁵.

⁵Though the CDF presented in this paper is restricted to the spatially averaged RGB-inputs, we believe that the same principles can be used for other cases. For non-RGB inputs (e.g., infrared or other wavelengths) the color projection direction used by CDF has to be adapted.

V. CONCLUSION

This paper introduces a novel and generic filtering method, “Color-Distortion Filter” (CDF), that significantly improves the performance of existing rPPG algorithms. Essentially, we exploit the fact that the cardiac- and motion- induced frequency components have different relative strengths in the RGB camera-outputs. We develop a metric to measure this characteristic for individual RGB frequency components, and weight them such that the components containing clear color distortions are suppressed before the actual pulse extraction. The benchmark in the challenging fitness use-case shows that CDF brings substantial and consistent improvements to all benchmarked rPPG algorithms, and drives all multi-channel approaches to a similar high quality-level. The result suggests that if CDF is standardized as a pre-processing tool, choosing a multi-channel rPPG algorithm becomes a trivial task.

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] D. J. McDuff *et al.*, “A survey of remote optical photoplethysmographic imaging methods,” in *Proc. IEEE Conf. Eng. Med. Biol. Soc. (EMBS)*, Milan, Italy, Aug. 2015, pp. 6398–6404.
- [3] Y. Sun and N. Thakor, “Photoplethysmography revisited: From contact to noncontact, from point to imaging,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.
- [4] 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. of Biomed. Opt.*, vol. 20, no. 4, p. 048002, 2015.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] W. Wang *et al.*, “A novel algorithm for remote photoplethysmography: Spatial subspace rotation,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1974–1984, Sept. 2016.
- [10] W. Wang *et al.*, “Algorithmic principles of remote-PPG,” *IEEE Trans. Biomed. Eng.*, Sept. 2016, DOI: 10.1109/TBME.2016.2609282.
- [11] X. Li *et al.*, “Remote heart rate measurement from face videos under realistic situations,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, June 2014, pp. 4264–4271.
- [12] 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.
- [13] M. Kumar *et al.*, “DistancePPG: Robust non-contact vital signs monitoring using a camera,” *Biomed. Opt. Exp.*, vol. 6, no. 5, pp. 1565–1588, May 2015.
- [14] S. Tulyakov *et al.*, “Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, June 2016, pp. 2396–2404.
- [15] M. Hülsbusch, “An image-based functional method for opto-electronic detection of skin perfusion,” Ph.D. dissertation (in German), Dept. Elect. Eng., RWTH Aachen Univ., Aachen, Germany, 2008.
- [16] C. Julien, “The enigma of Mayer waves: Facts and models,” *Cardiovascular Research*, vol. 70, no. 1, pp. 12–21, 2006.
- [17] F. Bousefsaf *et al.*, “Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate,” *Biomed. Sig. Proc. Control*, vol. 8, no. 6, pp. 568–574, 2013.