

A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation

Wenjin Wang, Sander Stuijk, and Gerard de Haan

Abstract—In this paper, we propose a conceptually novel algorithm, namely “Spatial Subspace Rotation” (2SR), that improves the robustness of remote photoplethysmography. Based on the assumption of (1) spatially redundant pixel-sensors of a camera and (2) a well-defined skin mask, our core idea is to estimate a spatial subspace of skin-pixels and measure its temporal rotation for pulse extraction, which does not require skin-tone or pulse-related priors in contrast to existing algorithms. The proposed algorithm is thoroughly assessed on a large benchmark dataset containing 54 videos, which includes challenges of various skin-tones, body-motions in complex illuminance conditions, and pulse-rate recovery after exercise. The experimental results show that given a well-defined skin mask, 2SR outperforms the popular ICA-based approach and two state-of-the-art algorithms (CHROM and PBV). When comparing the pulse frequency spectrum, 2SR improves on average the SNR of ICA by 2.22 dB, CHROM by 1.56 dB, and PBV by 1.95 dB. When comparing the instant pulse-rate, 2SR improves on average the Pearson correlation and precision of ICA by 47% and 65%, CHROM by 22% and 23%, PBV by 21% and 39%. ANOVA confirms the significant improvement of 2SR in peak-to-peak accuracy. The proposed 2SR algorithm is very simple to use and extend, i.e., the implementation only requires a few lines Matlab code.

Index Terms—Biomedical monitoring, photoplethysmography, remote sensing, colors.

I. INTRODUCTION

REMOTE photoplethysmography (rPPG) enables contactless monitoring of human cardiac activities by detecting the pulse-induced subtle color changes on skin surface using a regular RGB camera [1], [2]. This detection 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. A regular RGB camera can therefore identify the phase of the blood circulation based on minute color changes in skin reflections.

The core of rPPG is the algorithm used for pulse extraction. In recent years, several robust rPPG algorithms have been proposed. These include: (1) *Blind Source Separation* (e.g., PCA-based [3] and ICA-based [4]), which separates the temporal RGB traces into independent signal-sources using different

criteria and takes the most periodic source as the pulse; (2) *CHROM* [5], which computes the pulse-signal as a linear combination of chrominance-signals assuming a standardized skin-color to white-balance the camera; (3) *PBV* [6], which defines a \vec{P}_{bv} vector, the signature of blood volume change, to distinguish the pulse-induced color changes from motion noise in temporal RGB traces. A thorough review on the history of rPPG can be found in [7], [8]. Based on these algorithms, improvements have been made including enhancement of motion robustness [9]–[11] and rPPG-signal quality [12]. More recently, they also lead to further advances in vision-based intelligent systems, such as vital signs monitoring [13], living subject detection [14], facial expression analysis [15], mental stress detection [16], etc. All these techniques are based on the core rPPG algorithms.

In essence, all existing rPPG algorithms exploit a common spatio-temporal scheme for pulse extraction, which can be generalized as “*temporal combination of spatial color mean*”. It consists of two steps: given a video sequence containing a subject, it (1) spatially quantifies RGB values of the subject’s skin-pixels in each single frame, i.e., RGB mean, and (2) temporally creates RGB traces over multiple frames and combines them into a pulse-signal. The key difference between these algorithms is the different criteria/priors used to combine the RGB traces. The methods [3], [4], which are based on Blind Source Separation, first combine the RGB traces by linear projection and then select the most periodic independent signals as the pulse, which cannot deal with the case that motion is also periodic. For CHROM [5] and PBV [6], if the relative contribution of the blood volume pulse to the RGB channels is changed (e.g., due to different lighting spectra), the relations between RGB traces for deriving pulse will also change, and thus their fixed priors may be sub-optimal, i.e., CHROM assumes a standardized skin-color and PBV uses a pre-defined pulse signature, while both may vary a bit especially with extreme illumination spectra.

To this end, we propose a conceptually novel rPPG algorithm, namely “Spatial Subspace Rotation” (2SR), to solve the limitations in the conventional rPPG scheme. The core idea of our method is to measure the temporal rotation of the spatial subspace of skin-pixels for pulse extraction. It consists of two steps: (1) in the spatial domain, a subspace of skin-pixels is constructed in RGB space; (2) in the temporal domain, the rotation angle of spatial subspaces between subsequent frames is measured for pulse extraction. Our experiments demonstrate that when (1) multiple pixel-sensors of a regular RGB camera are used for skin sensing, and (2) a well-defined skin mask is available, 2SR outperforms the popular ICA-based method,

W. Wang and S. Stuijk are with the Electronic Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: (W.Wang@tue.nl, S.Stuijk@tue.nl).

G. de Haan is with the Philips Innovation Group, Philips Research, Eindhoven, The Netherlands, e-mail: (G.de.Haan@philips.com).

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

as well as two state-of-the-art methods (CHROM and PBV), especially for subjects with dark skin-tone or body-motions in complex illuminance conditions.

The remainder of this paper is as follows. In Section II, we analyze the concerned problems in detail and describe the proposed rPPG algorithm. In Section III and IV, the proposed algorithm is experimentally evaluated and compared. In Section V, we discuss future improvements. Finally in Section VI, we draw the conclusions.

II. METHOD

A. Spatial subspace formulation

In the spatio-temporal scheme of rPPG, the first step is to quantify RGB values of skin-pixels in each single frame as the spatial representation. All existing rPPG algorithms use the spatially averaged RGB to quantify skin-pixels, i.e., $(\bar{R}, \bar{G}, \bar{B})$. They do not consider the spatial distribution of skin-pixels in RGB space. As an alternative, we propose to take the spatial distribution of skin-pixels into account, which can be simply derived by the spatial RGB correlation:

$$C = \frac{V^T \cdot V}{N}, \quad (1)$$

where $N > 0$ is the number of skin-pixels; V is a $N \times 3$ matrix vectorized from RGB channels of skin-pixels in a video frame (e.g., each row of V is a skin-pixel while each column of V is a color channel); C is a 3×3 symmetric correlation matrix with non-negative values. Note that C is different from a covariance matrix in which the mean of V is subtracted. The reason is: $(\bar{R}, \bar{G}, \bar{B})$ contains important pulsatile information when being concatenated in the temporal domain, which is an essential element used by existing rPPG algorithms [3]–[6] for pulse extraction. If $(\bar{R}, \bar{G}, \bar{B})$ is removed from V , the pulsatile components will be eliminated in C , which makes it impossible to perform pulse extraction. By decomposing C , we can obtain the subspace of skin-pixels:

$$C \cdot U = \Lambda \cdot U \quad \text{subj. to} \quad \det(C - \Lambda \cdot I) = 0, \quad (2)$$

where $\det(\cdot)$ denotes the matrix determinant; U and Λ denote the eigenvectors and eigenvalues respectively. The eigenvalue decomposition is based on the QR algorithm [17]. Note that C is a 3×3 full-rank matrix, since the elements between RGB channels can hardly be identical in realistic situations. This is due to the 3D geometry of the skin surface, different hemoglobin and melanin concentrations in skin tissues, and the presence of independent sensor noise in RGB channels. Thus we can expand C as:

$$C = \lambda_1 \cdot u_1 \cdot u_1^T + \lambda_2 \cdot u_2 \cdot u_2^T + \lambda_3 \cdot u_3 \cdot u_3^T, \quad (3)$$

where u_i is the i -th column vector of U ; λ_i is the i -th diagonal element of Λ . In RGB space, we define U as a new axis system of skin-pixels, where (1) the principal eigenvector u_1 is the skin-vector dominating the cluster of skin-pixels (the main direction), which is a least square estimation that is robust to spatial outliers; (2) u_2 and u_3 are succeeding directions of variation that are orthogonal to u_1 .

Fig. 1 illustrates the subspaces of skin-pixels in different circumstances. There are two important properties that one

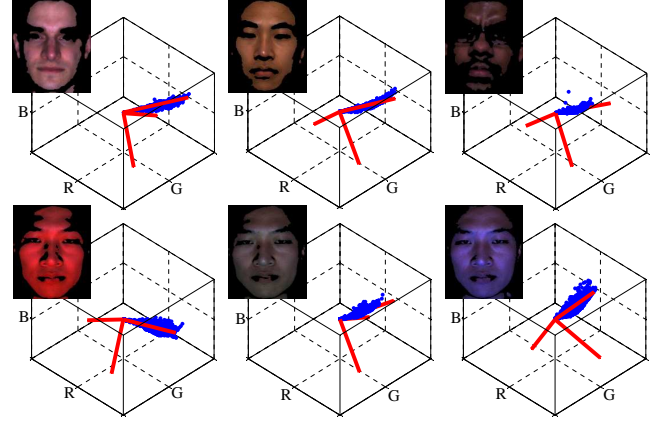


Fig. 1. An example of subspaces of skin-pixels formulated in different circumstances. The subspace (e.g., the eigenvector denoted in red) can adapt to different spatial distributions of skin-pixels (e.g., scatters denoted in blue) in variant skin-tones and illuminance conditions. The principal direction of subspace is along the main variance of the skin-pixel cluster.

should consider when looking at these figures: (1) skin-pixel values are non-negative, so u_1 can never point to the negative direction in RGB space; (2) human skin is relatively homogeneous in chromaticity (within a specific hue range) but varies in intensity due to shadows and specular reflections. Based on different intensity levels, the distribution of skin-pixels in RGB space can either be an ellipse that originates from the RGB origin or a compact sphere with uniform skin reflections. In both cases, u_1 will point to the skin-pixel cluster from the RGB origin. Since human skin is not a colorful surface, the distribution of skin-pixel can never be a sparse cloud that is spread everywhere in RGB space.

So far in this paper, the conventional spatial representation $(\bar{R}, \bar{G}, \bar{B})$ is replaced by U of skin-pixels. The essence of the proposed method is the notion to exploit the information contained in C (e.g., direction and energy) instead of the averaged skin-pixel values. Note that (1) U has to be estimated from multiple pixel-sensors using a regular camera and thus only works for rPPG, which is different from existing algorithms (e.g., CHROM and PBV) that also work for a single pixel-sensor similar to the contact-based PPG; (2) since U depends on the statistical distribution of skin-pixels, it requires a well-defined skin mask for measuring the single cluster of skin-pixels, otherwise it models the joint distribution of skin and background (e.g., multiple clusters), which would render our algorithm invalid.

B. Spatial subspace rotation

In the temporal domain, pulsatile blood causes variations in RGB channels and thus changes the subspace of skin-pixels. Since the spatial subspace is constructed from spatially redundant skin-pixels without temporal normalization (i.e., dividing the RGB channels by their temporal mean), we cannot directly use the subspace translation (i.e., distance shift of spatial RGB mean) to measure pulse. This is because that pulsatile variations without temporal normalization are proportional to the luminance intensity and thus in a multiplicative relationship. We model the temporal relation between two subspaces as an instantaneous rotation and scaling: (1) the

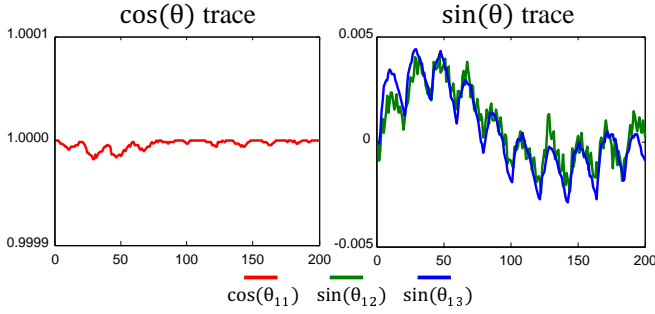


Fig. 2. A video sequence (with 200 frames) is used to illustrate the function of cosine and sine angles in the rotation matrix. The recorded subject has bright skin and remains stationary in this video. Regarding the first frame as the reference, we estimate the rotation matrix of two subspaces between the reference and subsequent frames using Eq. 4. The first row of the rotation matrix $[\cos(\theta_{11}), \sin(\theta_{12}), \sin(\theta_{13})]$ is temporally concatenated to generate $\cos(\theta)$ and $\sin(\theta)$ traces. Due to their different temporal variations, $\cos(\theta)$ trace and $\sin(\theta)$ trace are plotted in separate subfigures for visual comparison.

rotation between eigenvectors (direction change) is related to the different relative PPG-contributions in RGB channels; and (2) the change of eigenvalues (energy change) is related to the pulsatility of the measured skin.

Since the subspace U consists of eigenvectors with unit norm, its temporal changes can only cause a rotation. Thus we define a temporal stride with length l to analyze the subspace rotation. Considering the subspace in the first frame of a stride U_τ as the reference, the rotation between $U_{t,t \leq l}$ and U_τ is:

$$R = U_t^\top \cdot U_\tau = \begin{pmatrix} u_1^t \\ u_2^t \\ u_3^t \end{pmatrix}^\top \cdot (u_1^\tau \ u_2^\tau \ u_3^\tau), \quad (4)$$

where R is the rotation matrix; u_i^t denotes the i -th column of U_t . Since the L2-norm of the eigenvectors is normalized to 1, the entries in R essentially represent the cosine angles between eigenvectors:

$$\cos(\theta_{ij}) = \frac{u_i^t \cdot u_j^\tau}{\|u_i^t\| \cdot \|u_j^\tau\|}, \quad (5)$$

where i and j denote the column index; θ_{ij} denotes the rotation angle between u_i^t and u_j^τ ; $\|\cdot\|$ denotes the L2-norm; $\|u_i^t\|$ and $\|u_j^\tau\|$ are 1. Thus R can be rewritten as:

$$R = \begin{pmatrix} \cos(\theta_{11}) & \cos(\theta_{12}) & \cos(\theta_{13}) \\ -\cos(\theta_{12}) & \cos(\theta_{22}) & \cos(\theta_{23}) \\ -\cos(\theta_{13}) & -\cos(\theta_{23}) & \cos(\theta_{33}) \end{pmatrix}. \quad (6)$$

Due to the orthogonality of eigenvectors in U , $\cos(\theta_{ij,i \neq j})$ in the non-diagonal entries actually measures the sine angle changes, i.e., $\sin(\theta_{ij,i \neq j}) = \cos(\theta_{ij,i \neq j} - \frac{\pi}{2})$. Thus Eq. 6 can be rewritten as:

$$R = \begin{pmatrix} \cos(\theta_{11}) & \sin(\theta_{12}) & \sin(\theta_{13}) \\ -\sin(\theta_{12}) & \cos(\theta_{22}) & \sin(\theta_{23}) \\ -\sin(\theta_{13}) & -\sin(\theta_{23}) & \cos(\theta_{33}) \end{pmatrix}. \quad (7)$$

Since the pulse-induced color changes are minute, θ_{ij} are subtle angular changes varying around 0. Thus $\cos(\theta_{ii})$ in diagonal entries vary around 1, which cannot reflect the exact direction of rotation (always positive), i.e., the relative PPG-contributions in RGB channels become vague due to the unsigned rotation. Besides, the cosine angle is less sensitive to

subtle pulsatile fluctuations, i.e., $\cos(\cdot)$ can be considered as a kernel function here. In contrast, $\sin(\theta_{ij,i \neq j})$ in non-diagonal entries do not have the sign problem and exhibit the steepest changes when varying around 0. Fig. 2 shows an example of temporal $\cos(\theta)$ and $\sin(\theta)$ traces concatenated by the first row of R . The $\sin(\theta)$ trace presents much stronger pulsatile variations than the $\cos(\theta)$ trace, i.e., the standard deviations for $\cos(\theta_{11})$, $\sin(\theta_{12})$ and $\sin(\theta_{13})$ traces are respectively 4.3×10^{-6} , 1.6×10^{-3} and 1.9×10^{-3} . Therefore, only the non-diagonal entries in R will be used.

Pulsatile blood changes the skin-tone, which is in fact the dominant skin-vector u_1 in RGB space. Hence, only the temporal rotation of u_1 is concerned. However, as explained before, the rotation between u_1^t and u_1^τ (e.g., $\cos(\theta_{11})$) cannot be used. We only measure the rotation between the vector u_1^t and orthonormal plane $(u_2^\tau \ u_3^\tau)$ as:

$$R' = (u_1^t)^\top \cdot (u_2^\tau \ u_3^\tau) = (u_1^t)^\top \cdot u_2^\tau \ u_1^t \cdot u_3^\tau, \quad (8)$$

which are in fact $\sin(\theta_{12})$ and $\sin(\theta_{13})$. To understand the performance of this step, we show the signals produced by R' in Fig. 3 (b). In the stationary case, the blood volume pulse is the only signal-source causing the periodic rotation of the skin subspace, which produces two periodic in-phase signals. In the motion case, one direction in R' is somehow distorted by the head motion, which is related to the lighting spectrum. The other direction in R' is orthogonal to the distorted direction, and thus is more or less independent of motion distortions, i.e., the signal is still dominated by the pulse.

In addition to the subspace rotation, the eigenvalues corresponding to the variance/energy of the eigenvectors are also influenced by the pulsatile blood, which should be exploited as well. Since λ_i , decomposed from C , is powered variance, we square it and derive its scale changes as:

$$S = \sqrt{\lambda_1^t} \cdot \text{diag}\left(\begin{pmatrix} \sqrt{\lambda_2^\tau} & 0 \\ 0 & \sqrt{\lambda_3^\tau} \end{pmatrix}^{-1}\right) = \left(\frac{\sqrt{\lambda_1^t/\lambda_2^\tau}}{\sqrt{\lambda_1^t/\lambda_3^\tau}}\right), \quad (9)$$

where $\text{diag}(\cdot)$ denotes the diagonal entries of a matrix. The signals produced by S are shown in Fig. 3 (c): they represent the scale/energy change of the rotated subspace (always positive), which is in fact related to the pulsatility of measured skin. However, if skin-reflections also contain the spectra intensity changes (e.g., caused by motion distortion), S could be affected as well, i.e., the signals obtained on rotating subject are modulated by the head rotation and change in a larger range (e.g., ± 10) as compared to that of a stationary subject (e.g., ± 4). Thus we cannot only use the eigenvalues to derive the pulse. Since S is estimated with respect to the reference subspace, we can restrict its changes to the direction of subspace rotation by combining Eq. 9 with Eq. 8:

$$SR = S^\top \odot R' = \underbrace{\left(\frac{\sqrt{\lambda_1^t}}{\sqrt{\lambda_2^\tau}} \ \frac{\sqrt{\lambda_1^t}}{\sqrt{\lambda_3^\tau}}\right)}_{\text{scaling}} \odot \underbrace{(u_1^t)^\top \cdot u_2^\tau \ u_1^t \cdot u_3^\tau}_{\text{rotation}}, \quad (10)$$

where \odot denotes the element-wise multiplication. An intuitive explanation to Eq. 10 is: the dominant skin-vector u_1^t is projected and scaled on the orthonormal plane $(u_2^\tau \ u_3^\tau)$.

This step has two benefits: (1) it magnifies the amplitude of subspace rotation, i.e., pulsatility is emphasized; and (2) it suppresses the energy variations in S that do not align with the rotation direction, such as light intensity distortions. Although the magnitude of the scaling term (due to λ_1) is much larger than that of the rotation term, the scaling term is always positive and can thus be steered by the rotation term to positive/negative direction. The improved signals of the rotating subject are shown in Fig. 3 (d).

Considering a sliding window approach [5], the reference subspace U_τ is constantly changed in different temporal strides. In order to obtain the time-consistent SR over multiple strides, SR has to be analyzed in the same space, and thus is backprojected into the original RGB space:

$$\begin{aligned} SR' &= SR \cdot \begin{pmatrix} u_2^\tau \\ u_3^\tau \end{pmatrix} \\ &= \left(\sqrt{\frac{\lambda_1^\tau}{\lambda_2^\tau}} \quad \sqrt{\frac{\lambda_1^\tau}{\lambda_3^\tau}} \right) \odot \begin{pmatrix} u_1^{t\top} \cdot u_2^\tau & u_1^{t\top} \cdot u_3^\tau \end{pmatrix} \cdot \begin{pmatrix} u_2^\tau \\ u_3^\tau \end{pmatrix} \\ &= \sqrt{\frac{\lambda_1^\tau}{\lambda_2^\tau}} \cdot u_1^{t\top} \cdot u_2^\tau \cdot u_2^\tau + \sqrt{\frac{\lambda_1^\tau}{\lambda_3^\tau}} \cdot u_1^{t\top} \cdot u_3^\tau \cdot u_3^\tau \end{aligned} \quad (11)$$

The arbitrary sign problem of eigenvector decomposition is eliminated after the backprojection. In a single stride, multiple SR' between the reference frame and succeeding frames are estimated and concatenated into a trace $\overrightarrow{SR'}$. The first two traces in $\overrightarrow{SR'}$ are in anti-phase, as our example in Fig. 3 (e) shows. Similar to CHROM [5], we derive/boost the pulse-signal by combining the anti-phase traces as:

$$\vec{p} = \overrightarrow{SR'_1} - \frac{\sigma(\overrightarrow{SR'_1})}{\sigma(\overrightarrow{SR'_2})} \cdot \overrightarrow{SR'_2}, \quad (12)$$

where $\overrightarrow{SR'_i}$ is the i -th trace of $\overrightarrow{SR'}$; $\sigma(\cdot)$ denotes the standard deviation operator. Consequently, a long-term pulse-signal \vec{P} is estimated from successive strides using overlap-adding as:

$$\vec{P}^{t-l} = \vec{P}^{t-l} + (\vec{p} - \mu(\vec{p})), \quad (13)$$

where $\mu(\cdot)$ denotes the averaging operator; \vec{P} , an one-dimensional signal with length K (the total number of video frames), is initialized by zero-entries and constantly updated by \vec{p} . An example of \vec{P} is shown in Fig. 3 (f): the pulse-signal can be extracted even when RGB channels are seriously distorted by the head rotation in Fig. 3 (a).

In order to show the independent performance and improvement of the proposed method, we remove all the post-processing steps (e.g., signal smoothing) to keep the algorithm as clean as possible, i.e., even the commonly used band-pass filtering is rejected. The complete algorithm of 2SR is shown in Algorithm 1, which is very simple to use and extend, i.e., the implementation only requires a few lines Matlab code.

III. EXPERIMENTS

This section presents the experimental setup for evaluating the proposed rPPG algorithm. First, we introduce the benchmark video dataset. Next, four evaluation metrics are

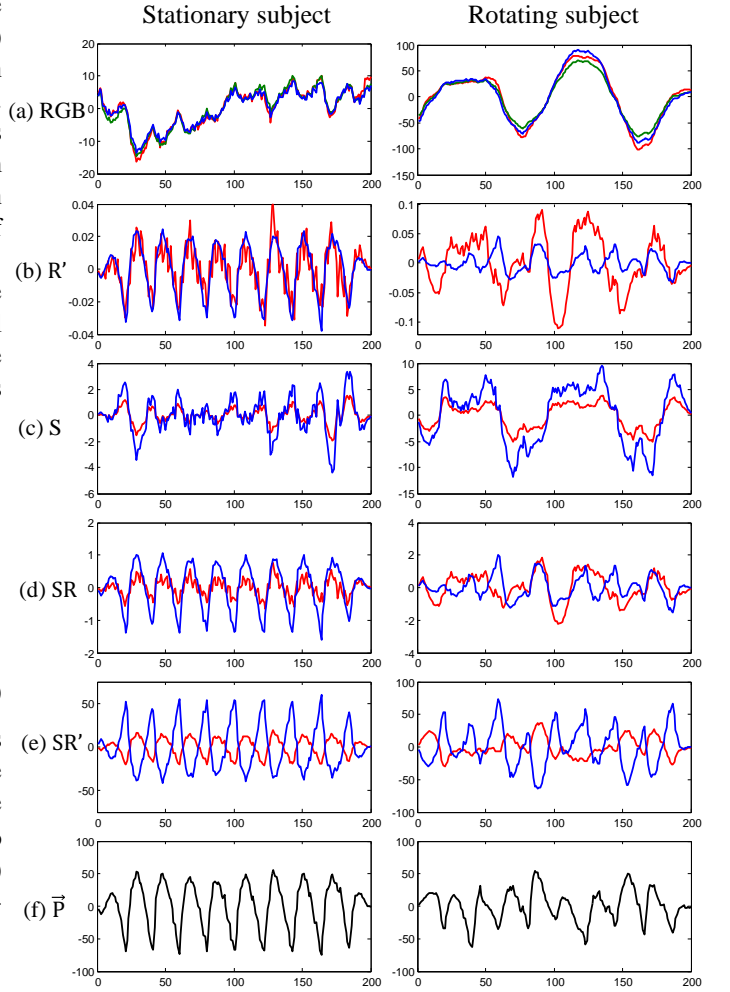


Fig. 3. Two video sequences (with 200 frames) are used to illustrate the results obtained by each step of 2SR. The subject either remains stationary or rotates head in these two videos to simulate an easy and a challenging scenario. The signals in the figure are concatenated by the vectors produced in each step of 2SR, which are respectively the (a) RGB mean traces, which is used in previous works but not this work, i.e., only for comparison purpose, (b) R' in Eq. 8, (c) S in Eq. 9, (d) SR in Eq. 10, (e) SR' in Eq. 11, and (f) pulse-signal in Eq. 13. The mean of all the signals are subtracted for visual comparison. Note that in the steps producing two signals, we use red/blue color to denote the first/second trace.

Algorithm 1 Spatial Subspace Rotation

Input: a video sequence containing K frames

```

1: Initialize:  $\vec{P} = \text{zeros}(1, K)$ 
2: for  $k = 1, 2, \dots, K$  do
3:    $C_k = \frac{V_k^\top \cdot V_k}{N}$ 
4:    $[U^k, \Lambda^k] = \text{eigs}(C^k)$  where  $U^k = \{u_i^k\}, \Lambda^k = \{\lambda_i^k\}$ 
5:   if  $\tau = k - l + 1 > 0$  then
6:     for  $t = \tau, \tau + 1, \dots, k$  do
7:        $SR' = \sqrt{\frac{\lambda_1^\tau}{\lambda_2^\tau}} \cdot u_1^{t\top} \cdot u_2^\tau \cdot u_2^\tau + \sqrt{\frac{\lambda_1^\tau}{\lambda_3^\tau}} \cdot u_1^{t\top} \cdot u_3^\tau \cdot u_3^\tau$ 
8:        $\overrightarrow{SR'} \leftarrow$  concatenated by  $SR'$ 
9:     end for
10:     $\vec{p} = \overrightarrow{SR'_1} - \frac{\sigma(\overrightarrow{SR'_1})}{\sigma(\overrightarrow{SR'_2})} \cdot \overrightarrow{SR'_2}$ 
11:     $\vec{P}^{t-l} = \vec{P}^{t-l} + (\vec{p} - \mu(\vec{p}))$ 
12:   end if
13: end for
Output: the pulse-signal  $\vec{P}$ 

```


adopted to assess the performance. Finally, the popular ICA-based approach and two state-of-the-art methods (CHROM and PBV) are compared in the benchmarking.

A. Benchmark dataset

A benchmark dataset containing 54 video sequences (with 108000 frames) has been built to evaluate the proposed rPPG algorithm. The videos are recorded with a regular RGB camera¹ in an uncompressed bitmap format, 768×576 pixels, 8 bit depth, and 20 FPS. During the recording, the subject sits in front of the camera with his/her face visible and wearing a finger-based transmissive pulse oximetry² for synchronized PPG-signal sampling - the *ground-truth*. The subject is illuminated by a fluorescent lamp³ (parallel illuminance source) that is placed 1.5 meters in front of the subject face. The non-uniform illumination produces shadows and specular reflections on the subject's face. Since the lighting condition is crucial to the rPPG technique, we simulate it as a challenge in the "body-motion" category, where the number of light-sources, color of the light-source, and types of the illumination will be introduced in details. Note that all the recordings are carried out indoors.

To thoroughly evaluate the robustness of the proposed rPPG algorithm in realistic scenarios, we performed three tests. In the first test we evaluate the effect of different skin-tones. In the second test we ask the subjects to perform head motions in front of the camera, while in the last test we record them during recovery from a running exercise. This allows us to investigate these challenges independently, as described below (the bold number in brackets indicates the number of frames simulated for this challenge):

- **Skin-tone (22500)** 15 subjects with various skin-tones are recorded and categorized into three skin-types based on the Fitzpatrick scale, i.e., participants are from West Europe (*skin-type I-II*, 5 subjects), East Asia (*skin-type III*, 5 subjects) and Sub-Sahara Africa/India (*skin-type IV-V*, 5 subjects).

- **Body-motion (31500)** To rPPG, the most significant challenge caused by body-motion is not the "motion tracking", but the modulated reflections of lighting spectra on the skin-surface, which disrupt subtle color changes induced by the pulse. Thus we combine the body-motion challenge with the illuminance challenge in this category. Ignoring the problem of tracking, three basic motion-types, i.e., *stationary*, *rotation* (rigid motion) and *talking* (non-rigid motion), are defined for a subject (*skin-type III*) to perform under 7 different illuminance conditions including single/mixture of colored light sources, i.e., fluorescent lamp, red LED lamp, green LED lamp, blue LED lamp, red-green LED lamps, red-blue LED lamps and green-blue LED lamps. Note that the colored LED lamps are point illuminance sources.

- **Recovery after exercise (54000)** In order to evaluate the robustness of rPPG to pulse-rate changes, a series of videos are recorded to analyze subjects recovering from a running exercise. In this category, 6 subjects (3 males and 3



Fig. 4. Snapshots of some recordings in the benchmark dataset. The frames in the first row are from the "skin-tone" category, where subjects have various skin colors; in the second row are from the "body-motion" category, where subjects perform head motions in different illuminance conditions; and in the third row are from the "recovery after exercise" category, where subjects achieve different levels of pulse-rate after a running exercise, i.e., all participants achieve their maximum limits in the high-level and breathe heavily or sweat profusely.

females) in skin-type I-III participate in the recordings. Each subject perform 3 different levels of running (with different intensities) by adjusting speed and gradient of the treadmill: *low* (gradient= 12° , speed=4-5 km/p), *medium* (gradient= 14° , speed=5-6 km/p), and *high* (gradient= 15° , speed=7-8 km/p). The duration of each running exercise is 3 minutes. After the exercise, the subject immediately sits in front of the camera for a video recording.

Fig. 4 shows snapshots of some recordings in our benchmark dataset. All videos are pre-processed by the OC-SVM classifier that has been used in [9] for selecting the skin-pixels. This study has been approved by the Internal Committee Biomedical Experiments of Philips Research, and informed consent has been obtained from each test subject.

B. Evaluation metrics

For comparison, the performance of rPPG is evaluated using four different metrics:

- **SNR of pulse frequency** In line with [5], the Signal-to-Noise-Ratio (SNR) of the pulse frequency is derived by the ratio between the energy around the first two harmonics and remaining parts in the frequency spectrum, where the location of the first two harmonics is determined by the reference PPG-signal. The SNR values are measured from each video and averaged in each challenge of a particular category.

- **Pearson correlation of instant pulse-rate** The Pearson correlation is applied to evaluate the correspondence of instant pulse-rates between rPPG and the PPG-reference. The instant pulse-rate, defined as the inverse of the peak-to-peak interval of the pulse-signal, is derived by a simple peak detector in the time-domain. It captures the instantaneous changes and reflects the real-time differences. For statistical analysis, the Pearson correlation is performed in each challenge per category and interpreted by the ρ -value.

- **Precision of instant pulse-rate** The instant pulse-rate of rPPG is also measured in terms of "precision", the percentage of frames where the absolute difference between the reference is under a threshold T (error tolerance). For statistical analysis,

¹Global shutter RGB CCD camera USB UI-2230SE-C from IDS.

²Model CMS50E from ContecMedical.

³Philips HF3319 - EnergyLight White.

TABLE I
SNR (μ)

Category	Challenge	ICA	CHROM	PBV	2SR
Skin-tone	Type I-II	6.51	6.47	5.57	7.44
	Type III	6.61	6.21	6.26	7.90
	Type IV-V	4.56	5.43	4.04	6.60
Body-motion	Stationary	11.61	9.42	6.57	10.53
	Rotation	4.04	3.63	6.36	6.16
	Talking	3.11	3.99	4.01	5.33
Recovery	low	1.78	2.66	1.95	4.93
	medium	1.64	3.62	3.15	5.26
	high	-0.82	3.52	3.52	4.84
Overall	Average	4.34	4.99	4.60	6.55

TABLE II
PEARSON CORRELATION (ρ)

Category	Challenge	ICA	CHROM	PBV	2SR
Skin-tone	Type I-II	0.88	0.92	0.89	0.93
	Type III	0.66	0.66	0.94	0.97
	Type IV-V	0.69	0.83	0.84	0.95
Body-motion	Stationary	0.94	0.95	0.62	0.97
	Rotation	0.27	0.29	0.82	0.89
	Talking	0.37	0.75	0.53	0.84
Recovery	low	0.76	0.73	0.64	0.98
	medium	0.73	0.93	0.85	0.99
	high	0.48	0.88	0.87	0.96
Overall	Average	0.64	0.77	0.78	0.94

TABLE III
PRECISION (AUC)

Category	Challenge	ICA	CHROM	PBV	2SR
Skin-tone	Type I-II	0.64	0.73	0.58	0.77
	Type III	0.56	0.58	0.67	0.74
	Type IV-V	0.46	0.54	0.45	0.65
Body-motion	Stationary	0.77	0.82	0.60	0.85
	Rotation	0.36	0.50	0.66	0.77
	Talking	0.31	0.57	0.45	0.65
Recovery	low	0.24	0.37	0.26	0.62
	medium	0.16	0.40	0.31	0.55
	high	0.12	0.35	0.32	0.38
Overall	Average	0.40	0.54	0.48	0.66

we estimate a precision curve by setting $T \in [0, 3]$ beats per minute (bpm) and use the Area Under Curve (AUC) to interpret the precision. Note that the AUC is normalized by 3, the total area. Similarly, the precision is evaluated in each challenge per category.

• **ANOVA** To investigate the significance of difference between compared rPPG methods, we apply the balanced one-way Analysis of Variance (ANOVA) to analyze the results obtained by each evaluation metric. The p -value of ANOVA is used for interpretation and a common threshold 0.05 is specified to determine whether the difference is significant, i.e., when $p < 0.05$, the difference is significant.

C. Compared method

The 2SR proposed in this paper, is intended as an algorithmic component in an rPPG monitoring-system [9]. Thus we compare it as clean as possible with direct algorithmic alternatives, such as the popular ICA-based [4], CHROM [5]

and PBV [6] (the state-of-the-art), but not with the complete rPPG system [9] that combines many other non-rPPG ingredients like face detection/tracking and signal smoothing. Among the compared rPPG algorithms, CHROM and PBV require skin-tone/pulse related priors, while ICA does not. All these methods 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 2SR strictly follows Algorithm 1 presented in this paper. The parameters in ICA, CHROM and PBV are set to the optimal values in the original papers, while the only parameter in 2SR is defined as $l = 20$ without tuning, i.e., the default setting according to the camera frame-rate (20 FPS) in our setup. For fair comparison, all the parameters remained identical when processing different videos.

IV. RESULTS

The experimental results of ICA, CHROM, PBV and 2SR on 54 benchmark video sequences are summarized in Table I-III. The performance of all four algorithms are compared under each challenge per category. The bold entries in tables indicate the best result obtained by the corresponding rPPG algorithm during the comparison. Fig. 5 shows the plots of Pearson correlation and precision curves.

A. Skin-tone robustness comparison

In the “skin-tone” category, the newly proposed 2SR shows the highest score in all skin-types in three evaluation metrics, which demonstrates its superior performance against ICA, CHROM and PBV in skin-tone robustness. When comparing the pulse frequency spectrum, 2SR improves on average the SNR of ICA by 1.42 dB, CHROM by 1.28 dB, PBV by 2.02 dB. When comparing the instant pulse-rate, 2SR improves on average the Pearson correlation and precision of ICA by 28% and 30%, CHROM by 18% and 17%, PBV by 7% and 27%.

All color-based rPPG methods including 2SR have problems in dealing with dark skin as the pulsatility is much lower. Our experiments show that 2SR has relatively stable performance across various skin types. The hypothesis for this improvement is: the skin subspace formulated by 2SR is adapted to the video content in real-time (e.g., a specific skin distribution), and the subspace rotation is estimated on the direction orthogonal to the skin-tone direction, where the pulse-induced subtle changes are maximized. In Fig. 5, (1) the Pearson correlation curves show that the performance of ICA, CHROM and PBV are more variant than 2SR in this category; and (2) the precision curves show that 2SR gains more improvements in skin-type IV-V (dark skin) than skin-type I-III, as compared to the benchmarked algorithms.

In 2SR, we also notice that the temporal changes of λ is much lower for dark skin as compared to bright skin. The reason is that the dark skin has higher melanin contents than the bright skin. It absorbs a portion of diffuse reflections carrying the pulse-signal, whereas the specular reflection is not reduced. The lower pulsatility of dark skin leads to lower amplitudes of pulse-induced color variations in RGB channels, and thus the smaller λ .

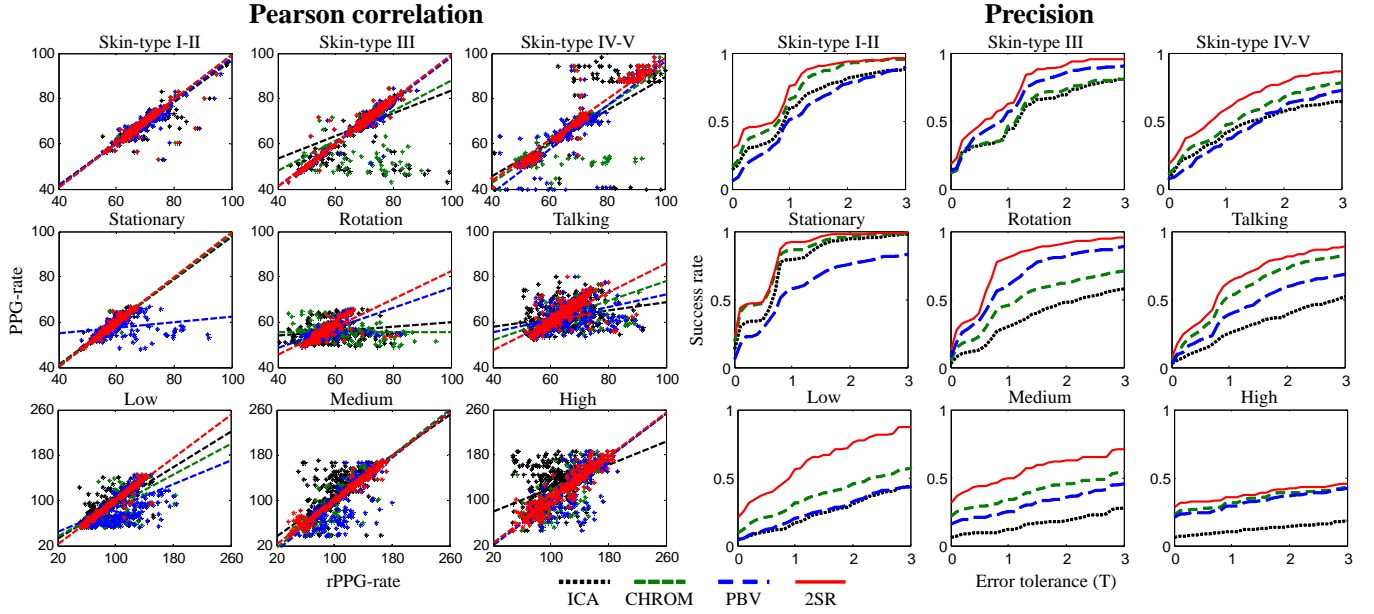


Fig. 5. In each challenge of a category, ICA, CHROM, PBV and 2SR are statistically compared using Pearson correlation and precision. Both metrics evaluate the instant pulse-rates of the rPPG-signal.

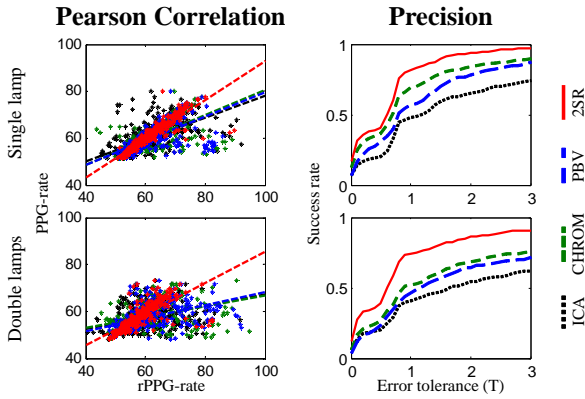


Fig. 6. Categorizing the experimental results of “body-motion” in terms of “illuminance condition”, four algorithms are compared in “single lamp” and “double lamps” challenges using Pearson correlation and precision.

B. Motion robustness comparison

In the “body-motion” category, 2SR outperforms ICA, CHROM and PBV in almost all three evaluation metrics except in the SNR comparison with ICA on stationary subjects and PBV on rotating subjects. When comparing the pulse frequency spectrum, 2SR improves on average the SNR of ICA by 1.09 dB, CHROM by 1.66 dB, PBV by 1.69 dB. When comparing the instant pulse-rate, 2SR improves on average the Pearson correlation and precision of ICA by 70% and 58%, CHROM by 36% and 20%, PBV by 37% and 33%.

As can be seen in Fig. 5, the most significant improvement of 2SR over (1) ICA and CHROM is in motion (e.g., rotation and talking), and (2) PBV is in stationary. In ICA, although the blind source separation step does not rely on the assumption of skin-tone or illuminance color, the selection of pulsatile component that based on periodicity is problematic for motion. The component containing motion frequency noise can be incorrectly selected. For example, ICA outperforms the other three algorithms in stationary subjects illuminated by different

colored light sources (e.g., SNR achieves 11.61 dB), but is much worse in rotation and talking (e.g., SNR is only 3-4 dB). The use of blood volume signature in PBV brings a modest loss in signal quality in stationary subjects, which is in line with the findings in [6], i.e., PBV performs better in motion.

Furthermore, the Pearson correlation curves in the “body-motion category” of Fig. 5 show that most errors produced by the benchmarked algorithms (e.g., scattering points far from the regression line) are on the side of higher pulse-rate (e.g., 75-85 BPM), especially in the videos containing rotation and talking. This is due to incorrectly detected peaks when estimating the instant pulse-rate, but not the pulse-rate of test subjects, i.e., the instant pulse-rate, derived from inter-beat interval, is highly sensitive to high-frequency noise or abrupt changes (e.g., motion artifacts) in the rPPG-signal. To further understand the algorithms’ performance in different lighting conditions, we categorize the results obtained in the “body-motion” in terms of “illuminance condition”: “single lamp” and “double lamps” categories, as shown in Fig. 6. It shows that 2SR performs better in both lighting categories, while its improvement in the “double lamps” category is more clear.

C. Recovery after exercise comparison

In the “Recovery after exercise” category, 2SR outperforms ICA, CHROM and PBV by showing an all-round improvement in all three evaluation metrics. When comparing the pulse frequency spectrum, 2SR improves on average the SNR of ICA by 4.14 dB, CHROM by 1.74 dB, PBV by 2.14 dB. When comparing the instant pulse-rate, 2SR improves on average the Pearson correlation and precision of ICA by 49% and 198%, CHROM by 15% and 38%, PBV by 24% and 74%.

In Fig. 5, the Pearson correlation curves suggest that 2SR has a better correlation with the PPG-reference than others. However, we notice that in the precision curves, when subjects’ pulse-rates are increased to medium and high levels

(e.g., more intensive exercise), the peak-to-peak accuracies of all benchmarked algorithms are improved at $T = 0$ (e.g., error tolerance is 0), but degraded at $T = 3$ (e.g., allowing 3-beats difference). The reason is that after running the subject's pulse amplitude is stronger, which helps estimation. On the other hand, the relatively fast change in pulse-rate renders the 3-beats error tolerance insufficient.

To further understand their performance, we show the pulse frequency spectrums obtained by PPG-reference and four rPPG algorithms on one subject (e.g., the subject 2) at three running levels in Fig. 7. We can see that (1) when the subject pulse-rate is increased from low to high levels, ICA shows noisier spectrums and performs worse. The respiratory component showing strong frequency is incorrectly selected as pulse, especially in medium and high levels where the subject breathes heavily after intensive exercise; (2) conversely, CHROM and PBV show cleaner spectrums in higher levels. Since the pulsatility of measured skin becomes stronger after subjects perform intensive exercise, it leads to easier pulse extraction. Besides, these two methods do not have problem of component selection as ICA; and (3) 2SR shows much cleaner spectrums than the other compared algorithms in all three levels, and obtains a performance similar to the PPG-reference. The improved robustness of 2SR over the compared algorithms are further validated by the statistics in "recovery" category of Table I-III. Moreover, we notice that in the medium and high levels of 2SR, the respiration frequency is well preserved in the spectrum, which does not interrupt the pulse frequency as ICA. This implies that 2SR could be useful for respiration extraction as well.

In the high-level running exercise, all subjects achieve their maximum limits before finishing the test. Although the recordings are performed immediately after the exercise, some subjects recover so quickly that their pulse-rates have already dropped prior to the recording, especially the subject 4. To investigate the algorithms' performance in this level, we show their frequency spectrums obtained on 6 subjects in Fig. 8. As can be seen, (1) ICA shows much noisier spectrums than the others, especially in subject 1 where the respiratory component confused with the pulse-rate for a long period; (2) the spectrums of 2SR are cleaner than that of CHROM and PBV, and comparable to that of PPG-reference, i.e., it even outperforms the PPG-reference in subject 2 and 6, where the PPG-sensor (finger-contacted pulse oximetry) is occasionally interfered by finger movement and produces motion artifacts.

Interestingly, we find that during the exercise recovery, the pulse-rates of (1) subject 1, 2, 3 and 6 decrease slowly and smoothly, and are not fully recovered within 150 seconds; and (2) subject 4 and 5 return to their normal states rapidly, within 80 and 120 seconds respectively. Particularly, their pulse-rates change dramatically and irregularly in a short period, as shown in both the PPG and rPPG approaches. These irregular changes in pulse-rates are not caused by motion artifacts or sensor noise, but are actual cardiac arrhythmias. These are not uncommon in endurance athletes [18], and our subject 4 and 5 are amateur endurance athletes that run a lot (e.g., at least three times per week), even half and full marathons.

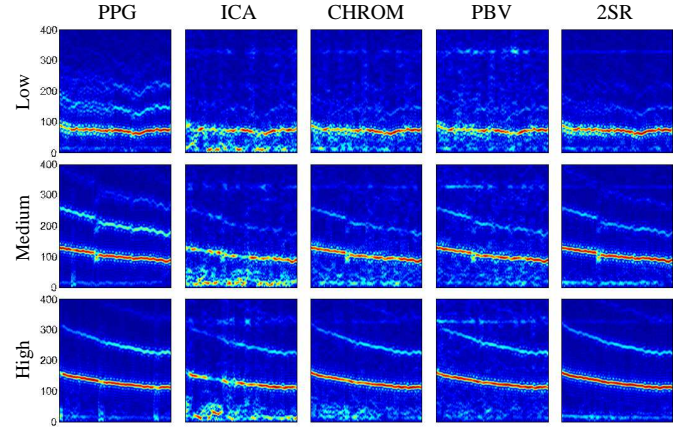


Fig. 7. The frequency spectrums obtained by PPG-reference and four rPPG algorithms on subject 2 (male) at 3 exercise levels, where the x-axis and y-axis denote the time and frequency respectively.

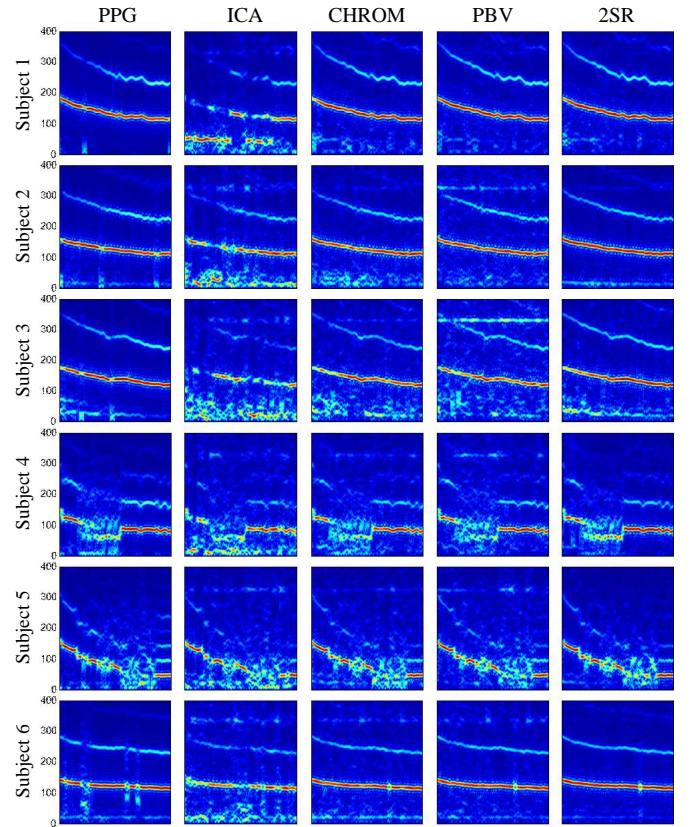


Fig. 8. The frequency spectrums obtained by PPG-reference and four rPPG algorithms on 6 subjects at the high-level exercise (most intensive), where the x-axis and y-axis denote the time and frequency respectively.

D. Overall comparison

Table IV shows the ANOVA results between three existing rPPG algorithms and 2SR in our benchmark dataset respectively, where the bold entries indicate significant improvements ($p < 0.05$). From Table IV, we conclude that in our benchmark dataset: (1) in SNR evaluation, 2SR outperforms the other three algorithms, although the improvements are not significant except in the comparison with PBV ($p = 0.031$). Since the estimation of SNR is based on the frequency spectrum energy, it is less sensitive to modest improvements; (2) in Pearson correlation and precision evaluations, the improvements of

TABLE IV
ANOVA (P-VALUE)

Comparison	SNR	Pearson correlation	Precision
ICA vs 2SR	0.121	0.001	0.009
CHROM vs 2SR	0.115	0.027	0.098
PBV vs 2SR	0.031	0.005	0.017

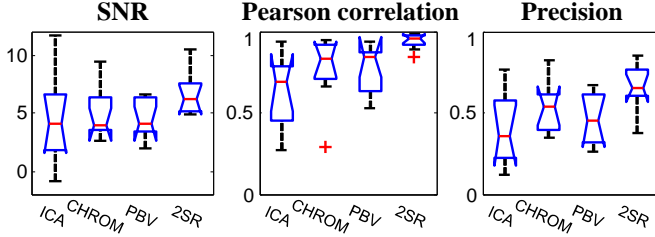


Fig. 9. The overview of performance comparison between ICA, CHROM, PBV and 2SR in three evaluations, which compares the median (red bar), standard deviation (blue box), minimum and maximum (black bar) values.

2SR over the other three algorithms are significant except the comparison with CHROM in precision ($p = 0.098$). Since these two metrics are based on the instant pulse-rate derived from inter-beat interval, it suggests that 2SR can eliminate high-frequency noises that degrade the accuracy of the peak-detector used to determine the interval; (3) the most significant improvement obtained by 2SR is in the comparison with ICA. This is in line with the findings in CHROM [5] that ICA is in general less robust to motion distortions, especially to periodic motions; and (4) the improvements of 2SR over PBV is slightly larger than that over CHROM. In the model of PBV, the blood volume pulse signature depends on the optical filters of the camera and illumination spectrum [6], which is not adapted to the different illumination conditions incorporated in our recordings. Our hypothesis is that its noise suppression may be sub-optimal.

Fig. 9 shows the overall comparison between four rPPG algorithms in our benchmark dataset. Given a well-defined skin mask, 2SR shows improved robustness over ICA, CHROM and PBV in all-round evaluations. When comparing the pulse frequency spectrum, 2SR improves on average the SNR of ICA by 2.22 dB, CHROM by 1.56 dB, and PBV by 1.95 dB. When comparing the instant pulse-rate, 2SR improves on average the Pearson correlation and precision of ICA by 47% and 65%, CHROM by 22% and 23%, PBV by 21% and 39%.

V. DISCUSSION

A. Skin-pixel cluster

As discussed in Section II, 2SR requires two preliminary conditions for reliable pulse estimation: (1) multiple skin-pixels in the spatial domain, and (2) the measured skin-pixels need to be in a single cluster. This implies that the performance of 2SR may drop when either the number of skin-pixels decreases or the skin mask is noisy, i.e., including non-skin pixels. To thoroughly understand the limitation/weakness of 2SR, we perform another quantitative comparison between benchmarked algorithms. In these experiments, we (1) reduce the number of measured skin-pixels, and (2) introduce non-skin pixels into the skin mask. Examples of the simulated

challenges are shown in Fig. 10, where each challenge is simulated in two different ways and also multiple times by using a different percentage of skin-pixels or non-skin pixels with respect to the original measurement.

Fig. 11 shows the SNR comparisons between the algorithms under the simulated challenges. In the challenge of reduced number of skin-pixels, all four algorithms show a similar quality drop in both down-sampling methods. The smaller number of skin-pixels leads to larger quantized RGB errors, which is a general problem for any rPPG algorithms, but not a particular challenge for 2SR. When increasing the number of non-skin pixels, all four algorithms suffer from performance degradations, while the quality of 2SR drops more dramatically when the percentage of non-skin pixels becomes larger, i.e., it has an obvious SNR drop when the percentage of non-skin pixels arrives between 10% and 30%. This is understandable because 2SR is designed for measuring the subspace of a single cluster distribution. The occurrence of another non-skin cluster will distort the direction of the skin-pixels' subspace, and the degree of distortion depends on the percentage of non-skin pixels. In comparison, ICA, CHROM and PBV have a lower quality drop when significant non-skin distortions occur, i.e., when the percentage of non-skin pixels is between 30% and 50%. Since the simulated non-skin cluster does not introduce extra frequencies, ICA can properly separate and select the signal source for pulse estimation. In addition, we notice that PBV performs worse than CHROM and 2SR when the skin-mask is clean, but better when skin-similar noise is introduced, which is due to its design in suppressing noise variations. The simulated comparisons suggest that the number of skin-pixels is not so critical for 2SR, but a well-defined skin mask is essential, i.e.,

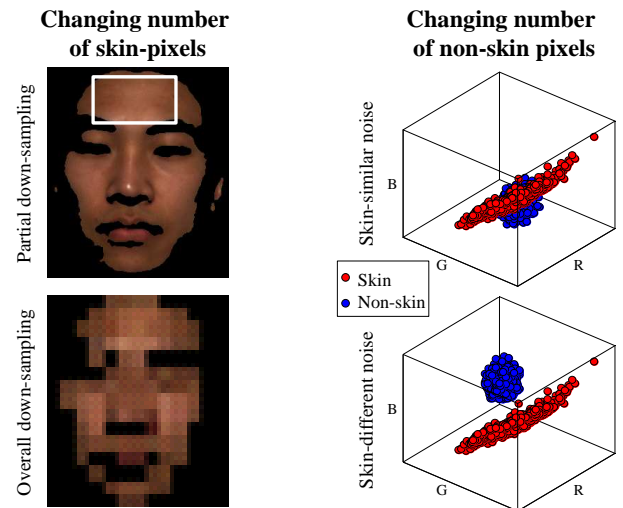


Fig. 10. To verify the limitations of 2SR, we simulate two challenges in the original recordings by (1) decreasing the number of skin-pixels, and (2) increasing the number of non-skin pixels. The number of skin-pixels is decreased in two different ways: a. subsampling a small region in face (e.g., forehead), and b. downsampling the complete face uniformly (e.g., nearest-neighbor interpolation). The number of non-skin pixels is also increased in two different ways: a. introducing a skin-similar noise cluster with RGB vector $[0.7, 0.5, 0.3]$ (e.g., skin-vector), and b. introducing a skin-different noise cluster with RGB vector $[0.5, 0.5, 0.5]$ (e.g., white light vector). Note that the simulated non-skin cluster is also spatially and temporally variant in RGB space.

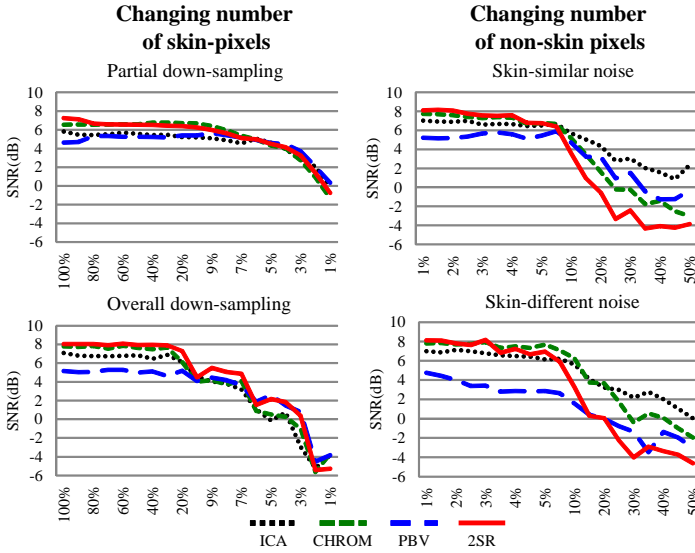


Fig. 11. The comparison between rPPG algorithms in the simulated challenges of skin cluster according to Fig. 10. Since the instant pulse-rate that measures the peak-to-peak accuracy cannot be used for measuring the extremely noisy signals (e.g., signals with negative SNR), we only use the SNR that based on frequency spectrum for comparison. The x-axis denotes the percentages of skin-pixels or non-skin pixels with respect to the original measurement.

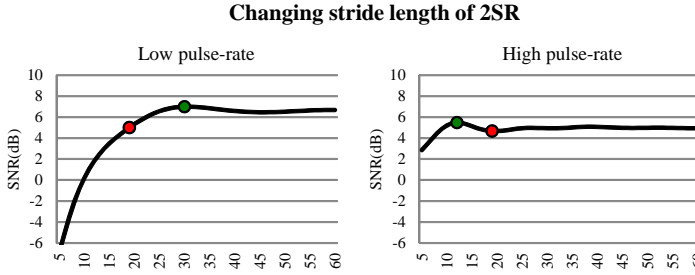


Fig. 12. The comparison between different l in 2SR on subjects with a low or a high pulse-rate. The stride length of l is changed in [5, 60] when processing the same video, as denoted in the x-axis. Similar to Fig. 11, SNR is used to interpret the quality of rPPG-signal. The red dot represents the SNR obtained by $l = 20$ (default setting in our experiment), and the green dot represents the maximal SNR can be achieved by changing l .

when a large non-skin cluster occurs in RGB space, 2SR could be worse than ICA, CHROM and PBV.

To further improve the robustness of 2SR, we propose to either (1) employ advanced techniques for finding the skin-regions, i.e., detecting pulsatile regions [14], or (2) use a better method for subspace segmentation, which can simultaneously separate multiple clusters into independent subspaces for measurement. However, it has to be mentioned that due to the integration effect of a camera sensor, some pixels may always contain a combination of skin and non-skin in a scenario when the skin-region is moving at high-speed (e.g., motion blur), which will be inherently challenging to address.

B. Temporal window stride

Although 2SR does not rely on pulse/skin-tone priors, there is still one parameter l fixed for processing, which defines the temporal stride for measuring the subspace rotation. For pulse extraction, each stride should include at least a half cardiac cycle for capturing the maximal subspace rotation induced by the blood volume pulse. Thus the optimal l is in fact based

on the camera frame-rate and subject pulse-rate. To simplify the illustration, we perform a test on two benchmark videos recorded on subjects with a low pulse-rate (around 45 BPM) and a high pulse-rate (around 93 BPM). The frame-rates of both videos are 20 FPS. We constantly change the stride length of 2SR when processing these two videos and show the corresponding SNR in Fig. 12.

As can be seen in Fig. 12, $l = 20$ (default setting) used in our experiment is not optimal for both videos: the maximal SNR in case of (1) the low pulse-rate achieves 7.00 dB when $l = 31$ (e.g., SNR = 5.33 dB when $l = 20$); and (2) the high pulse-rate achieves 5.49 dB when $l = 12$ (e.g., SNR = 4.68 dB when $l = 20$). It shows that 2SR requires longer stride for subject with low pulse-rate to obtain reasonably good results, as compared to that of subject with high pulse-rate. Besides, it shows that too short strides produce worse results than that of long strides in both videos, i.e., the SNR is even negative for subjects with low pulse-rate when $l < 10$. This is to be expected, as the rotation angle becomes very small when the reference frame is rapidly/frequently updated in every short stride. However, a longer stride (e.g., $l > 50$) does not guarantee a quality improvement, since long-term motion distortions or other frequency sources (e.g., respiration) could enter the estimation.

As a short conclusion, there is no fixed optimal l for all video recordings. One could improve upon this is to adjust l to a specific video content (e.g., based on the camera frame-rate and subject pulse-rate) and seek the optimal stride length algorithmically, i.e., l can even be adapted to time-varying pulse-rate in real-time processing. Since this work focuses on elaborating a novel concept in rPPG algorithm, we leave the steps of optimization as future work, such as parameter tuning or more complex and sophisticated rPPG systems.

VI. CONCLUSION

In this paper, we propose a conceptually novel rPPG algorithm for pulse extraction, namely ‘‘Spatial Subspace Rotation’’ (2SR). The core idea of the proposed algorithm is to estimate the temporal rotation of skin-pixels’ subspace in RGB for deriving the pulse. It exploits the benefit of statistical measurement of multiple pixel-sensors provided by a remote camera, and requires a well-defined skin mask for measuring the single cluster distribution of skin-pixels. Numerous experiments demonstrate that given a well-defined skin mask, the proposed method outperforms the popular ICA-based approach and two state-of-the-art algorithms (CHROM and PBV) in challenges of skin-tone, body-motion in complex illuminance conditions, and pulse-rate recovery after exercise.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ihor Kirenko, Dr. Wim Verkruijsse, and Dr. Bert den Brinker at Philips Research for their support, and also the volunteers from Eindhoven University of Technology for their efforts in creating the benchmark dataset.

REFERENCES

- [1] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Medical Engineering and Physics*, vol. 29, no. 8, pp. 853–857, Oct. 2007.
- [2] W. Verkruijsse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21 434–21 445, Dec. 2008.
- [3] M. Lewandowska, J. Ruminski, T. Kocajko, and J. Nowak, "Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, Sept. 2011, pp. 405–410.
- [4] M.-Z. Poh, D. McDuff, and R. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [5] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [6] G. de Haan and A. van Leest, "Improved motion robustness of remote-ppg by using the blood volume pulse signature," *Physiological Measurement*, vol. 35, no. 9, pp. 1913–1922, Oct. 2014.
- [7] Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging," *Biomedical Engineering, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [8] D. J. McDuff, J. R. Estep, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, Aug. 2015, pp. 6398–6404.
- [9] W. Wang, S. Stuijk, and G. de Haan, "Exploiting spatial redundancy of image sensor for motion robust rppg," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 2, pp. 415–425, Feb. 2015.
- [10] L. Feng, L. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, no. 5, pp. 879–891, May 2015.
- [11] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, no. 7, pp. 077 010–077 010–9, July 2011.
- [12] Y. Hsu, Y.-L. Lin, and W. Hsu, "Learning-based heart rate detection from remote photoplethysmography features," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 4433–4437.
- [13] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement*, vol. 35, no. 5, p. 807, May 2014.
- [14] W. Wang, S. Stuijk, and G. de Haan, "Unsupervised subject detection via remote ppg," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 11, pp. 2629–2637, Nov. 2015.
- [15] H. E. Tasli, A. Gudi, and M. Den Uyl, "Integrating remote ppg in facial expression analysis framework," in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI '14. New York, USA: ACM, 2014, pp. 74–75.
- [16] B. Kaur, S. Moses, M. Luthra, and V. N. Ikonomidou, "Remote stress detection using a visible spectrum camera," *Proc. SPIE*, vol. 9496, pp. 949 602–949 602–13, May 2015.
- [17] J. G. Francis, "The QR transformation—part 2," *The Computer Journal*, vol. 4, no. 4, pp. 332–345, 1962.
- [18] H. Heidbuchel, D. L. Prior, and A. La Gerche, "Ventricular arrhythmias associated with long-term endurance sports: what is the evidence?" *British Journal of Sports Medicine*, vol. 46, no. Suppl 1, pp. i44–i50, 2012.



Wenjin Wang received the B.Sc. degree in Biomedical Engineering (in top class) from Northeastern University, Shenyang, China, in 2011 and the M.Sc. degree in Artificial Intelligence (with full scholarship) from University of Amsterdam, The Netherlands, in 2013. Currently, he is a Ph.D. candidate at Eindhoven University of Technology, The Netherlands, and cooperates with the Vital Signs Camera project at Philips Research Eindhoven.

Wenjin Wang works on problems in computer vision, i.e., remote photoplethysmography (rPPG).



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



Gerard de Haan received BSc, MSc, and PhD degrees from Delft University of Technology in 1977, 1979 and 1992, respectively. He joined Philips Research in 1979 to lead research projects in the area of video processing/analysis. From 1988 till 2007, he has additionally taught post-academic courses for the Philips Centre for Technical Training at various locations in Europe, Asia and the US. In 2000, he was appointed "Fellow" in the Video Processing & Analysis group of Philips Research Eindhoven, and "Full-Professor" at Eindhoven University of

Technology. He has a particular interest in algorithms for motion estimation, video format conversion, image sequence analysis and computer vision. His work in these areas has resulted in 3 books, 3 book chapters, 180 scientific papers and more than 180 patent applications, and various commercially available ICs. He received 5 Best Paper Awards, the Gilles Holst Award, the IEEE Chester Sall Award, bronze, silver and gold patent medals, while his work on motion received the EISA European Video Innovation Award, and the Wall Street Journal Business Innovation Award. Gerard de Haan serves in the program committees of various international conferences on image/video processing and analysis, and has been a Guest-Editor for special issues of Elsevier, IEEE, and Springer.