Exploiting Spatial-redundancy of Image Sensor for Motion Robust rPPG

Wenjin Wang, Sander Stuijk, and Gerard de Haan

***Abstract*—Remote photoplethysmography (rPPG) techniques can measure cardiac activity by detecting pulse-induced colour variations on human skin using an RGB camera. State-of-the-art rPPG methods are sensitive to subject body motions (e.g., motion-induced colour distortions). This study proposes a novel framework to improve the motion robustness of rPPG. The basic idea of this work originates from the observation that a camera can simultaneously sample multiple skin regions in parallel, and each of them can be treated as an independent sensor for pulse measurement. The spatial-redundancy of an image sensor can thus be exploited to distinguish the pulse-signal from motion-induced noise. To this end, the pixel-based rPPG sensors are constructed to estimate a robust pulse-signal using motion-compensated pixel-to-pixel pulse extraction, spatial pruning, and temporal filtering. The evaluation of this strategy is not based on a full clinical trial, but on 36 challenging benchmark videos consisting of subjects that differ in gender, skin-types and performed motion-categories. Experimental results show that the proposed method improves the SNR of the state-of-the-art rPPG technique from 3.34dB to 6.76dB, and the agreement ($\pm 1.96\sigma$) with instantaneous reference pulse-rate from 55% to 80% correct. ANOVA with post-hoc comparison shows that the improvement on motion robustness is significant. The rPPG method developed in this study has a performance that is very close to that of the contact-based sensor under realistic situations, while its computational efficiency allows real-time processing on an off-the-shelf computer.**

***Index Terms*—Biomedical monitoring, photoplethysmography, remote sensing, motion analysis.**

I. Introduction

CARDIAC activity is measured by medical professionals to monitor patients' health and assist clinical diagnosis. The conventional *contact-based* monitoring methods, i.e., electrocardiogram (ECG) and photoplethysmography (PPG), are somewhat obtrusive and may cause skin-irritation in sensitive subjects (e.g., skin-damaged patients, neonates). In contrast, *camera-based* vital signs monitoring triggers a growing interest for non-invasive and non-obtrusive healthcare monitoring.

Earlier progress made in camera-based vital signs monitoring can be categorised into two trends: (1) detecting the minute optical absorption variations of the human skin induced by blood volume changes during the cardiac cycle, i.e., remote-PPG (rPPG) [1], [2], [3]; (2) detecting the periodic head motions caused by the blood pulsing from heart to head via the abdominal aorta and carotid arteries [4]. However, both the colour-based and motion-based approaches are sensitive to body motions, since these can dramatically change the light reflected from the skin surface and also corrupt the subtle head motion driven by the cardiovascular pulse. Although significant progress has been reported in the rPPG-category for a fitness setting recently [3], the Signal-to-Noise Ratio (SNR) of the pulse-signals obtained by all existing methods are still reduced when the subject is moving relative to the camera.

The goal of this paper is to significantly improve the SNR of the rPPG pulse-signal by better exploiting the spatial redundancy of the image-sensor. To some extent, the spatial-redundancy of the image-sensor has already been exploited in previous rPPG methods [1], [2], [3] as they extract the pulse-signal from the *averaged* pixel value in a skin region. Such averaging of independent sensors is optimal only if the (temporal) noise-level in skin pixels is comparable and has a Gaussian distribution. However, the image-to-image variations in skin pixels from a face may be very strong in the mouth region of a talking subject, while relatively low on the stationary forehead. If the outliers (pixels near the mouth) could be removed from the average, the quality of the extracted pulse-signal is expected to be improved significantly.

To this end, a motion robust rPPG method is proposed to treat each skin pixel in an image as an independent rPPG sensor and extract/combine multiple rPPG-signals in a way that is immune to noise. The proposed method consists of three steps: (1) creating pixel-based rPPG sensors from motion-compensated image pixels, (2) rejecting motion-induced spatial noise, and (3) optimising temporally extracted pulse-traces into a single robust rPPG-signal. To demonstrate the effectiveness, it has been evaluated on 36 challenging videos with an equal number of male and female subjects in 3 skin-type categories and 6 motion-type categories.

The contributions of this work are threefold: (1) a new strategy is proposed to track pixels in the region of interest (e.g., a subject's face) for rPPG measurement using global and local motion compensation; (2) exploiting the spatial-redundancy of an image sensor, i.e., pixel-based rPPG sensors, is proved to lead to a considerable gain in accuracy as compared to the common approach that takes a single averaged colour trace; and (3) a novel algorithm is introduced to optimise the pixel-based rPPG sensors in spatial and temporal domain.

The rest of this paper is organised as follows. Section II provides an overview of the related work. Section III analyses the problem concerning this study and describes the proposed method. The experimental setup is discussed in Section IV while the proposed method is evaluated and discussed in Section V. Finally, the conclusions are drawn in Section VI.

W. Wang and S. Stuijk are with the Electronic Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Einhoven, 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).

Manuscript received ?, 2014; revised ?, 2014.

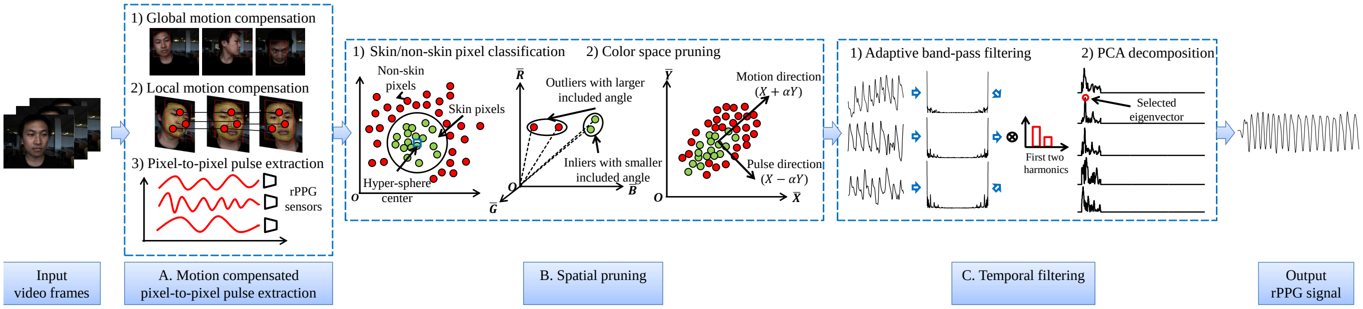

1) Global motion compensation
2) Local motion compensation
3) Pixel-to-pixel pulse extraction
rPPG sensors
1) Skin/non-skin pixel classification
2) Color space pruning
Non-skin pixels
Skin pixels
Hyper-sphere center
Outliers with larger included angle
Inliers with smaller included angle
Motion direction $(X+\alpha Y)$
Pulse direction $(X-\alpha Y)$
$\bar{R}$
$\bar{G}$
$\bar{B}$
$\bar{X}$
$\bar{Y}$
O
1) Adaptive band-pass filtering
2) PCA decomposition
Selected eigenvector
First two harmonics
Input video frames
A. Motion compensated pixel-to-pixel pulse extraction
B. Spatial pruning
C. Temporal filtering
Output rPPG signal

Fig. 1. The flowchart of the proposed motion robust rPPG framework: A. a video sequence with a manually selected RoI is the input to the framework. The global and local motion of the RoI are compensated, and pixel-based rPPG sensors between adjacent frames are constructed using motion-compensated pixel-to-pixel correspondences; B. the outliers among the pixel-based rPPG sensors, i.e., the ones without skin information or distorted by motion noise, are pruned in the spatial domain; and C. the spatially pruned inliers are chained up in the temporal domain as multiple pulse-traces, which are filtered and further optimised into a single robust rPPG-signal.

II. Related work

In the cardiovascular system, the blood pulse propagating throughout the body changes the blood volume in the vessels. Given the fact that the optical absorption of haemoglobin varies across the light spectrum, a specific cardiovascular event can be revealed by measuring the colour variations of skin reflections [1]. In 2008, Verkruysse *et al.* found that in an ambient light condition, the PPG-signal has different relative strength in three colour channels of an RGB camera that senses the human skin [5]. Based on this finding, Poh *et al.* proposed a linear combination of RGB channels defining three independent signals with Independent Component Analysis (ICA) using non-Gaussianity as the criterion for separating independent resource signals [1]. As an alternative, Lewandowska *et al.* suggested a Principal Component Analysis (PCA) based solution to define three independent linear combinations of RGB channels [2]. In 2012, MIT developed a method called “Eulerian video magnification” to amplify the subtle colour changes through band-pass filtering the temporal pyramidal image differences [6]. However, any motion-induced colour distortions within the same frequency band as that of the pulse are unfortunately amplified. More recently, de Haan *et al.* introduced the chrominance-based rPPG method (CHROM) to consider the pulse as a linear combination of three colour channels under a standardised skin-tone assumption [3]. This method demonstrates the highest accuracy of all existing rPPG methods. Based on a comparison of the state-of-art rPPG methods, this study relies on the CHROM method as the baseline to develop a motion robust rPPG method.

III. Method

The overview of the proposed motion robust rPPG framework is shown in Figure 1, which takes a video sequence containing a subject’s face as the input and returns the extracted pulse-signal as its output. There are three main steps in the processing chain: motion-compensated pixel-to-pixel pulse extraction, spatial pruning, and temporal filtering. Each step is discussed in detail in the following subsections.

A. Motion-compensated pixel-to-pixel pulse extraction

To extract parallel pulse-signals from spatial-redundant pixels, the pixels belonging to the same part of skin should be concatenated temporally. So this method compensates for the subject motion and relates temporally corresponding pixels.

1) Global motion compensation: In previous rPPG methods [1], [2], [3], the subject’s face is typically used as the Region of Interest (RoI) for pulse measurement. The motion of the face can be interpreted as a linear combination of global rigid motion (head translation and rotation) and local non-rigid motion (e.g., eye blinking and mouth talking). The common approach to compensate for the global motion of a face is using the Viola-Jones face detector to locate the face in consecutive frames with a rectangular bounding-box [7]. However, a classifier that has for example been trained with only the frontal-face samples cannot detect the side-view faces. This fundamental limitation may lead to a discontinuous face localisation across subsequent video frames.

As an alternative, a “Tracking-by-Detection” approach, which enables the online updating of the target appearance model while tracking the object, demonstrates the capability of adapting to occasional appearance changes of the target as well as handling the challenging environmental noise (e.g., partial occlusions and background clutter). According to the latest benchmark results of online object tracking presented in 2013 [8], the Circulant Structure of Tracking-by-detection with Kernels (CSK) developed by Henriques *et al.* [9] has the highest tracking speed among the top 10 accurate trackers, which can achieve hundreds of frames-per-second [8]. Considering that no significant accuracy difference can be observed among the state-of-the-art trackers in the setting of this study, the fastest CSK method is chosen to compensate for the global motion of the subject’s face instead of a Viola-Jones face detector.

2) Local motion compensation: Based on the globally tracked face, the pixels’ displacements can be more precisely estimated in this step. The implementation of the Farneback dense optical flow algorithm [10] in OpenCV 2.4 [11] is utilised to measure the translational displacement of each image pixel between adjacent frames. In addition, the idea of forward-backward flow tracking proposed by Kalal *et al.* [12] is adopted to detect the pixel-based tracking failures: in a bi-directional tracking procedure, the motion vectors with larger spatial errors yielded by abrupt motion are removed as noise, whereas the consistent motion vectors are retained to associate the temporal corresponding pixels via spatial bilinear interpolation.

3) Pixel-to-pixel pulse extraction: After global and local motion compensation, the pixels between adjacent frames have been aligned into pairs. By concatenating them in a longer frame interval, multiple pixel trajectories can be generated. However, there is a problem in creating such longer pixel trajectories: pixels belonging to the same trajectory may disappear due to occlusions (e.g., face rotation).

In fact, under a constant lighting environment, the pixels in different locations of the skin show the same *relative* PPG-amplitude. It implies that if the pulse-induced colour changes in each aligned pixel pair are temporally normalised, they can be concatenated in *an arbitrary order* to derive a long-term signal. Since the pixel-based motion vectors only need to be estimated between two frames (the smallest possible interval), it minimises the occlusion problem and also prevents the propagation of errors in local motion estimation.

The temporally normalised RGB differences of the ith pixel between frame t and $t+1$ is denoted by a vector $\overline{C}_i^{t\to t+1}$, which is defined as:

$$\overline{C}_i^{t\to t+1} = \overline{C}_i^{t+1} - \overline{C}_i^{t} = \begin{pmatrix} \overline{R}_i^{t\to t+1} \\ \overline{G}_i^{t\to t+1} \\ \overline{B}_i^{t\to t+1} \end{pmatrix}. \quad (1)$$

Assuming the spatial displacement of the ith pixel from frame t to $t+1$ is $\overrightarrow{d}(d_x, d_y)$, Eq. (1) can be written as:

$$\overline{C}_i^{t\to t+1} = \begin{pmatrix} \frac{R_i^{t+1}(x+d_x, y+d_y) - R_i^{t}(x,y)}{R_i^{t+1}(x+d_x, y+d_y) + R_i^{t}(x,y)} \\ \frac{G_i^{t+1}(x+d_x, y+d_y) - G_i^{t}(x,y)}{G_i^{t+1}(x+d_x, y+d_y) + G_i^{t}(x,y)} \\ \frac{B_i^{t+1}(x+d_x, y+d_y) - B_i^{t}(x,y)}{B_i^{t+1}(x+d_x, y+d_y) + B_i^{t}(x,y)} \end{pmatrix}. \quad (2)$$

Figure 2 shows the histogram distribution of $\overline{C}_i^{t\to t+1}$ on three different skin-tones: the Gaussian-shaped distribution of $\overline{R}_i^{t\to t+1}$, $\overline{G}_i^{t\to t+1}$ and $\overline{B}_i^{t\to t+1}$ on different skin-tones are all within the range $[-0.02, 0.02]$, which is very concentrated compared to its theoretical variation range $[-1, 1]$. Thus it can be concluded that in all skin pixels, pulse-induced colour variations roughly exhibit the same strengths in temporally normalised colour channels.

After that, the temporally normalised RGB differences are projected onto the chrominance plane using the CHROM method [3], which defines the pulse-signal as a linear combination of RGB channels as:

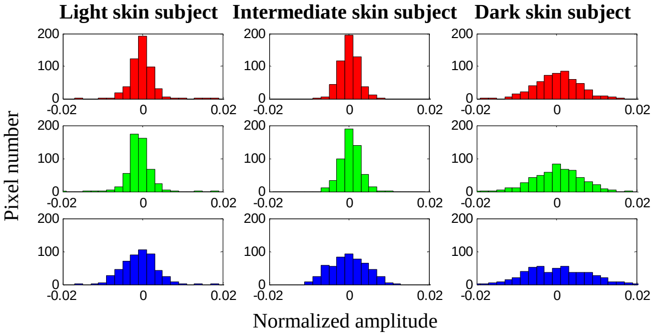


Fig. 2. The histograms of temporally normalised RGB differences between frame t and $t+1$ of three skin-types in a homogeneous lighting condition. The histogram distributions show that all skin pixels describe the similar pulse-induced RGB changes after temporal normalisation.

$$\begin{aligned} \overline{X}_i^{t\to t+1} &= 3\overline{R}_i^{t\to t+1} - 2\overline{G}_i^{t\to t+1} \\ \overline{Y}_i^{t\to t+1} &= 1.5\overline{R}_i^{t\to t+1} + \overline{G}_i^{t\to t+1} - 1.5\overline{B}_i^{t\to t+1} \end{aligned}. \quad (3)$$

By temporally concatenating $(\overline{X}_i^{t\to t+1}, \overline{Y}_i^{t\to t+1})$ estimated from pixel pairs between adjacent frames and integrating them, multiple chrominance-traces can be derived as:

$$\begin{aligned} \tilde{X}_i^{t\to t+l} &= 1 + \textstyle\sum_0^l \overline{X}_i^{t\to t+1} \\ \tilde{Y}_i^{t\to t+l} &= 1 + \textstyle\sum_0^l \overline{Y}_i^{t\to t+1} \end{aligned}, \quad (4)$$

where l is the interval length of the chrominance trace defined by a temporal sliding window. In line with [3], l is specified as 64 frames in case of a 20 FPS video recording rate. The pulse-trace in the temporal window can be calculated as:

$$\tilde{P}_i^{t\to t+l} = \tilde{X}_i^{t\to t+l} - \alpha \tilde{Y}_i^{t\to t+l}, \quad (5)$$

with

$$\alpha = \frac{\sigma(\tilde{X}_i^{t\to t+l})}{\sigma(\tilde{Y}_i^{t\to t+l})}, \quad (6)$$

where $\sigma(\cdot)$ corresponds to the standard deviation operator. In order to avoid the signal drifting/explosion in a long-term accumulation, the pulse-traces estimated from the sliding window are overlap-added together with a Hann window [3].

Note that the spatial averaging of local pixels can reduce quantisation errors during the temporal colour normalisation. The face RoI is down-sampled starting from the local motion compensation step, which not only reduces the noise sensitivity of pixel-based rPPG sensors, but also increases the processing speed of the dense optical flow. There is a trade-off in selecting the optimal down-scaling size considering the accuracy and efficiency. Since the size of all subjects' face used in this study are approximately 200×250 pixels, the RoI is uniformly down-sampled to 36×36 pixels.

B. Spatial pruning

Since the temporal noise-level in pixel-based rPPG sensors is not Gaussian distributed, the next step is to optimally select the inliers (reliable sensors) from a set of spatially redundant sensors for a robust rPPG-signal measurement. In practice, there are mainly two kinds of noise degrading the quality of rPPG sensors: (1) non-skin pixels (e.g., eyebrow, beard and nostril) that do not present pulse-signals; (2) skin pixels that contain motion-induced colour distortions. Based on this observation, a spatial pruning method including skin/non-skin pixel classification and colour space pruning is designed to pre-select the reliable sensors.

1) Skin/non-skin pixel classification: Most skin segmentation methods use pre-defined thresholds of skin colour composition or model a binary boundary between foreground and background. However, these approaches suffer from dilemmas in choosing suitable thresholds or defining foreground/background. As a matter of fact, most of the pixels inside a well-tracked face region represent the skin while only a small number of them are not skin. Since the skin pixels that share some similarities are bound in one cluster, a clustered feature-space can be constructed to detect the pixels that are further away from the cluster centre as novelties (non-skin pixels). In this method, the One Class Support Vector Machine (OC-SVM) [13] is employed to estimate such a hyper-plane, which encircles most of the pixel samples as a single class (skin class) without any prior skin colour information.

In order to train an OC-SVM, a list of feature descriptors $x_1, x_2, x_3, ..., x_n$ should be created to represent the skin pixels. Inspired by [14] that using the intensity-normalised rgb and YCrCb to discriminate skin and non-skin regions, this method represents each vector x_i with four components: $r-g$, $r-b$, $Y-Cr$ and $Y-Cb$. The OC-SVM is only trained with the first few frames to adapt to the subject skin-tone; then it is used to predict the skin pixels in the subsequent frames, i.e., the pixels with the positive and negative response for $f(x)$ are classified as skin and non-skin pixels respectively. This step significantly removes the pixel-based rPPG sensors that are not pointing at the subject's skin, and its performance is invariant to different skin-tones, as shown in Figure 3.

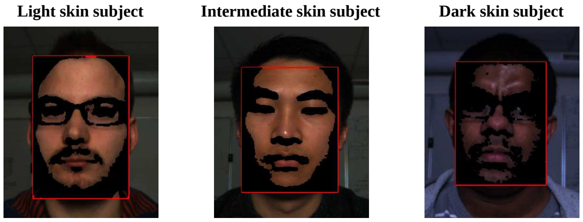


Fig. 3. An example of skin/non-skin pixel classification on three subjects with different skin colours. The red bounding-box is the tracked face, and the non-skin pixels inside the bounding-box are masked by black colour.

2) Colour space pruning: As explained before, the pulse-induced colour variations exhibit similar changes in $\overline{C}_i^{t\rightarrow t+1}$ under a homogeneous lighting environment, i.e., in temporally normalised colour space, the transformation between $(\overline{R}_i^t, \overline{G}_i^t, \overline{B}_i^t)$ and $(\overline{R}_i^{t+1}, \overline{G}_i^{t+1}, \overline{B}_i^{t+1})$ should ideally be the translation. However, motion-induced colour distortions enter this translation by adding additional residual transformations, such as rotation. Therefore by checking the geometric transformation of pixel-based rPPG sensors in the temporally normalised colour space, a number of unreliable sensors distorted during the transformation can be found and pruned. To realise this step, the inner product ϕ of the unit colour vectors between frame t and $t+1$ is simply calculated as:

$$\phi_i^{t\rightarrow t+1} = < \frac{\overline{C}_i^t}{||\overline{C}_i^t||}, \frac{\overline{C}_i^{t+1}}{||\overline{C}_i^{t+1}||} >, \tag{7}$$

where $<,>$ denotes the inner product operation; $||\cdot||$ corresponds to the L2-normalisation. When $\phi_i^{t\rightarrow t+1}$ is more deviated from 1, the angle between $\overline{C}_i^t$ and $\overline{C}_i^{t+1}$ is larger, which implies that the colour transformation is more likely to be motion-induced. In this manner, all the rPPG sensors are sorted based on their inner products and a fraction β (e.g., $\beta = \frac{1}{8}$) of them ranking closest to 0 (orthogonal) are pruned as outliers. Figure 4 shows an example of spatially pruned results in this space: subject motion yields a more sparse distribution of rPPG sensors in the spatial domain as compared to the stationary scenario.

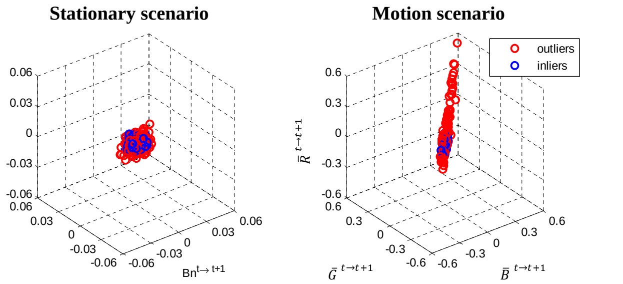


Fig. 4. An example of spatial pruning in the temporally normalised RGB space. The distribution of pixel-based rPPG sensors in this space is different between the stationary and motion scenarios. This step removes the sensors containing explicit motion-induced colour distortions.

Furthermore, the remaining rPPG sensors are pruned in the temporally normalised XY space. On the projected chrominance plane using Eq. (3), it can be observed that when the subject is perfectly stationary, $X-\alpha Y$ (pulse direction) is the principal direction while the projections are densely distributed as an ellipse; when motion appears, the direction orthogonal to $X-\alpha Y$ starts to dominate the space and the projections are sparsely distributed like a stripe, as shown in Figure 5. The direction orthogonal to the pulse direction on this chrominance plane is named as the "motion direction", which can be expressed as:

$$\overline{M}_i^{t\rightarrow t+1} = \overline{X}_i^{t\rightarrow t+1} + \alpha \overline{Y}_i^{t\rightarrow t+1}, \tag{8}$$

where α is identical to the one calculated in Eq. (6). The criterion to prune sensors on the chrominance plane is: selecting the sensors containing the least motion signals but the most likely pulse-signals. Therefore in the first round, all sensors are sorted in an ascending order based on the magnitude of their motion signal $|X+\alpha Y|$. The ones ranking at the high end are more affected by motion and are thus pruned. In the second round, the remaining sensors are sorted again in an ascending order based on their pulse-signal $X-\alpha Y$. The ones ranking in the median position represent the most probable pulse-signal and are thus selected. Similarly, this step uses the same fraction β to prune the outliers on the chrominance plane.

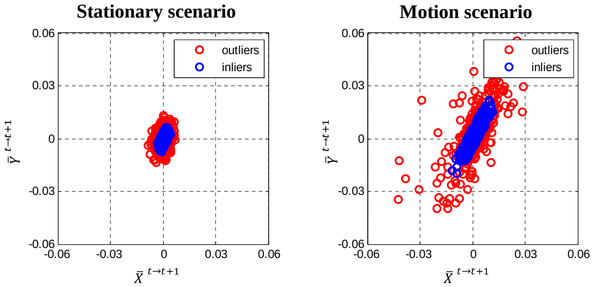


Fig. 5. An example of spatial pruning in the temporally normalised chrominance space. This step removes the sensors containing implicit motion-induced distortion residues, but retains the sensors with the most likely pulse-signal.

C. Temporal filtering

Till this step, there are two alternatives to use the spatially pruned rPPG sensors: (1) averaging the inliers for subsequent pulse estimation that is further identical to previous rPPG methods; (2) first extracting independent pulse-signals from the inliers in parallel, and then combining them into a single robust pulse-signal after post-processing. Due to the residual errors in motion estimation, the noise in spatial inliers still shows no Gaussian distribution and is not zero mean. Furthermore, concatenating the local rPPG sensors separately allows the local optimisation of the α in Eq. (5) when deriving the pulse-signals. Consequently, option (2) is adopted to separate the pulse-signal and noise by generating parallel pulse-traces.

Given the fact that the pulse derivatives in local rPPG sensors are temporally normalised, they can be randomly concatenated for creating long-term traces. But generating all possible concatenations is an impossible task (e.g., $(600!)^{64}$ different ways of concatenation in case of 600 skin pixels over 64 frames), so a simple solution is proposed to find favourable concatenations: first sort all the pulse derivatives (sensors) based on their distance to the mean, and concatenate them in the sorted order. The signal-traces ranking at the top are expected to be fairly reliable pulse-signals, whereas the ones ranking at the bottom are likely to be sub-optimal. Afterwards, the adaptive band-pass filtering and PCA decomposition steps are designed to further enhance and combine the multiple pulse-traces into a single robust rPPG-signal.

1) Adaptive band-pass filtering: Essentially, the pulse-rate of a healthy subject falls within the frequency range $[40, 240]$ beats per minute (bpm), so the parts of signal that are not in this frequency band can be safely blocked, i.e., in a temporal sliding window with 64 frames length, the in-band frequency range corresponds to $[2, 12]$. For a given moment, the instant pulse frequency should be even more concentrated in a smaller range such as $[80, 90]$ bpm. So using the real-time pulse-rate statistics, an *adaptive* band-pass filtering method is developed to better limit the band-pass filter range.

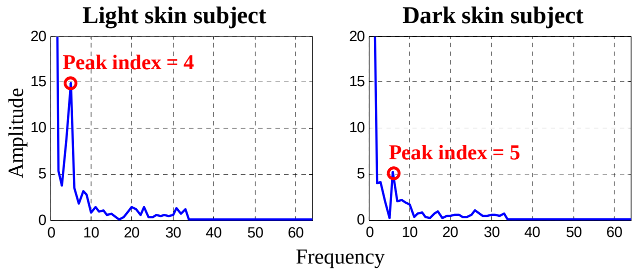


Fig. 6. An example of using the adaptive band-pass filtering on frequency spectrums obtained from two subjects (e.g., light skin and dark skin subjects) shown in Figure 3.

An example is shown in Figure 6: the mean frequency-peak position of all pulse-traces in the current temporal window is found as the most probable instantaneous pulse-frequency of the subject, then a fraction (β) of pulse-traces whose frequency-peak position has a large distance to the most probable instantaneous pulse-frequency are pruned. After that, the original pulse frequency band is adapted to the first two harmonics derived from the mean frequency-peak position, i.e., if the most probable peak position is at 4, the pulse frequency range is reduced from original $[2, 12]$ to $[3, 5] \cup [6, 10]$ (first two harmonics). Similarly, if the most probable peak position is at 5, the pulse frequency band is narrowed down to $[4, 6] \cup [8, 12]$.

Note that the proposed adaptive band-pass filtering method adjusts the pulse-frequency bandwidth based on instantaneous statistics in the current sliding window, which does not rely on any prior assumptions or previous observations (e.g., Kalman filter) of a specific subject's pulse-rate.

2) PCA decomposition: To derive a robust rPPG-signal from multiple band-passed pulse-traces, the robust pulse-signal is defined as a periodic signal with the highest variance. The reasons are: (1) the subject motions are often occasional and unintentional in a hospital/clinical use-case, i.e., non-periodic motions; (2) the motion-induced variance has been reduced by motion compensation, so the pulse-induced periodicity is more obvious in a cleaner signal-trace.

Based on this observation, the periodicity of a pulse-signal is defined as a ratio between the maximum power and total power of the signal spectrum in the pulse-frequency band. When the signal is more periodic, this ratio is larger. Similarly, the pulse-traces are sorted based on their periodicity, and a fraction β of traces with low periodicity are pruned.

Finally, PCA is performed on the periodic pulse-traces to obtain the eigenvectors, which has two benefits: (1) the decomposed eigenvectors are orthogonal to each other in the subspace, which clearly separates the pulse-signal and noises; (2) the eigenvectors are ordered in term of variance, which simplifies the procedure of selecting the most variant trace. In the temporal sliding window, the eigenvector (among the top 5 eigenvectors) that has the best correlation with the mean pulse-trace is selected to be the rPPG-signal after correcting the arbitrary sign of the eigenvector as:

$$\tilde{P}^{t\to t+l}_{selected} = \frac{<\tilde{P}^{t\to t+l}_{eigen}, \tilde{P}^{t\to t+l}_{mean}>}{|<\tilde{P}^{t\to t+l}_{eigen}, \tilde{P}^{t\to t+l}_{mean}>|} \times \tilde{P}^{t\to t+l}_{eigen}, \qquad (9)$$

where $\tilde{P}^{t\to t+l}_{eigen}$ and $\tilde{P}^{t\to t+l}_{mean}$ represent the eigenvector and mean pulse-trace respectively; $<,>$ corresponds to the inner product (correlation) between two vectors; and $|\cdot|$ denotes the absolute value operator.

IV. Experiment

This section presents the experimental setup for evaluating the proposed rPPG method. First, it shows the way of creating the benchmark video dataset. Next, it introduces two metrics for evaluating the performance of rPPG methods. Finally, it includes 5 (r)PPG methods for performance comparison.

A. Benchmark dataset

To evaluate the proposed rPPG method, 6 healthy subjects (students) are recruited from Eindhoven University of Technology. The study is approved by the Internal Committee Biomedical Experiments of Philips Research, and the informed consent is obtained from each subject. The video sequences are recorded with a global shutter RGB CCD camera (type USB UI-2230SE-C of IDS) in an uncompressed data format, at a frame rate of 20Hz, 768×576 pixels, 8 bit depth and has a duration of 90 seconds per motion-category. During the video recording, the subject wears a finger-based transmissive pulse oximetry (model CMS50E from Contec Medical) for obtaining the reference pulse-signal, which is synchronised with the recorded video frames using the USB protocol available on the device. The subjects sit in front of the camera with their

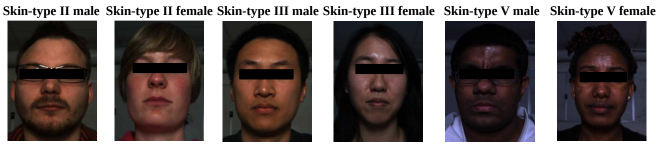


Fig. 7. A snapshot of the skin-types of six subjects in the benchmark dataset. The subjects' eyes are covered for protecting their identity only in the printing.

face visible and illuminated by a fluorescent light source (type: Philips HF3319 - EnergyLight White).

Figure 7 shows a snapshot of the recorded subjects from three skin-type categories according to the Fitzpatrick skin scale [15]: *Skin-category I* with 'Skin-type II' male/female; *Skin-category II* with 'Skin-type III' male/female; and *Skin-category III* with 'Skin-type V' male/female. All subjects are instructed to perform 6 different types of head motion: *stationary*, *translation*, *scaling*, *rotation*, *talking* and *mixed motion* (mixed motion is the mixture of all motions). For each recording, the subject remains stationary in the first 15 seconds and then performs a specific motion till the end by repeating it. There is no guidance to restrict the amount of motion, so it leads to displacements up to the maximum 35 pixels per picture-period in practice. This is intended to better mimic the practical use-cases and make the videos sufficiently challenging for rPPG. Figure 8 shows some uniformly sampled frames in the rotation video sequence of skin-category II male.

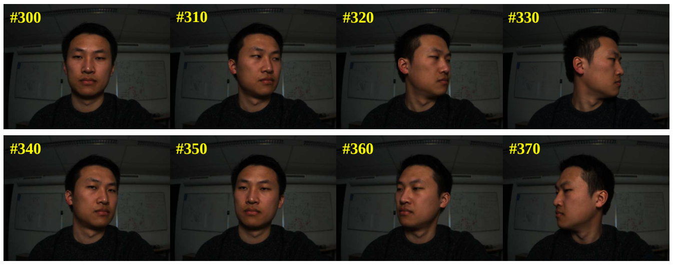


Fig. 8. An example of frames in skin-category II male rotation video. In this video, the subject performs in-plane and out-of-plane rotations.

The goal of this study is aimed to improve the "motion robustness" of rPPG, "motion" is considered the key variable that is varied in the dataset. (As mentioned before, the gender and skin-type are also varied.) So when recording each video sequence, the subject is asked to perform a specific type of motion repeatedly. Each motion is repeated approximately 15 times in each video sequence. Since motion is the most important variable affecting the rPPG performance in a single constant luminance environment, the measurement of the whole video sequence with repeated subject motion can be considered as a composition of multiple repeated short-term measurements. Hence, the video sequences allow studying the measurement repeatability. The Bland-Altman plots in Figure 10 shows for example the within-measurement repeatability comparison between rPPG and PPG, in which each scatter point represent the measurement of one complete pulse. To prevent an explosion of test data, the subjects selected for recording are representative/typical in each skin-category. There are no subjects at all intermediate skin-types, which makes it impossible for us to draw thorough conclusions on skin-tone invariance of the rPPG methods.

B. Evaluation metrics

This study adopts the same SNR metric as used in [3] to measure the signal quality for comparing the strength and weakness of rPPG methods. In this SNR metric, a temporal sliding window is utilised to segment the whole pulse-signal into intervals for deriving the SNR-trace, i.e., the temporal window has a 300 frames stride and a 1 frame sliding-step. In the sliding window, the signal interval is transformed to the frequency domain using FFT. The SNR is measured as the ratio between the energy around the first two harmonics (pulse in-band frequency) and the remaining energy (noises out-of-band frequency) of the spectrum, which is defined as:

$$SNR = 10\log_{10}\left(\frac{\sum_{f=40}^{220}(U_t(f)\tilde{S}(f))^2}{\sum_{f=40}^{220}(1-U_t(f)\tilde{S}(f))^2}\right), \tag{10}$$

where f is the pulse frequency in bpm; $\tilde{S}(f)$ is the spectrum of the pulse-signal; $U_t(f)$ is a defined binary window to pass the pulse in-band frequency and block the noisy out-of-band frequencies. Consequently, the *SNRa*, an averaged value of the SNR-trace, is used to summarise the quality of the pulse-signal.

Additionally, Bland-Altman plots are included to show the agreements of the *instantaneous pulse-rate* between the rPPG and reference PPG-sensor. The instantaneous 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. The reasons of using it for signal comparison are twofold: (1) the primitive pulse-signals obtained by rPPG and PPG have good alignment with each other, thus their instantaneous rates are comparable; (2) it captures the instantaneous changes of the pulse-signal and reflects the occasional differences between compared signals, as an example shown in Figure 10. In the standard Bland-Altman plot, the Cartesian coordinate of a pulse-rate's sample s_i is calculated as:

$$s_i(x,y) = \left(\frac{PR_i + RR_i}{2}, PR_i - RR_i\right). \tag{11}$$

where PR_i and RR_i are ith instantaneous pulse-rates obtained by rPPG and PPG respectively. And RR_i is smoothed by a 5-point mean filter for suppressing the noise effect. Furthermore, the Bland-Altman agreement A between PR_i and RR_i is calculated as:

$$A = \frac{\sum_{i=1}^{n} a_i}{n}, \tag{12}$$

with

$$a_i = \begin{cases} 1 & \text{if } |PR_i - RR_i| < 1.96\sigma \\ 0 & \text{if } |PR_i - RR_i| \geq 1.96\sigma \end{cases}, \tag{13}$$

where n is the total number of samples in a pulse-rate; σ denotes the standard deviation of the difference between PR_i and RR_i.

Finally, the Analysis of Variance (ANOVA) is applied on SNRa values to analyse the significance of difference between (r)PPG methods under certain categories (e.g., skin or motion), i.e., to show whether the main variation in SNRa is "between" groups (rPPG methods) or "within" groups (video sequences). Based on the results of ANOVA, the post-hoc comparison is used to further evaluate the posteriori pairwise comparisons between individual methods to see which one is significantly better than the other. The ANOVA with post-hoc comparison gives a clear overview of statistical comparison between investigated (r)PPG methods.

C. Compared methods

Based on the benchmark dataset and evaluation metrics, three comparisons have been performed for the evaluation: (1) comparing the proposed method to the state-of-the-art rPPG method CHROM [3]; (2) comparing the separate steps in the developed framework to show their independent improvements and contributions to the complete solution, since these separate steps involve innovations that are not addressed in previous rPPG studies; and (3) comparing the rPPG methods to the PPG method to show the disparity between camera-based and contact-based approaches. The details of the compared (r)PPG methods are described below:

- **Face-Detect-Mean** (FDM) is a re-implementation of the CHROM method. It uses the Viola-Jones face detector to locate the face, and applies the OC-SVM method to select the skin-pixels to derive the averaged RGB traces for pulse-signal estimation.
- **Face-Track-Mean** (FTM) is the included sub-step of the proposed method. It replaces the Viola-Jones face detector in FDM with the CSK tracker for the better face localisation.
- **Pixel-Track-Mean** (PTM) is the included sub-step of the proposed method. It extends FTM with spatial redundancy by creating pixel-based rPPG sensors, but takes the averaged values of the temporally normalised colour differences to derive the pulse-signal.
- **Pixel-Track-Complete** (PTC) is the complete version of the proposed method, which adds the spatio-temporal optimisation procedure (spatial pruning and temporal filtering) to the PTM.
- **Contact-Based-Sensor** (CBS) is a finger-based pulse oximetry. It is used to record the reference pulse-signal for comparison.

V. Results and discussion

The proposed method is implemented in Java using the OpenCV 2.4 libaray [11] and ran on a laptop with an Intel Core i7 2.70 GHZ processor and 8 GB RAM. All 5 methods are evaluated on 36 video sequences from the benchmark dataset. For fair comparison, only the RoI (e.g., subject's face) needs to be manually initialised while the other parameters remained identical when processing different videos.

The results show that the gender is not the key factor which needs to be investigated in this dataset, i.e., the differences between stationary male and female from the same skin-category are rather small. Thus the results obtained by the different genders in the same skin-category and motion-type are averaged. Table I and Table II summarise the gender-averaged SNRa and Bland-Altman agreements respectively. Moreover, the SNRa values in Table I are further averaged over (1) the three skin-categories for comparing the motion robustness; (2) the six motion-types for comparing the skin-tone invariance, as shown in Figure 9 (the standard deviation of SNRa is also calculated to show the methods' variability in each category).

1) Stationary scenario: Figure 9a shows that all (r)PPG methods gain similar performance on stationary subjects, i.e., the standard deviations of their SNRa are below 1.0dB. The reason is that these methods are all using the chrominance-based method [3] for pulse extraction. Their main difference is in motion estimation and outlier rejection. No significant improvements can be expected for static subjects.

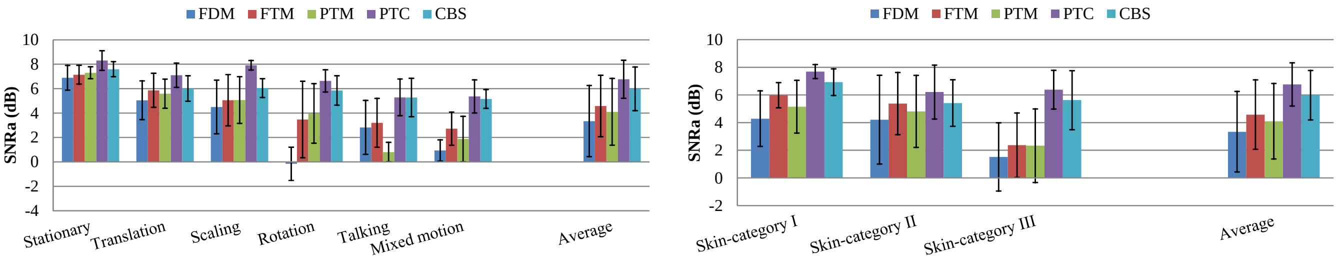
2) Motion scenarios: In videos where the subjects' frontal face can be detected by the Viola-Jones method (e.g., translation, scaling and talking), FDM still works properly, whereas FTM that relies on the online object tracker is approximately

TABLE I
SNRa results gained by (r)PPG methods on benchmark videos (averaged over genders). Bold entries indicate the best performance of rPPG methods in each category.

Videos	FDM	FTM	PTM	PTC	CBS
Skin-category I stationary	6.54	6.65	6.73	**7.18**	6.80
Skin-category I translation	6.20	6.75	6.33	**8.40**	7.16
Skin-category I scaling	3.90	5.48	5.44	**8.26**	7.14
Skin-category I rotation	1.53	6.83	6.78	**7.91**	8.72
Skin-category I talking	5.69	5.94	1.34	**7.25**	5.81
Skin-category I mixed motion	1.86	4.24	4.30	**7.18**	5.92
Skin-category II stationary	8.26	8.24	7.93	**8.80**	7.68
Skin-category II translation	6.13	**6.95**	6.52	6.91	4.64
Skin-category II scaling	7.43	7.39	7.20	**8.11**	5.48
Skin-category II rotation	-0.20	4.29	4.30	**5.90**	7.46
Skin-category II talking	2.49	2.42	1.39	**3.60**	3.13
Skin-category II mixed motion	1.18	2.97	1.53	**3.97**	4.09
Skin-category III stationary	5.87	6.55	7.24	**8.93**	8.30
Skin-category III translation	2.81	3.89	3.90	**5.97**	6.24
Skin-category III scaling	2.16	2.29	2.55	**7.37**	5.52
Skin-category III rotation	-1.80	-0.70	0.83	**6.09**	1.38
Skin-category III talking	0.30	1.24	-0.32	**5.00**	6.88
Skin-category III mixed motion	-0.24	0.94	-0.21	**4.93**	5.44
Average	3.34	4.58	4.10	**6.76**	5.99

TABLE II
Agreements gained by rPPG methods on benchmark videos (averaged over genders). Bold entries indicate the best performance of rPPG methods in each category.

Videos	FDM	FTM	PTM	PTC
Skin-category I stationary	96%	95%	96%	**96%**
Skin-category I translation	80%	77%	86%	**97%**
Skin-category I scaling	63%	79%	80%	**97%**
Skin-category I rotation	41%	75%	71%	**95%**
Skin-category I talking	66%	70%	45%	**93%**
Skin-category I mixed motion	36%	62%	57%	**88%**
Skin-category II stationary	98%	98%	97%	**99%**
Skin-category II translation	65%	62%	58%	**76%**
Skin-category II scaling	83%	**84%**	80%	83%
Skin-category II rotation	27%	57%	54%	**84%**
Skin-category II talking	57%	58%	47%	**65%**
Skin-category II mixed motion	48%	67%	57%	**78%**
Skin-category III stationary	74%	79%	84%	**85%**
Skin-category III translation	45%	52%	52%	**75%**
Skin-category III scaling	31%	53%	50%	**68%**
Skin-category III rotation	19%	25%	33%	**43%**
Skin-category III talking	40%	49%	34%	**65%**
Skin-category III mixed motion	24%	32%	28%	**49%**
Average	55%	65%	62%	**80%**



(a) Motion SNRa

(b) Skin SNRa

Fig. 9. In each category, the colour bar is the averaged SNRa while the black bar is the standard deviation. (a) Motion SNRa: it compares the SNRa obtained by the (r)PPG methods in different motion-types (averaged over genders and skin-categories). (b) Skin SNRa: it compares the SNRa obtained by the (r)PPG methods in different skin-categories (averaged over genders and motion-types).

1.0dB better. The improvement is due to the object tracker, which leads to a smoother face localisation between consecutive frames compared to the face detector by exploiting the target's appearance consistency and position coherence.

However, the comparison between FTM and PTM implies that only exploiting the spatial-redundancy cannot consistently improve the signal quality, i.e., in talking videos that containing local non-rigid mouth/lips motions, PTM increases the noise sensitivity in local pixel-based rPPG sensors and thus exhibits more quantisation errors (even 2.4dB less than FTM). This problem is solved in PTC that incorporates an outliers pruning procedure to remove the motion-distorted sensors.

In videos with vigorous motions (e.g., rotation and mixed motion), PTC including its substeps (FTM and PTM) show superior performance against FDM in Figure 9a. The failure of FDM in these two types of motion (-0.15dB and 0.93dB respectively) is mainly caused by the face detector, which cannot locate the side-view faces in some frames. Another significant challenge is from the large motion-induced colour distortions on the skin surface, i.e., both the magnitude and orientation of skin-reflected light are dramatically changed during the rotation. In such a case, PTC achieves the largest improvement over FDM compared to other motion-types (6.79dB and 4.43dB more respectively), which indicates that the proposed method can better deal with the subject motions in challenging use-cases. Comparing the subject variability (standard deviation) between the videos with and without motion, FDM, FTM and PTM increase around ± 2.0dB while PTC increases around ± 0.7dB, which is fairly stable.

Figure 10 shows the instantaneous pulse-rate and Bland-Altman plots of 6 motion-types in Skin-category II male. In videos with regular motions (e.g., stationary, translation, scaling and talking), all rPPG methods are able to precisely capture the instantaneous abrupt changes of pulse-rate and have good alignments with corresponding reference-signal. In videos with vigorous motions (e.g., rotation and mixed motion), PTC particularly outperforms other rPPG methods, i.e., the agreement of PTC achieves 98% and 89% respectively.

3) Different skin-categories: In addition to the motion robustness comparison, the skin-tone invariance of rPPG methods is analysed. Figure 9b shows that FDM, FTM and PTM have difficulties in dealing with the darker skin-type (Skin-category III) as compared to the brighter skin-types ((Skin-category I and II) (around 3dB less). The performance degradation is caused by using the skin-chromaticity based method for pulse extraction: the higher melanin contents in darker skin absorbs part of the diffuse light reflections that carry the pulse-signal, whereas the specular reflection is not reduced [3]. In contrast, PTC obtains a relatively consistent performance across the different skin-categories, since the skin pixels with specular reflections caused by either the subject motion or skin absorption are all pruned as outliers. Besides, its temporal filtering suppresses the out-of-band frequency noise and strengthens the pulse-frequency. Figure 12 shows

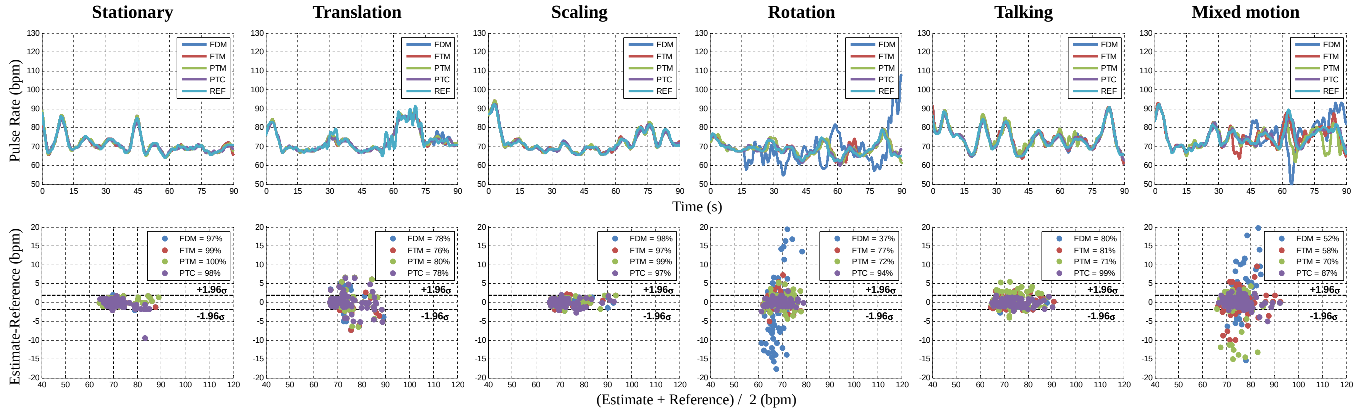


Fig. 10. The instantaneous pulse-rate plot (first row) and Bland-Altman plot (second row) for six motion-types of the male subject in skin-category II. The subject's appearance is shown in Figure 8. The Bland-Altman agreements are calculated between rPPG-signals and reference-signals (REF), where the reference-signals are the smoothed signals recorded by CBS. To visually compare the agreements between rPPG methods and reference, the Bland-Altman plots of four rPPG methods are put in one graph and use the σ of $\pm 1.96\sigma$ obtained between PTC and the reference to denote the variance range.

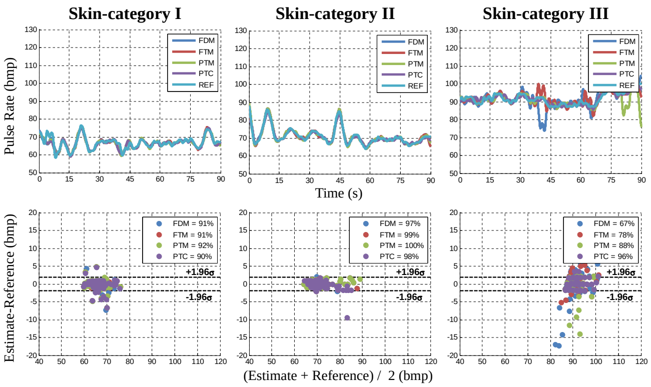


Fig. 12. An example of instantaneous pulse-rate plot (first row) and Bland-Altman plot (second row) for stationary male subjects in three skin-categories. The facial appearance of three male subjects are shown in Figure 3.

the instantaneous pulse-rate and the Bland-Altman agreement of the stationary male subjects in three skin-categories. It is apparent that only PTC shows consistently high agreements with the reference-signal.

4) ANOVA with post-hoc comparison: To analyse the significance of differences in motion and skin-tone robustness between methods, the SNRa values in Table I are grouped into five categories: the skin-categories (I, II and III), the stationary-category and the overall-category. In each of the skin categories, the significance of differences between methods on motion robustness is measured (results on moving videos). In the stationary category, the significance of differences between methods on skin-tone robustness is investigated. Finally in the overall category, the overall significance of difference between methods is shown using the entire dataset. This paper applied the balanced one-way ANOVA on these five categories, and post-hoc comparison using Tukeys honestly significant difference criterion. In each category, a common significance threshold (p-value < 0.05) is used. Figure 11 shows the results, while Table III lists the main ANOVA statistics.

TABLE III

THE STATISTICS OBTAINED BY ANOVA IN FIVE CATEGORIES. BOLD ENTRIES INDICATE THE CATEGORY WITH p-VALUE LARGER THAN 0.05.

Categories	*MS*-within	*MS*-between	*F*-ratio	*p*-value
Skin-category I	2.42	12.62	5.2	**0.0049**
Skin-category II	5.89	3.71	0.63	0.6466
Skin-category III	3.07	28.74	9.36	**0.0002**
Stationary	0.86	0.88	1.03	0.4398
Overall	5.90	35.23	5.97	**0.0003**

In skin-categories I and III, the compared methods have significant differences (both are < 0.05). In skin-category II, the differences are not significant (p-value $= 0.6466$). This high p-value reflects a limited variation between groups (3.71) as compared to that within groups (5.89). Indeed the subjects in this group caused rather large motion variations as compared to subjects in the other groups. This could happen as limited instructions for the precise movements to be made were given to the subjects. In Figure 11, the ANOVA plots show that PTC achieves the best performance in all three skin-categories with respect to the subject motion. The post-hoc plots show that PTC is the only method that is significantly different from the baseline method (FDM) for skin-categories I and III. CBS, the contact-based reference method, only has significant difference with FDM in skin-category I. FTM and PTM have no significant pairwise differences with FDM in any skin-category, i.e., their possible motion-robustness improvement is very limited.

In the stationary-category, the p-value is 0.4398 (> 0.05) and thus the differences between methods are not significant in terms of the skin-tone robustness. Figure 11 shows that on average PTC does score best.

Also in the overall-category, the differences between methods in the complete benchmark dataset are significant 0.0003 (< 0.05). Figure 11 shows that PTC again yields the largest improvement over the baseline method (FDM) and has a performance that is similar to the contact-based method (CBS), i.e., PTC and CBS have significant pairwise differences with FDM in the post-hoc comparison.

It can be concluded that the proposed method, PTC, leads to significantly improved motion robustness, while for stationary

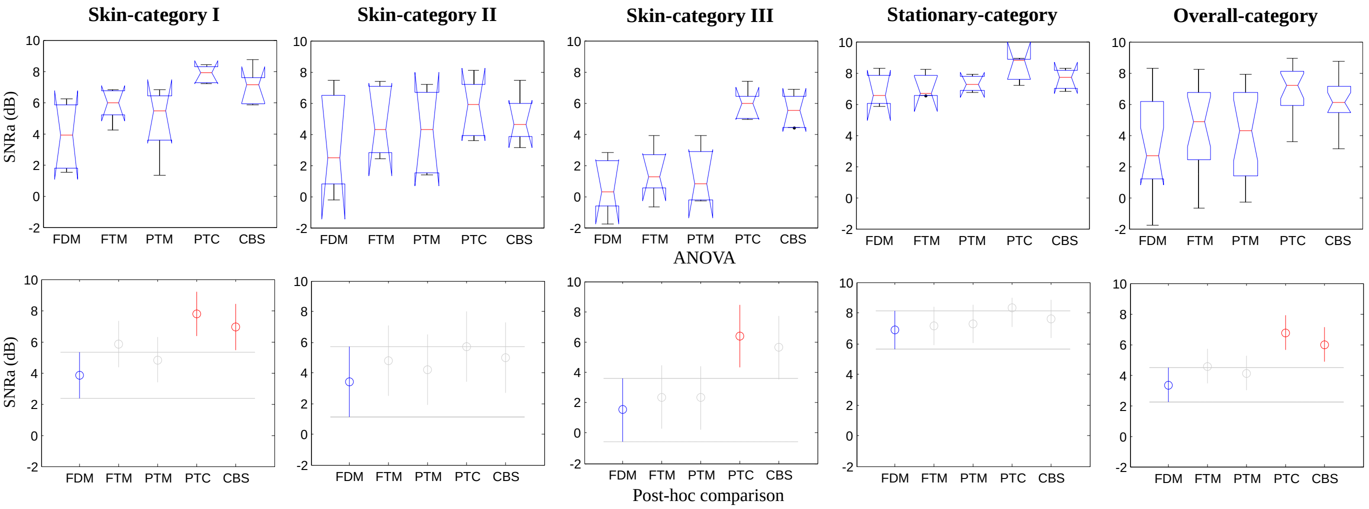


Fig. 11. The statistical comparison between five (r)PPG methods in five categories using ANOVA with post-hoc analysis. The ANOVA plots in the first row show the overview of performance variation between methods in each category, i.e., median (red bar), standard deviation (blue box), minimum and maximum (black bar) SNRa values. The post-hoc plots in the second row show the pairwise differences between the methods in each category and highlight the pairs that are significantly different (in blue and red).

videos the skin-tone robustness on average is the best though the differences with other methods are not significant.

VI. Conclusion

This study introduces a motion robust rPPG method that enables the remote detection of a pulse-signal from subjects using an RGB camera. This work integrates the latest methods in motion estimation and pulse extraction, and proposes novel algorithms to create and optimise pixel-based rPPG sensors in the spatial and temporal domain for robust pulse measurement. Experimental results on 36 challenging benchmark video sequences show that the proposed method significantly improves the SNR of the state-of-the-art rPPG method from 3.34dB (± 2.91) to 6.76dB (± 1.56), and improves the Bland-Altman agreement ($\pm 1.96\sigma$) with instantaneous reference pulse-rate from 55% to 80% correct, i.e., a performance that is very close to the contact-based sensor. ANOVA with post-hoc comparison shows that the proposed method, PTC, leads to significantly improved motion robustness, while on stationary videos with skin-tone variance it is also the best on average though the difference with the baseline method is not significant.

Acknowledgement

The authors would like to thank Ihor Kirenko, Erik Bresch, Joyce Westerink, Bert den Brinker, Wim Verkruijsse and Vincent Jeanne at Philips Research for their support. Also, we are grateful for the help of volunteers from Eindhoven University of Technology in creating the benchmark video dataset.

References

[1] M.-Z. Poh, D. McDuff, and R. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Trans. on*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[2] M. Lewandowska, J. Ruminski, T. Kocejko, 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.

[3] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *Biomedical Engineering, IEEE Trans. on*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[4] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 3430–3437.

[5] W. Verkruysse, 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.

[6] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 65:1–65:8, July 2012.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition (CVPR), 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, Dec. 2001, pp. I–511–I–518.

[8] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 2411–2418.

[9] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *European Conference on Computer Vision (ECCV), 2012 on*, ser. Lecture Notes in Computer Science. Springer, Oct. 2012, vol. 7575, pp. 702–715.

[10] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, ser. Lecture Notes in Computer Science. Springer, 2003, vol. 2749, pp. 363–370.

[11] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2000.

[12] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Aug. 2010, pp. 2756–2759.

[13] Y. Chen, X. S. Zhou, and T. Huang, "One-class SVM for learning in image retrieval," in *Image Processing, 2001. Proceedings of the 2001 International Conference on*, vol. 1, Oct. 2001, pp. 34–37.

[14] N. A. Ibraheem, R. Z. Khan, and M. M. Hasan, "Comparative study of skin color based segmentation techniques," *International Journal of Applied Information Systems*, vol. 5, no. 10, pp. 24–34, Aug. 2013.

[15] T. Fitzpatrick, "The validity and practicality of sun-reactive skin types i through vi," *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871, 1988.

Wenjin Wang received his BSc from Northeastern University, China in 2011 and his MSc from University of Amsterdam, Netherlands in 2013. Currently, he is a PhD candidate at Eindhoven University of Technology and cooperates with the Vital Signs Camera project at Philips Research Eindhoven.

Wenjin Wang works on computer vision and related problems.

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, 2 book chapters, 170 scientific papers and more than 130 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.