Enhancing Remote-PPG Pulse Extraction in Disturbance Scenarios Utilizing Spectral Characteristics

Kai Zhou, Simon Krause, Timon Blöcher
FZI Forschungszentrum Informatik
Karlsruhe, Germany
{zhou, skrause, bloecher}@fzi.de

Wilhelm Stork
Institute of Information Processing Technologies
Karlsruhe Institute of Technology KIT
Karlsruhe, Germany
wilhelm.stork@kit.edu

Abstract

In recent years, several approaches for remote Photoplethysmography (rPPG) have been proposed, and the recently proposed methods have achieved substantial improvement in measurement accuracy. However, none of the methods has investigated the possibility of using the spectral characteristics for the design of rPPG signal extraction algorithms. In this paper, we propose a new rPPG measurement method which exploits the spectral characteristics of rPPG signals. We validated the freshly proposed method on a benchmark dataset including seven scenarios and 26 participants. The results of the validation experiment demonstrates the feasibility to use spectral characteristics to extract rPPG signal. By combining with the constraint plane, the new proposed method provides better overall performance.

1. Introduction

Remote Photoplethysmography (rPPG) as a means for contactless vital sign measurement has become more and more popular in recent years. It enables an analysis of microvascular blood volume variation in upper skin tissues using a simple vision system.

rPPG has found its application in many fields, such as fitness monitoring [1], stress monitoring [2], driver state monitoring [3] [4], neonatal care [5], and product response analysis [6]. rPPG systems don't require a complicated placement process or direct contact to skin like ECG or common PPG. Dassel *et al*. [7] has shown that pressure between skin tissue and photo sensors of traditional contact PPG can affect waveform of PPG signals. Moreover, contact PPG sensors tend to cause skin irritation and thereby reduce comfort for long term use.

The microvascular blood volume in skin tissue is modulated by the cardiac activity, which leads to variations in the hemoglobin amount thereby causing fluctuation in the skin color. The object of an rPPG system is to recover the blood volume pulse signal (BVP) from videos and extract vital parameters (*e.g*., pulse rate). Generally, extraction of pulse signal using rPPG is comprised of three common steps: ROI detection, BVP signal extraction and calculation of vital parameters. ROI detection is completed by detecting the face region in each video frame. Raw signals are calculated as pixel intensity in the ROI. Then, a core rPPG algorithm is adopted to recover the blood volume pulse (BVP) from raw signals. Vital parameters like heart rate can be retrieved either in the time domain by calculating the inter-beats-interval [8] [9] or searching the maximum peak in the FFT spectrum [10]. This general process is illustrated in Figure 1.

An early approach for the detection of pulse rates in video recordings of human faces is given by Verkruysse *et al*. [11]. The blood volume pulse signal was revealed using only ambient light and a low-cost camera. His work demonstrated that the signal extracted from the green channel has the highest pulsatility. Besides pulsatile and AC-components related to the intrinsic property of the skin color and the illumination condition, the signal trace extracted from a signal color channel also contains disturbance components due to the quantization noise of the camera as well as abrupt variations in illumination and motion.

Blind Source Separation (BSS) methods accomplish pulse retrieval by de-mixing raw signals into different sources. Tsouri and Li [12] performed the principal component analysis (PCA) on the raw traces and determined the pulse signal as the component with the maximal variance.

Methods using independent component analysis (ICA) to solve the rPPG problem have been investigated as well [9]. Unlike PCA studying the variance of the signal, ICA separates the pulsatile signal from noise by minimizing the Gaussianity within the de-mixed signal. In [8], McDuff *et al*. extracted the blood volume pulse signal using ICA on

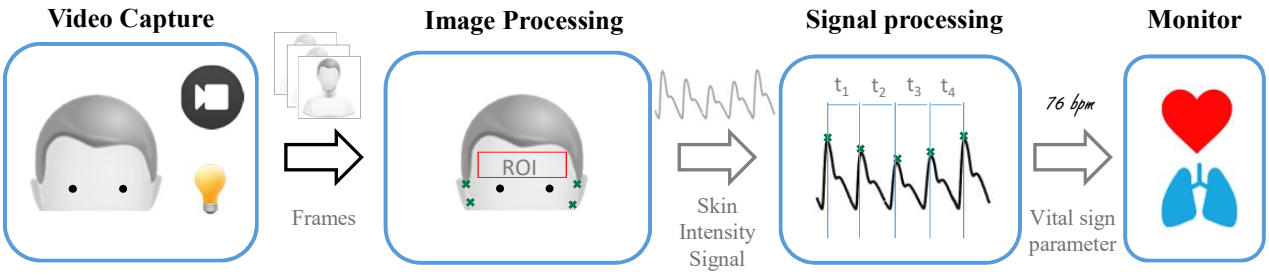

Figure 1. A computer vision algorithm registers ROI by detecting the face region in the video frame sequence and outputs raw color signals. A core kernel rPPG algorithm is adopted to recover the BVP signals from the raw signals, and finally, vital parameters are retrieved from the extracted BVP signals.

color traces captured by a five-band camera (RGBCO). One of the ambiguities deteriorating the quality of signals extracted by ICA is the indeterminacy of the order of separated signals due to the linearity of the de-mixing transform. Poh *et al*. [9] chose the signal within the second component after applying ICA and Mcduff *et al*. [8] selected the signal with the highest peak in the frequency domain. Besides Gaussianity of the source signal, Macwan *et al*. [10] also considered the periodicity of the BVP signal. They incorporated it as an extra term into the objective function for Gaussianity minimization, achieving performance boost upon the conventional ICA. In his method, autocorrelation was used as the measure for the periodicity.

PCA and ICA are data-driven solutions which exploit only statistical characteristics of the blood volume pulse signal. Methods exploiting physical prior knowledge of rPPG were also proposed and brought higher measurement accuracy and robustness above the BSS-methods. Haan and Jeanne [13] proposed a chrominance-based method (CHROM) to extract the blood volume pulse signal by assuming a standardized skin-color to white-balance the images. Wang *et al*. [14] proposed an approach (POS) which extracts BPV signals by projecting color traces onto a plane where intensity components cancel out.

In this work, we investigate the spectral characteristics of rPPG and propose a method to incorporate these characteristics into the rPPG signal extraction procedure. Section 2 introduces the problem definition and theory of the proposed approach. In Section 3 the proposed method will be mathematically described. The validation experiment is conducted in Section 4. In the final Section 5, we draw conclusions.

2. Theory

The raw signal of each color channel is extracted by spatially averaging pixel values inside the ROI in each individual video frame then temporally concatenating them into a signal trace. Stacking the signal trace of each color channel gives a raw signal matrix $\mathbf{C}$ with a size of $l \times 3$ for an RGB camera, where l stands for the length of the time window.

CHROM and POS extract pulse signals from the zero-mean skin color signals $\tilde{\mathbf{C}}(t)$, which are obtained by temporally normalizing $\mathbf{C}$ and subtracting DC components. $\tilde{\mathbf{C}}(t)$ are first projected onto a plane in the color space defined based on the prior knowledge of the physical model and skin characteristics.[1] The first step of projection compensates some disturbance energy and gives two signal traces $\mathbf{S}_1$ and $\mathbf{S}_2$. The yielded signals $\mathbf{S}_1$ and $\mathbf{S}_2$ have opposite relative phase relationships with respect to the pulsatile signal and the remaining disturbance. For example in POS, the pulsatile components in $\mathbf{S}_1$ and $\mathbf{S}_2$ are in accordant phase with each other and the remaining disturbance components are in the counter-phase.

This opposite phase-alignment property in disturbance and target pulsatile components makes it possible to separate the pulsatile signal by deploying the so-called alpha-tuning [13]:

$$\mathbf{h} = \mathbf{S}_1 + a\frac{\sigma(\mathbf{S}_1)}{\sigma(\mathbf{S}_2)} \cdot \mathbf{S}_2 \tag{1}$$

where $\sigma(\cdot)$ stands for the standard deviation of the signal. a is -1 for CHROM and 1 for POS. Alpha-tuning cancels out the disturbance component in $\mathbf{S}_1$ and $\mathbf{S}_2$ and retains the pulsatile component.

However, in some cases in presence of disturbance, the opposite phase-alignment property does not always hold due to the strong non-linearity of the light-skin interaction. Thus it can happen in POS that the disturbance components in $\mathbf{S}_1$ and $\mathbf{S}_2$ are actually in phase. In this situation, applying alpha-tuning with a positive value of a can only enhance the disturbance strength instead and fails to extract the signal correctly. For rPPG signal extraction in this case, we propose to define a projection vector based on spectral

[1]The projection matrix used by CHROM is [3, -2, 0; 1.5, 1, -1.5], and [0, 1, 1; -2, 1, 1] by POS.

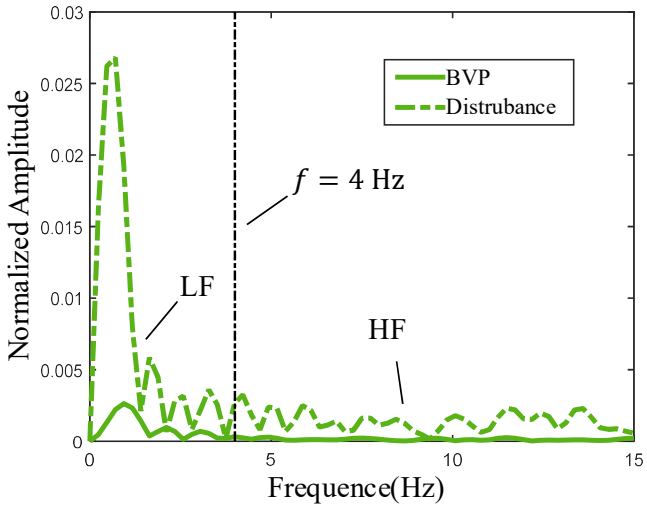

Figure 2. Comparison between spectra of signals repectively dominated by disturbance and pulsatile.

characteristics of signals.

We first investigate raw signals in two different scenarios. Figure 2 plots spectra of two signal snippets of the green channel. The signals snippets were cropped with a time window of 1.2 s and the frame rate of videos was 30 fps. The dash-dotted line represents a signal which was heavily polluted by disturbance and the solid one was extracted from a stationary scenario.

From the figure it can be observed that the polluted signal has a much higher amplitude than the signal extracted from the stationary scenario. For the scenario in presence of disturbance, we can consider the signal power in the high frequency range almost induced by the disturbance. If the disturbance has the same waveform in each color channel, the projection vector which cancels out the disturbance in high frequency (HF) range should be able to suppress disturbance in low frequency (LF) range as well, which means that the projection vector is orthogonal to the disturbance component. Assuming that the pulsatile component and the disturbance are physically separable in the normalized color space, there should be a big enough angle between the pulsatile component and the disturbance, therefore the pulsatile component and the yielded projection vector are not orthogonal to each other. Projecting the color traces onto the disturbance minimizing vector will give a pulsatile signal with an observable amplitude. Thus, for disturbance scenarios, we determine the projection direction as the vector in which the high frequency energy is minimized.

For the stationary scenario, the signal essentially consists of the target blood volume pulse. The signal energy is mainly distributed in the low frequency range. The projection direction in this case should be able to enhance low frequency components and simultaneously suppress high frequency components. The optimal projection vector can be computed by minimizing the ratio of HF components to LF components.

Since the human relative pulsatile amplitude varies in a specific range and the relative amplitude (AC/DC) is much larger [15], we can easily distinguish the stationary and disturbance scenarios by setting a threshold on the relative amplitude.

3. Method

In this section, we will describe our method in more details. The purpose of the method is to find an optimal projection vector $\mathbf{v}^* \in \mathbb{R}^{3\times 1}$ so that the pulsatile signal can be extract as:

$$\mathbf{h}\left(t\right) = \tilde{\mathbf{C}}\left(t\right)\mathbf{v}^* \tag{2}$$

As different objectives are adopted for stationary and disturbance scenarios, methods for these two scenarios will be introduced separately. After that we propose to improve performance of our method by putting a constraint on the projection vector.

The optimization process is performed in the frequency domain. Performing Fast Fourier Transform on the temporally normalized color traces $\tilde{\mathbf{C}}\left(t\right)$ gives a complex matrix $\mathbf{F} \in \mathbb{C}^{f\times 3}$, where f denotes the number of frequency bins. We decompose $\mathbf{F}$ into two sub-matrices by setting a frequency threshold f_c:

$$\mathbf{F}^T = [\mathbf{F}_l^T, \mathbf{F}_h^T], \tag{3}$$

where $\mathbf{F}_h$ denotes all components with frequency of $(0, f_c] \cup [f_s - f_c, f_s)$ and $\mathbf{F}_l$ denotes components in the frequency range $(f_c, f_s - f_c)$. f_s is the frame rate of recordings. The setting of f_c should ensure that the HF energy is mainly contributed by disturbance if disturbance occurs. We choose f_c as 4 Hz which corresponds to the upper limit of human heart rates.

3.1. Scenario with disturbance

In scenarios of disturbance, HF components in color traces are almost induced by disturbance. To suppress the disturbance energy, we search a unit vector on which the energy of HF components is minimized. Mathematically, it can expressed as:

$$\begin{aligned} \mathbf{v}^* &= \underset{\mathbf{v}}{\operatorname{argmin}}\, \mathbf{v}^T\mathbf{F}_h^H\mathbf{F}_h\mathbf{v} \\ &\triangleq \underset{\mathbf{v}}{\operatorname{argmin}}\, \mathbf{v}^T\mathbf{M}_h\mathbf{v}, \\ s.t. \quad &\|\mathbf{v}\|_2 = 1 \end{aligned} \tag{4}$$

where $\mathbf{M}_h \triangleq \mathbf{F}_h^H\mathbf{F}_h$. $\mathbf{M}_h$ is a real matrix due to symmetry of $\mathbf{F}_h$ in the frequency domain. $\mathbf{F}_h^H$ is the Hermitian transpose of $\mathbf{F}_h$. The projection vector $\mathbf{v}^*$ can be determined

as the eigenvector of $\mathbf{M}_h$ corresponding to the least eigenvalue.

3.2. Stationary scenario

In stationary scenarios, the optimal projection vector $\mathbf{v}^*$ should be able to enhance the energy in the low frequency domain and reduce the high frequency noise. Determination of $\mathbf{v}^*$ can be expressed as:

$$\begin{aligned} \mathbf{v}^* &= \underset{\mathbf{v}}{\operatorname{argmin}} \frac{\mathbf{v}^T \mathbf{F}_h^H \mathbf{F}_h \mathbf{v}}{\mathbf{v}^T \mathbf{F}_l^H \mathbf{F}_l \mathbf{v}} \\ &\triangleq \underset{\mathbf{v}}{\operatorname{argmin}} \frac{\mathbf{v}^T \mathbf{M}_h \mathbf{v}}{\mathbf{v}^T \mathbf{M}_l \mathbf{v}} \\ s.t. \quad & \|\mathbf{v}\|_2 = 1 \end{aligned} \tag{5}$$

with $\mathbf{M}_l \triangleq \mathbf{F}_l^H \mathbf{F}_l$, which is in form of a generalized Rayleigh quotient. Decomposing $\mathbf{M}_l$ as:

$$\mathbf{M}_l = \mathbf{U}_l \mathbf{\Lambda} \mathbf{U}_l^T \tag{6}$$

where columns of $\mathbf{U}_l$ are the right eigenvectors of $\mathbf{M}_l$ and $\mathbf{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues. The denominator of (5) can be written as:

$$\begin{aligned} \mathbf{v}^T \mathbf{M}_l \mathbf{v} =& \mathbf{v}^T \mathbf{U}_l \mathbf{\Lambda} \mathbf{U}_l^T \mathbf{v} \\ =& \mathbf{v}^T \mathbf{U}_l \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{U}_l^T \mathbf{v} \\ \triangleq& \boldsymbol{\alpha}^T \boldsymbol{\alpha} \end{aligned} \tag{7}$$

with $\boldsymbol{\alpha} \triangleq \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{U}_l^T \mathbf{v}$. Combining this result with equation (5) and replacing $\mathbf{v}$ we have:

$$\begin{aligned} \boldsymbol{\alpha}^* &= \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{\boldsymbol{\alpha}^T \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}_l^T \mathbf{M}_h \mathbf{U}_l \mathbf{\Lambda}^{-\frac{1}{2}} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{\alpha}} \\ &\triangleq \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{\boldsymbol{\alpha}^T \mathbf{S} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{\alpha}} \end{aligned} \tag{8}$$

with $\mathbf{S} \triangleq \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}_l^T \mathbf{M}_h \mathbf{U}_l \mathbf{\Lambda}^{-\frac{1}{2}}$. $\boldsymbol{\alpha}^*$ can be determined as the eigenvector of $\mathbf{S}$ corresponding to the least eigenvalue. The optimal projection vector can be calculated using:

$$\mathbf{v}^* = \frac{\mathbf{U}_l \mathbf{\Lambda}^{-\frac{1}{2}} \boldsymbol{\alpha}^*}{\|\mathbf{U}_l \mathbf{\Lambda}^{-\frac{1}{2}} \boldsymbol{\alpha}^*\|_2} \tag{9}$$

Then we calculate the blood pulse signal as $\mathbf{h}(t) = \tilde{\mathbf{C}}(t)\mathbf{v}^*$. $\mathbf{h}(t)$ is calculated for a short time window l. The long term blood volume pulse signal can be derived by stacking the short interval pulse signal $\mathbf{h}(t)$ together using overlap-adding [13].

3.3. Constraint plane

The above introduced method determines the optimal projection vector in the global color space based on the spectral characteristics. Inspired by POS, we refine this method by constraining the projection vector on a predefined plane. In Section 4, we will investigate the impact of the constraint plane on the algorithm performance. To differentiate the methods, we refer the refined method with constraint plane as PSCc (Projection vector based on Spectral Characteristics) and the original method as PSCg (global PSC).

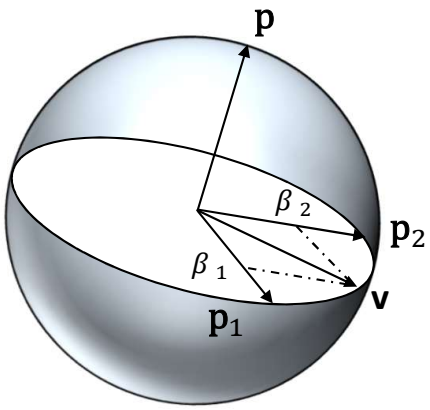

Figure 3. Constraining the optimal projection direction $\mathbf{v}$ on a plane which is denoted by its normal vector $\mathbf{p}$.

For stationary scenario, we choose the plane orthogonal to $\mathbf{1}$ in the temporally normalized color space as the predefined constraint plane, as in POS.

For cases with strong disturbance, we limit the projection vector onto the plane whose normal vector is in the direction of the signal variance of each channel. Projection onto this plane can compensate more disturbance energy than onto the POS-plane, since it is orthogonal to the signal energy.

We represent the constraint plane with its normal vector, as shown in Figure 3. Mathematically, the constraint plane can be expressed as:

$$\mathbf{p} = \begin{cases} [1,1,1], & \text{if } \max(|\mathbf{s}|) > \delta \\ sqrt\left(Diag\left(\mathbf{F}\mathbf{F}^T\right)\right), & \text{otherwise} \end{cases} \tag{10}$$

where $\mathbf{s} = [0,1,-1] \cdot \mathbf{F}$ and δ denotes the threshold used to distinguish between the two compensation cases. $Diag(\cdot)$ gives a vector with the diagonal elements of a matrix.

Then we represent the projection vector $\mathbf{v}$ as a linear combination of two basis vectors of $\mathbf{p}$:

$$\mathbf{v} = [\mathbf{p}_1, \mathbf{p}_2]\boldsymbol{\beta} \tag{11}$$

where $\mathbf{p}_1, \mathbf{p}_2$ are an arbitrary pair of basis vectors in the plane $\mathbf{p}$ and $\boldsymbol{\beta} \in \mathbb{R}^{2\times 1}$ denotes the coefficients. We write the basis vectors in matrix form as $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2]$ ($\mathbf{P}^T\mathbf{P} =$

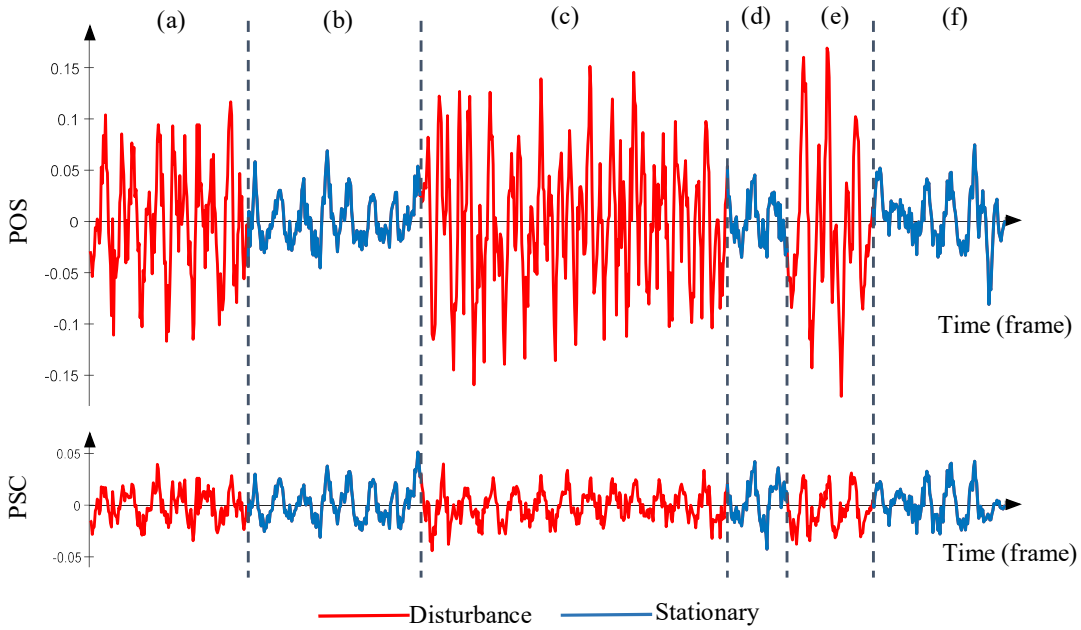

Figure 4. A comparison example between PSC and POS.

$\mathbf{1}, \mathbf{P}^T\mathbf{p} = 0$). Thus the determination of the projection vector $\mathbf{v}$ can be solved by finding the optimal coefficient vector $\boldsymbol{\beta}$. Substituting (11) into (4) we have:

$$\begin{aligned} \boldsymbol{\beta}^* &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \boldsymbol{\beta}^T\mathbf{P}^T\mathbf{M}_h\mathbf{P}\boldsymbol{\beta}, \\ s.t. \quad & \|\boldsymbol{\beta}\|_2 = 1 \end{aligned} \tag{12}$$

and substituting (11) into (5) gives:

$$\begin{aligned} \boldsymbol{\beta}^* &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \frac{\boldsymbol{\beta}^T\mathbf{P}^T\mathbf{M}_h\mathbf{P}\boldsymbol{\beta}}{\boldsymbol{\beta}^T\mathbf{P}^T\mathbf{M}_l\mathbf{P}\boldsymbol{\beta}} \\ s.t. \quad & \|\boldsymbol{\beta}\|_2 = 1 \end{aligned} \tag{13}$$

As $\mathbf{P}$ can be arbitrary chosen on the constraint plane, $\mathbf{P}^T\mathbf{M}_l\mathbf{P}$ and $\mathbf{P}^T\mathbf{M}_h\mathbf{P}$ in (12) and (13) can be considered as modified versions of $\mathbf{M}_l$ and $\mathbf{M}_h$ in (4) and (5). Thus, the determination of $\boldsymbol{\beta}^*$ can follow the same steps as in (4) and (5). The optimal projection direction $\mathbf{v}^*$ is then calculated using (11).

Figure 4 displays two signal segments extracted from a video captured using a webcam. The upper segment was obtained by POS and the lower one extracted by the refined PSC. The participants were instructed to conduct head motion in segments (a)(c)(e) and sit still in (b)(d)(f). In the figure we can see that in segments (a)(c)(e), the blood volume pulse signal extracted by POS was severely contaminated by the motion disturbance and has an overlarge signal amplitude. In comparison, the BVP extracted by the proposed method was effectively recovered and the disturbance was suppressed.

4. Experiment

In this section, we compare the proposed method with three state of the art methods: ICA, CHROM and POS. To demonstrate performance improvement by introducing the constraint plane, we also compared performance of the refined method PSCc and the method without the constraint plane PSCg. Section 4.1 introduces the dataset on which we evaluated the algorithms. Section 4.2 illustrates the pipeline in which the evaluation was conducted. In Section 4.3 we explain the metrics used for the performance evaluation and comparison results will be discussed in Section 4.4.

4.1. Dataset

The algorihtms were validated on a dataset comprising 26 participants and 7 scenarios, including one stationary scenario, two scenarios which simulate situations of evironment illumination change and four scenarios of body motion, as listed in Table 1. The videos were recorded using Logitech Pro Webcam HD C920 with a resolution of 640 x 480 pixels and a bit depth of 8 Bit. The frame rate of the recording was 30 fps. Each video has a length of 2 minutes. The subjects were made of 20 males and 6 females and in age between 23 and 33 years old. The reference signal was provided by an finger clip Photoplethysmography (PPG) sensor.

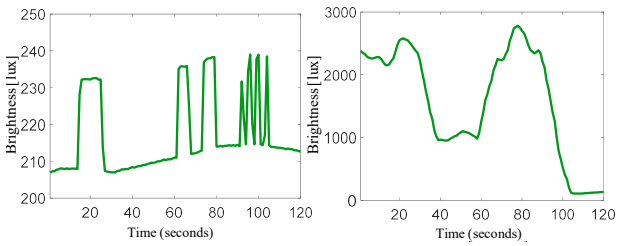

Figure 5. Example course of light changes for Scenario 103 and 104.

Figure 6. Screenshot of motion scenarios 201-203.

In the first scenario, natural daylight was used as illumination source. In scenarios 103 and 104, slow and fast lighting changes were induced by means of REST API based smart home light shutter control. A lux meter was used to measure the brightness during the recordings. Figure 5 illustrates the brightness change of two video examples of scenarios 103 and 104 measured by the lux meter.

In scenarios 201 – 203, the subjects were instructed to carry out certain movements (translatory, rotational and scaling) at specified points in time. The screen-shots are shown in Figure 6. In scenario 204, all subjects were asked to write a text, which induces random movements.

4.2. Pipeline

For a fair comparison, all methods should share the same processing pipeline, namely, the identical ROI segmentation and postprocessing steps.

ROI segmentation was performed by a face tracker implemented using supervised descent method (SDM) [16]. We first defined a mean face shape by performing Procrustes analysis on the label data in the LFPW dataset [17], which is a dataset for face alignment tasks. Since the landmarks above the forehead are not included in defined landmarks, we extended two points at the eyebrow along the direction from the nose tip to the nose bridge, to obtain two extra points on the forehead, as plotted in Figure 7(a). The mean face shape was then divided into subregions using Delaunay triangulation. Then, we uniformly sampled pixels inside of the face shape as signal sensors. The total number of pixel sensors was 490. For a given image frame, vertice

Table 1. Overview of dataset scenarios

Nr.	Scenario	
	Name	Description
101	Natural lighting	Daylight without movement
103	Abrupt changing lighting	Scenario with rapid and smaller light changes
104	Slowly changing lighting	Scenario with slow and larger light changes
201	Rotatory movement	Motion scenario with rotatory head movement
202	Scaling movement	Motion scenario with scaling head movement
203	Translatory movement	Motion scenario with translatory head movement
204	Text writing	Motion scenario by writing a text

in each triangle mesh defined in (a) were detected by the face tracker, as shown in Figure 7(b). Each pixel (x, y) inside the mesh can be tracked based on vertex locations of the triangle in which it locates, by utilizing piecewise affine warping [18], as shown in Figure 7(c). The tracked pixels are displayed in Figure 7(d).

For each tracked pixel in the face image, we calculated raw signals by averaging pixel intensity in a local patch. The obtained signals were then concatenated temporally into traces of length l. For CHROM, POS and the proposed methods, we chose the trace length as $1.2s$. As ICA uses the statistical characters of signals and requires effectively enough samples of observed variable, we set for ICA a longer trace length of $10s$. Since the defined base face shape mesh also covers non-skin regions, we applied the one-class support vector machine [19] to prune non-skin pixels, similarly in [20].

The extracted raw signals were then processed by the above listed core rPPG algorithms. We implemented CHROM and POS according to the framework proposed in [14]. For ICA, JADE [21] was adopted as the core algorithm. The extracted signal extracted by the core rPPG algorithms was then overlap-added into a long term blood pulse signal with the overlap length of $l - 1$.

To compare the performance of the algorithms, we evaluated the signal quality of the long term blood pulse signals. The evaluation was performed with a sliding window of 30s. The sliding step was chosen as $1s$, which gives 91 signal snippets for each video of $120s$.

4.3. Metrics

We used the Signal Noise Ratio (SNR) and Area under Curve (AUC) as the evaluation metrics in our experiments.

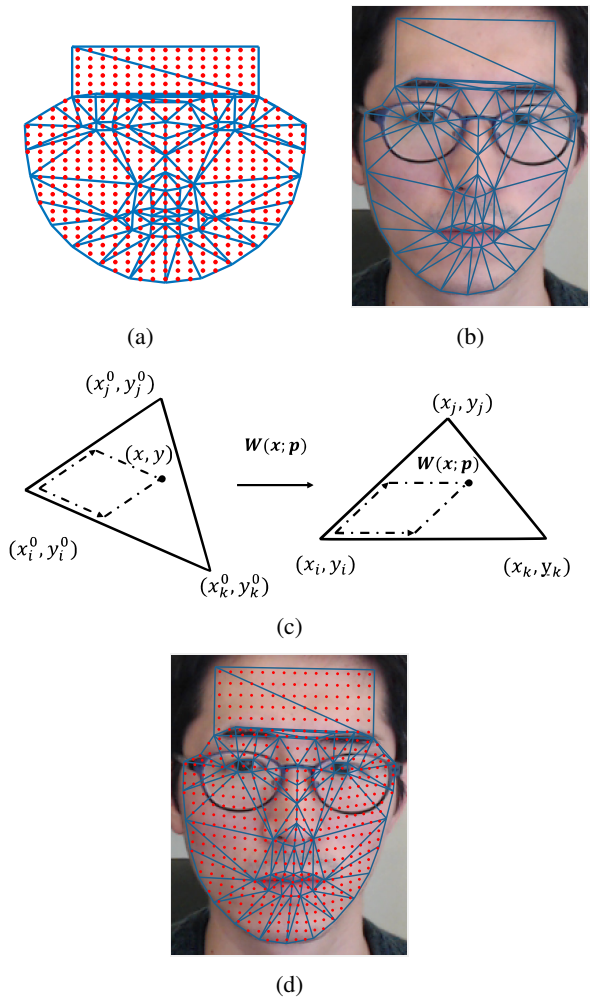

Figure 7. Alignment of pixels in ROI (a) pixels are uniformly sampled in a mean shape; (b) facial landmarks are registered by a face alignment algorithm; (c) pixels in each triangle is registered through affine warping, $(x^0_{(i,j,k)}, y^0_{(i,j,k)})$ denotes the vertex location in the base mean shape, and $(x_{(i,j,k)}, y_{(i,j,k)})$ represents their location in the aligned face image. $\mathbf{W}(x, y;\ \mathbf{p})$ is the warped location of (x, y) in the image frame; (d) aligned pixels in the ROI.

The SNR was estimated by the following formula, also proposed in [13]:

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{E_{\mathrm{signal}}}{E_{\mathrm{noise}}}\right) \tag{14}$$

where E_{signal} is the energy around the peak frequency of the reference signal and its first harmonic frequency in the spectrum, calculated from the extracted blood pulse signal, while E_{noise} includes the remaining energy in the spectrum.

AUC indicates the percentage of estimated heart rate within an error tolerance of the reference. We evaluated the heart rate from the spectrum by selecting the frequency peak. The maximal tolerance for AUC was set as 5 bpm.

Table 2. SNR of algorithms on the benchmarked dataset

Scenario	ICA	CHROM	POS	PSCg	PSCc
101	2.74	3.08	**3.54**	2.23	**3.43**
103	**0.28**	-0.40	-0.49	**-0.10**	-0.26
104	2.42	2.33	**2.63**	2.15	**3.06**
201	-3.17	-3.01	-2.65	**-1.40**	**-1.13**
202	0.91	1.21	**1.64**	0.80	**1.56**
203	-0.06	-0.33	**-0.04**	-0.30	**0.52**
204	-1.51	-1.46	**-0.79**	-1.28	**-0.77**

Table 3. AUC of algorithms on the benchmarked dataset

Scenario	ICA	CHROM	POS	PSCg	PSCc
101	0.705	0.711	**0.714**	0.711	**0.715**
103	**0.647**	**0.655**	0.640	0.625	0.634
104	0.702	**0.714**	0.712	0.705	**0.716**
201	0.413	0.468	0.472	**0.524**	**0.541**
202	0.605	**0.637**	**0.623**	0.603	0.622
203	0.542	0.543	**0.562**	0.557	**0.578**
204	0.555	0.564	0.602	**0.614**	**0.615**

4.4. Results and discussion

Table 2 and 3 summarizes the SNR and AUC values of the rPPG algorithms obtained in each scenario, where the red and blue entries denote the best (red) and second best (blue) results obtained by corresponding methods.

As shown in the tables, PSCg did not outperform the other state-of-the-art methods in all scenarios but provided comparable results. This validates the feasibility to use the spectral characteristics to design the projection vector. By combining spectral characteristics and the constraint plane, PSCc has achieved much better results and outperforms other algorithms in the most scenarios.

Comparing results cross scenarios, we can find that all algorithms performed well in Scenario 101. POS and PSCg show the best SNR and AUC respectively.

In Scenario 103, ICA gives the highest SNR value and CHROM shows the highest AUC. PSCg and PSCc fail to bring an obvious performance improvement. The abrupt illumination change disturbance in this scenario was simulated by fast switching on and off the ambient light source. The disturbance induced by the ambient light source has different waveform in each color channel, leading to different frequency distribution in the frequency domain. This makes the separation direction cross the frequency domain not uniform, thereby the compensation direction based on the HF components can not cancel out disturbance components in

the low frequency range.

Results for Scenario 104 show that slow and moderate illumination disturbance has a limited impact on performance of all the methods. In this scenario, PSCc achieves the best SNR and AUC.

In the scenarios in presence of motion disturbance, all algorithms have a performance drop. In Scenario 202 for scaling head motion, CHROM shows slightly better heart rate estimation accuracy than other methods and POS achieves the best SNR. In Scenario 201,203 and 204, PSCc outperformed other methods. Especially in Scenario 201, where the most complicate skin-light interaction was involved, both PSCg and PSCc achieved performance advancement upon other algorithms. This can be explained by the fact that both illumination and secular components produced strong variations in this scenario when head motion occurs, which sabotages the assumed linearity of the proposed model in [14] and makes the opposite phase-alignment property invalid for CHROM and POS.

In summary, we can say that the spectral characteristics can be used to define a projection vector for rPPG signal extraction, if the disturbance has the same waveform in each color channel. Combining with the constraint plane, the proposed method provides the best overall performance.

5. Conclusion and limitation

In this paper, we have proposed a new approach PSC to extract the pulsatile signal by defining a projection vector based on spectral characteristics of rPPG. Inspired by POS algorithm, we refine the method by limiting the projection vector onto a constraint plane.

The proposed method was evaluated on a benchmark dataset including 7 scenarios. The evaluation shows that the proposed method performs as well as or better than the state of the methods in the stationary scenario or scenarios where disturbance has the same waveform in all color channels.

One limitation of the proposed algorithm is that the proposed method considers the stationary and disturbance scenarios separately, resulting in increased complexity of parameter setting for the algorithm. Moreover, the proposed method determines the optimal projection vector by solving an optimization problem, which is slower than the one-step alpha-tuning used by POS and CHROM. Since the optimization problem has a maximal dimension of 3 and can be solved analytically, the proposed method can still achieve a real-time operation.

References

[1] R.-Y. Huang and L.-R. Dung, "A motion-robust contactless photoplethysmography using chrominance and adaptive filtering," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, 2015. 1

[2] D. J. McDuff, S. Gontarek, and R. W. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2014, pp. 2957–2960, 2014. 1

[3] T. Blöcher, J. Schneider, M. Schinle, and W. Stork, "An online ppgi approach for camera based heart rate monitoring using beat-to-beat detection," in *2017 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2017. 1

[4] B.-F. Wu, Y.-W. Chu, P.-W. Huang, M.-L. Chung, and T.-M. Lin, "A motion robust remote-ppg approach to driver's health state monitoring," in *COMPUTER VISION - ACCV 2016 WORKSHOPS, PT I*, vol. 10116, pp. 463–476, 2016. 1

[5] A. C. Kevat, D. V. R. Bullen, P. G. Davis, and C. O. F. Kamlin, "A systematic review of novel technology for monitoring infant and newborn heart rate," *Acta Paediatrica*, vol. 106, no. 5, pp. 710–720, 2017. 1

[6] P. Pham and J. Wang, "Understanding emotional responses to mobile video advertisements via physiological signal sensing and facial expression analysis," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 67–78, 2017. 1

[7] A. Dassel, R. Graaff, W. Zijlstra, and J. Aarnoudse, "Reflectance pulse oximetry at the forehead of newborns : The influence of varying pressure on the probe," *Journal of Clinical Monitoring and Computing*, vol. 12, no. 6, pp. 421–428, 1996. 1

[8] D. J. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2593–2601, 2014. 1, 2

[9] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011. 1, 2

[10] R. Macwan, Y. Benezeth, and A. Mansouri, "Heart rate estimation using remote photoplethysmography with multi-objective optimization," *Biomedical Signal Processing and Control*, vol. 49, pp. 24–33, 2019. 1, 2

[11] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light.," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008. 1

[12] G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras," *Journal of Biomedical Optics*, vol. 20, no. 4, pp. 48002–48002, 2015. 1

[13] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013. 2, 4, 7

[14] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017. 2, 6, 8

[15] W. Wang, A. den Brinker, S. Stuijk, and G. de Haan, "Amplitude-selective filtering for remote-ppg," *Biomedical Optics Express*, vol. 8, no. 3, pp. 1965–1980, 2017. 3

[16] X. Xiong and F. D. la Torre, "Supervised descent method and its applications to face alignment," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2013, pp. 532–539, 2013. 6

[17] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2930–2940, 2013. 6

[18] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004. 6

[19] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004. 6

[20] W. Wang, S. Stuijk, and G. de Haan, "Exploiting spatial redundancy of image sensor for motion robust rppg," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 415–425, 2015. 6

[21] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999. 6