

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 Verkruyse *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

利用干扰场景增强远程PPG脉冲提取 光谱特性

Kai Zhou, Simon Krause, Timon Bl'ocher
FZI计算机科学研究中心
Karlsruhe, 德国

{zhou (对不起, 先生, 先生) @ fz

威廉·斯托克信息处理技术学院

卡尔斯鲁厄理工学院
德国卡尔斯鲁厄

wilhelm.stork@kit.edu

摘要

近年来, 已经提出了几种用于远程光电容积脉搏波描记(*rPPG*)的方法, 并且最近提出的方法在测量精度上取得了实质性的改进。然而, 没有一种方法研究了使用频谱特性来设计*rPPG*信号提取算法的可能性。在本文中, 我们提出了一种新的*rPPG*测量方法, 利用*rPPG*信号的频谱特性。我们在一个基准数据集上验证了新提出的方法, 该数据集包括7个场景和26个参与者。验证实验的结果证明了利用频谱特征提取*rPPG*信号的可行性。通过与约束平面相结合, 新提出的方法提供了更好的整体性能。

1.近年来, 远程光电容积描记术(*rPPG*)作为一种非接触式生命体征测量手段越来越受欢迎。它能够使用简单的视觉系统分析上部皮肤组织中的微血管血容量变化。

*rPPG*已在许多领域中得到应用, 例如健康监测[1]、压力监测[2]、驾驶员状态监测[3][4]、新生儿护理[5]和产品反应分析[6]。*rPPG*系统不需要复杂的放置过程, 也不需要像ECG或普通PPG那样直接接触皮肤。Dassel等人[7]已经表明, 皮肤组织与传统接触式PPG的光传感器之间的压力会影响PPG信号的波形。此外, 接触式PPG传感器往往会引起皮肤刺激, 从而降低长期使用的舒适度。

皮肤组织中的微血管血容量由心脏活动调节, 这导致微血管血容量的变化。

血红蛋白量, 从而导致皮肤颜色的波动。*rPPG*系统的目的是从视频中恢复血容量脉搏信号(BVP)并提取生命参数(例如, 脉搏率)。通常, 使用*rPPG*提取脉搏信号包括三个常见步骤: ROI检测、BVP信号提取和生命参数计算。ROI检测通过检测每帧视频中的人脸区域来完成。原始信号被计算为ROI中的像素强度。然后, 采用核心*rPPG*算法从原始信号中恢复血容量脉搏(BVP)。心率等重要参数可以在时域中通过计算心跳间隔[8][9]或搜索FFT频谱中的最大峰值[10]来检索。这个一般过程如图1所示。

Verkruyse等人[11]给出了在人脸视频记录中检测脉搏率的早期方法。血容量脉搏信号仅使用环境光和低成本相机来显示。他的工作表明, 从绿色通道提取的信号具有最高的脉动性。除了与皮肤颜色和照明条件的固有特性相关的脉动和AC分量之外, 从信号颜色通道提取的信号轨迹还包含由于相机的量化噪声以及照明和运动的突然变化而引起的干扰分量。盲源分离(Blind Source Separation, BSS)方法通过将原始信号解混为不同的源来实现脉冲恢复。Tendon和Li[12]对原始轨迹进行了主成分分析(PCA), 并将脉搏信号确定为具有最大方差的分量。

还研究了使用独立分量分析(伊卡)解决*rPPG*问题的方法[9]。与研究信号方差的PCA不同, 伊卡通过最小化解混信号内的高斯性来将脉动信号与噪声分离。在[8]中, McDuff等人使用伊卡提取血容量脉搏信号,

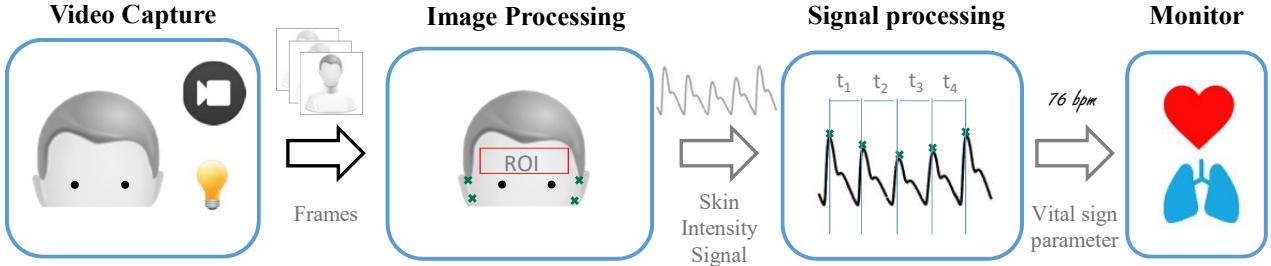


Figure 1. A computer vision algorithm registers ROI by detecting the face region in the video frame sequence and outputs raw color signals. A core kernel rPPG algorithm is adopted to recover the BVP signals from the raw signals, and finally, vital parameters are retrieved from the extracted BVP signals.

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.

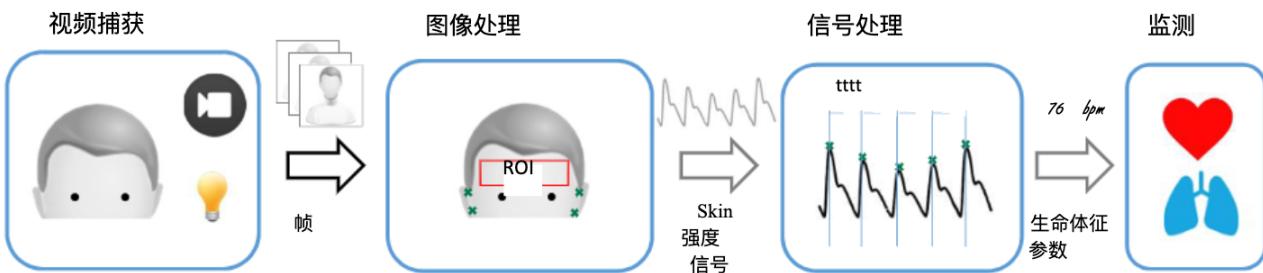
This opposite phase-alignment property in disturbance and target pulsatile components makes it possible to separate the pulsatile signal by deploying the so-called alpha-tuning [13]:

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

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

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



使用图1中的伊卡提取血容量脉搏信号。计算机视觉算法通过检测视频帧序列中的人脸区域来配准ROI，并输出原始彩色信号。采用核心核rPPG算法从原始信号中恢复BVP信号，最后从提取的BVP信号中提取重要参数。

由五波段相机（RGBCO）捕获的颜色轨迹。由于解混变换的线性性，分离信号的阶数不确定性是影响伊卡提取信号质量的模糊性之一。Poh等人[9]选择了应用伊卡后第二个分量内的信号，Mcduff等人[8]选择了频域中峰值最高的信号。除了源信号的高斯性之外，Macwan等人[10]还考虑了BVP信号的周期性。他们将其作为高斯最小化的目标函数的额外项，实现了传统伊卡的性能提升。在他的方法中，自相关被用作周期性的度量。

PCA和伊卡是仅利用血容量脉搏信号的统计特性的数据驱动的解决方案。还提出了利用rPPG的物理先验知识的方法，并且在BSS方法之上带来了更高的测量精度和鲁棒性。哈安和Jeanne [13]提出了一种基于色度的方法（CHROM），通过假设标准化的肤色对图像进行白平衡来提取血容量脉搏信号。Wang等人[14]提出了一种方法（POS），该方法通过将颜色轨迹投影到强度分量抵消的平面上来提取BPV信号。

在这项工作中，我们调查的频谱特性的rPPG，并提出了一种方法，将这些特性的rPPG信号提取过程。第二节介绍了问题的定义和所提出的方法的理论。在第3节中，将对所提出的方法进行数学描述。验证实验在第4节中进行。在第五部分，我们得出结论。

每个颜色通道的原始信号是通过对每个单独视频帧中ROI内的像素值进行空间平均，然后将它们在时间上连接成信号轨迹来提取的。将每个颜色通道的信号轨迹叠加，得到RGB相机的原始信号矩阵C，大小为 $I \times 3$ ，其中I代表时间窗口的长度。

CHROM和POS从零均值皮肤颜色信号 $\langle C(t) \rangle$ 中提取脉搏 $C(t)$ 投影到基于物理模型和皮肤特征的先验知识定义的颜色空间中的一个平面上，第一步投影补偿了一些干扰能量，并给出了两个信号迹S和 S' 。产生的信号Sand Shave相对于脉动信号和剩余扰动具有相反的相对相位关系。例如，在POS中，沙萨雷中的脉动分量彼此处于一致相位，而其余的扰动分量处于相反相位。

干扰和目标脉动分量中的这种相反的相位对准特性使得可以通过部署所谓的双调谐来分离脉动信号[13]：

$$h = s + a \frac{\sigma(s)}{\sigma(s')} \cdot s \quad (1)$$

其中 $\sigma(\cdot)$ 代表信号的标准偏差。 a 对于CHROM为-1，对于POS为1。阿尔法调谐抵消了沙沙中的干扰分量，保留了脉动分量。

然而，在某些情况下，在干扰的存在下，相反的相位对准属性并不总是保持由于光皮肤相互作用的强非线性。因此，在POS中可能发生的情况是，沙萨雷中的扰动分量实际上是同相的。在这种情况下，应用具有正值 α 的 α 调谐只能增强干扰强度，而不能正确地提取信号。对于在这种情况下的rPPG信号提取，我们提出基于频谱特征来定义投影向量。

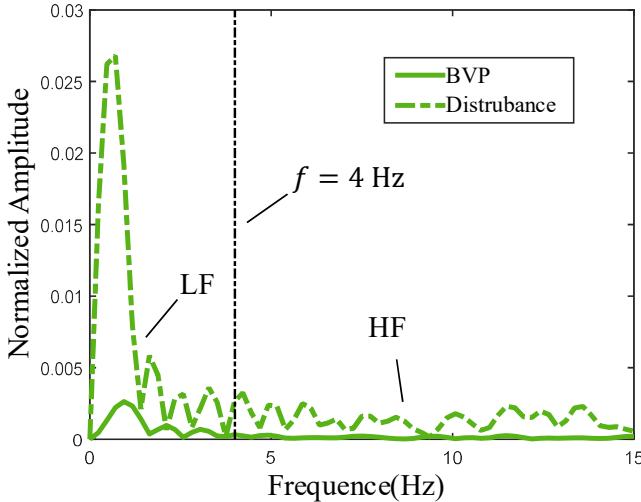


Figure 2. Comparison between spectra of signals respectively dominated by disturbance and pulsatile.

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 projec-

tion 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:

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

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 $\tilde{\mathbf{C}}(t)$ 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 :

$$\mathbf{F}^T = [\mathbf{F}_l^T, \mathbf{F}_h^T], \quad (3)$$

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:

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

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

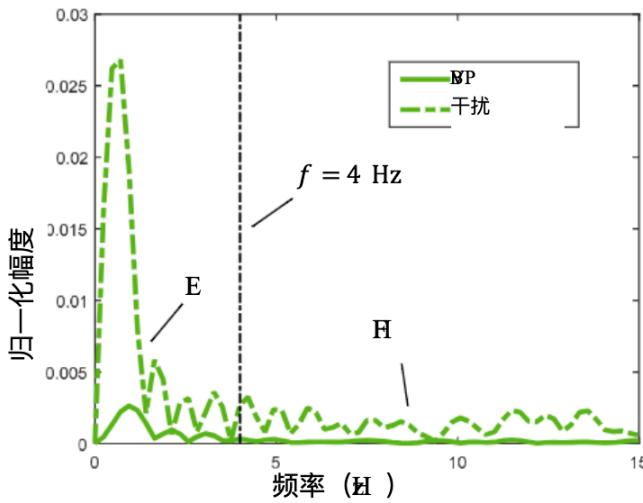


图2.扰动信号和脉动信号的频谱比较。

信号的特征。

我们首先研究两种不同情况下的原始信号。图2绘制了绿色通道的两个信号片段的光谱。信号片段以1.2秒的时间窗口裁剪，视频的帧速率为30 fps。点划线表示被干扰严重污染的信号，而实线是从静止场景中提取的。

从图中可以观察到，污染信号具有比从静止场景提取的信号高得多的幅度。对于存在干扰的情况，我们可以考虑几乎由干扰引起的高频范围内的信号功率。如果干扰在每个颜色通道中具有相同的波形，则抵消高频(HF)范围中的干扰的投影矢量应该也能够抑制低频(LF)范围中的干扰，这意味着投影矢量与干扰分量正交。假设脉动分量和扰动在归一化颜色空间中是物理上可分离的，脉动分量和扰动之间应该有足够大的角度，因此脉动分量和产生的投影向量彼此不正交。将彩色迹线投影到干扰最小化向量上将给予具有可观察幅度的脉动信号。因此，对于干扰场景，我们将投影方向确定为高频能量最小化的矢量。

对于静止情况，信号基本上由目标血量脉冲组成。信号能量主要分布在低频范围

这种情况下的投影方向应该能够增强低频分量并同时抑制高频分量。最佳投影向量可以通过最小化HF分量与LF分量的比率来计算。

由于人体相对脉动幅度在特定范围内变化，并且相对幅度(AC/DC)要大得多[15]，因此我们可以通过设置相对幅度的阈值来轻松区分静止和干扰场景。

3.方法在本节中，我们将更详细地描述我们的方法。该方法的目的是找到最佳投影向量 $v \in \mathbb{R}$ ，使得脉动信号可以被提取为：

$$h(t) = C(t)^* v \quad (2)$$

由于稳态和扰动情景采用不同的目标，因此将分别介绍这两种情景的方法。之后，我们建议通过对投影向量施加约束来提高我们的方法的性能。

优化过程在频域中执行。对时间归一化的颜色轨迹执行快速傅立叶变换给出复矩阵 $F \in \mathbb{C}$ ，其中 f 表示频率仓的数量。我们通过设置频率阈值 f 将 F 分解为两个子矩阵：

$$F = [F, F], \quad (3)$$

其中 F 表示频率为 $(0, f) \subset [f-f, f]$ 的所有分量， F 表示频率范围 $(f, f-f)$ 内的分量。 f 是录制的帧速率。 f 的设置应确保如果发生扰动，高频能量主要由扰动贡献。我们选择 $f=4$ Hz，这对应于人类心率的上限。

3.1.在有干扰的情况下，彩色迹线中的高频分量几乎都是由干扰引起的。为了抑制干扰能量，我们寻找一个单位向量上的HF分量的能量是最小的。在数学上，它可以表示为：

$$\begin{aligned} v &= \operatorname{argmin}_v \quad v^T F^T F v \\ &\quad , \quad \operatorname{argmin}_v \quad v^T M v, \\ &\quad \text{S.T. } v \end{aligned} \quad (4)$$

其中 $M, F^T F$ 是一个真实的矩阵，这是由于 F 在频域中的对称性。 F 是 F 的厄米特转置。可以确定投影向量 v

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:

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

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:

$$\mathbf{M}_l = \mathbf{U}_l \boldsymbol{\Lambda} \mathbf{U}_l^T \quad (6)$$

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

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

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

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

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

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

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

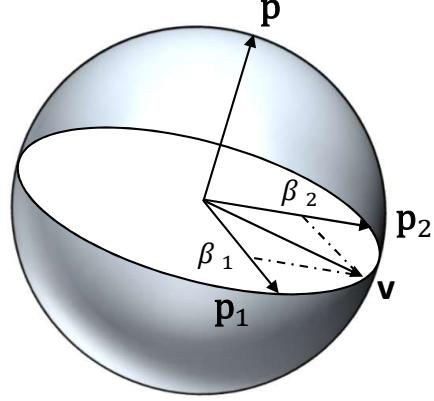


Figure 3. Constraining the optimal projection direction \mathbf{v} on a plane which is denoted by its normal vector \mathbf{p} .

spectral characteristics. Inspired by POS, we refine this method by constraining the projection vector on a pre-defined 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).

For stationary scenario, we choose the plane orthogonal to 1 in the temporally normalized color space as the pre-defined 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:

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

where $\mathbf{s} = [0, 1, -1] \cdot \mathbf{F}$ and δ denotes the threshold used to distinguish between the two compensation cases. $\text{Diag}(\cdot)$ 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} :

$$\mathbf{v} = [\mathbf{p}_1, \mathbf{p}_2]\beta \quad (11)$$

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

投影向量 v 可以被确定为对应于最小特征值的 M 的特征向量。
3.2在平稳场景中，最佳投影向量 v 应该能够增强低频域中的能量并降低高频噪声。 V 的确定可以表示为：

$$\begin{aligned} v &= \underset{v}{\operatorname{argmin}} \frac{v^T F F v}{v^T F v} \\ &\quad , \underset{v}{\operatorname{argmin}} \frac{v^T M v}{v^T M v} \end{aligned} \quad (5)$$

S.T. v

其中 M , FF , 其为广义瑞利商的形式。分解马斯：

$$M = U \Lambda U \quad (6)$$

其中 U 的列是 M 的右特征向量， Λ 是对应特征值的对角矩阵。
(5) 的分母可以写为：

$$\begin{aligned} v^T M v &= v^T U \Lambda U v = v^T \Lambda v \\ &= \alpha \alpha \end{aligned} \quad (7)$$

其中 α , Λ UV。将该结果与等式 (5) 组合并替换 v , 我们得到：

$$\begin{aligned} \alpha &< \sum_{i=1}^n \lambda_i < \infty \\ \alpha &= \frac{\alpha \alpha}{\alpha \alpha} \end{aligned} \quad (8)$$

其中 S , $\sum_{i=1}^n \lambda_i < \infty$ 。 α 可以确定为 S 的最小特征值对应的特征向量。最佳投影向量可以使用以下公式计算：

$$v = \frac{U \Lambda \alpha}{\|U \Lambda \alpha\|} \quad (9)$$

然后, 我们将血液脉搏信号计算为 $h(t) = C(t)v \cdot h(t)$ 是针对短时间窗口 l 计算的。长期血容量脉搏信号可以通过使用叠加将短间隔脉搏信号 $h(t)$ 叠加在一起来导出[13]。

3.3.上面介绍的方法基于约束平面确定全局颜色空间中的最佳投影向量。

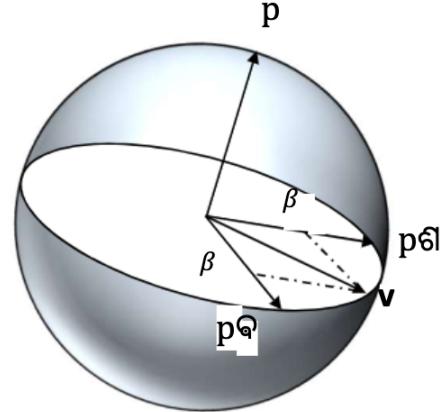


图3.将最佳投影方向 v 约束在由其法向量 p 表示的平面上。

光谱特性受POS的启发, 我们通过将投影向量约束在预定义的平面上来改进该方法。在第4节中, 我们将研究约束平面对算法性能的影响。为了区分这两种方法, 我们将具有约束平面的改进方法称为PSCc (基于谱特征的投影向量), 将原始方法称为PSCg (全局PSC)。

对于静止场景, 我们选择时间归一化颜色空间中与1正交的平面作为预定义的约束平面, 如POS中那样。

对于强干扰的情况, 我们将投影向量限制在其法向量在每个通道的信号方差方向上的平面上。投影到该平面上可以比投影到POS平面上补偿更多的干扰能量, 因为它与信号能量正交。

我们用法向量表示约束平面, 如图3所示。在数学上, 约束平面可以表示为:

$$p = \begin{cases} [1, 1, 1], & \text{如果 } \max(|S|) > \delta \\ \sqrt{(\text{Diag}(FF))}, & \text{否则} \end{cases} \quad (10)$$

其中 $S = [0, 1, -1] \cdot F$, δ 表示用于区分两种补偿情况的阈值。Diag(·)给出一个具有矩阵对角元素的向量。

然后, 我们将投影向量 v 表示为 p 的两个基向量的线性组合:

$$p = [p] \beta \quad (11)$$

其中 p , 帕雷平面 p 中的任意一对基向量, $\beta \in \mathbb{R}$ 表示系数。我们将矩阵形式的基向量写为 $P = [p, p]$ ($PP =$

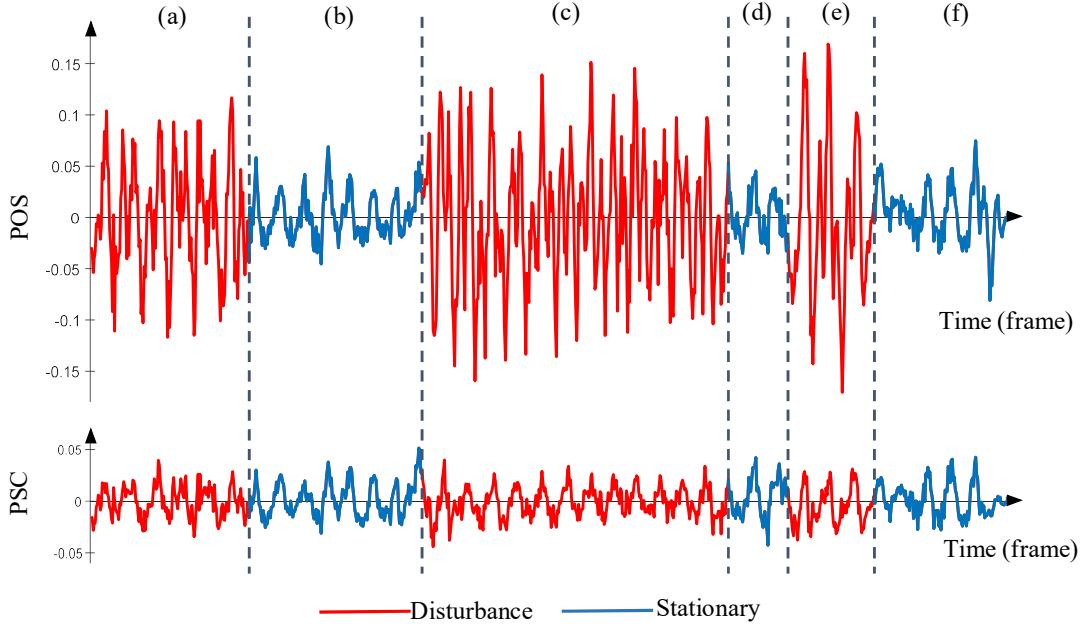


Figure 4. A comparison example between PSC and POS.

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

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

and substituting (11) into (5) gives:

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

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 β^* 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 environment 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.

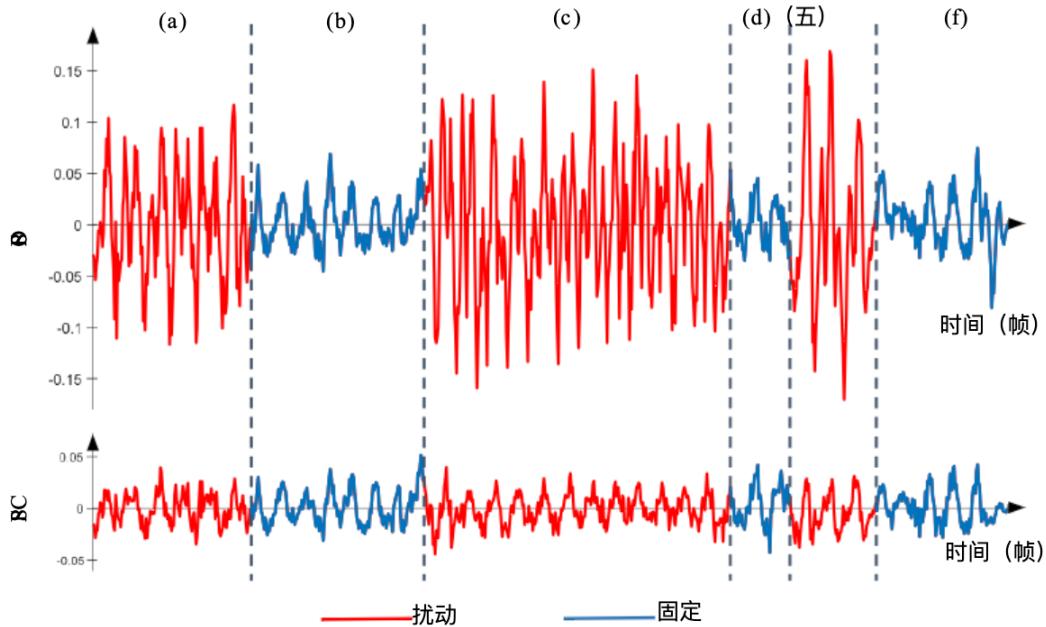


图4. PSC和POS的比较示例。

p_1 ($P_P = 1$, $P_p = 0$)。因此,可以通过找到最佳系数向量 β 来解决投影向量 v 的确定。将(11)代入(4),我们得到:

$$\begin{aligned} \beta &= \underset{\beta}{\operatorname{argmin}} \quad \beta P M P \beta, \\ \text{S.T. } \beta & \end{aligned} \quad (12)$$

将(11)代入(5)得到:

$$\begin{aligned} \beta^* &= \underset{\beta}{\operatorname{argmin}} \quad \beta P M P \beta \\ \text{S.T. } \beta & \end{aligned} \quad (13)$$

由于 P 可以在约束平面上任意选择,因此(12)和(13)中的 PMP 和 PMP 可以被认为是 M 和 Min (4)和(5)的修改版本。因此, β 的确定

可以遵循与(4)和 $^*(5)$ 中相同的步骤。然后使用(11)计算最佳投影方向维斯。

图4显示了从使用网络摄像头捕获的视频中提取的两个信号段。通过POS获得上段,通过改进的PSC提取下段。要求参与者在(a)(c)(e)段中进行头部运动,在(B)(D)(F)段中静坐。在图中,我们可以看到,在段(a)(c)(e)中,通过POS提取的血量脉搏信号被运动扰动严重污染,并且具有过大的信号幅度。相比

该方法提取的边值问题得到了有效的恢复,抑制了干扰。

4.实验在本节中,我们将所提出的方法与三种最先进的方法进行比较:伊卡、CHROM和POS。为了证明通过引入约束平面的性能改进,我们还比较了改进方法PSC和没有约束平面PSCg的方法的性能。第4.1节介绍了我们评估算法的数据集。第4.2节说明了进行评价的管道。在第4.3节中,我们解释了用于性能评估的指标,比较结果将在第4.4节中讨论。

4.1数据集算法在包括26名参与者和7个场景的数据集上进行验证,包括一个静止场景、两个模拟环境照明变化的场景和四个身体运动场景,如表1所列。这些视频使用Logitech Pro Webcam HD C920录制,分辨率为 640×480 像素,位深度为8位。记录的帧速率为30 fps。每个视频的长度为2分钟。受试者男20例,女6例,年龄23~33岁。参考信号由指夹式光电体积描记(PPG)传感器提供。

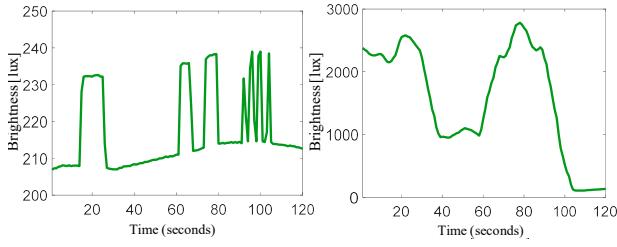


Figure 5. Example course of light changes for Scenario 103 and 104.



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, vertex

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.

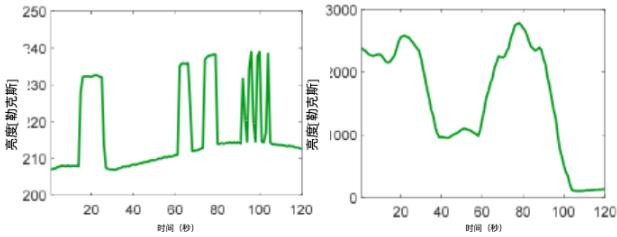


图5.场景103和104的灯光变化过程示例。



图6.运动场景201-203的屏幕截图。

在第一种情况下，使用自然光作为照明源。在场景103和104中，通过基于REST API的智能家庭光快门控制来引起缓慢和快速照明变化。勒克斯计用于测量记录期间的亮度。图5示出了由勒克斯计测量的场景103和104的两个视频示例的亮度变化。

在场景201 - 203中，受试者被指示在指定的时间点进行某些运动（平移、旋转和缩放）。屏幕截图如图6所示。在场景204中，所有受试者都被要求写一篇文章，这会引起随机运动。

4.2管道为了公平比较，所有方法都应该共享相同的处理管道，即相同的ROI分割和后处理步骤。

ROI分割由使用监督下降法（SDM）实现的面部跟踪器执行[16]。我们首先通过对LFPW数据集[17]中的标签数据执行Procrustes分析来定义平均人脸形状，LFPW数据集是用于人脸对齐任务的数据集。由于前额上方的标志不包括在定义的标志中，我们在眉毛处沿着从鼻尖到鼻梁的方向延伸两个点，以获得前额上的两个额外点，如图7（a）所示。然后使用Delaunay三角剖分将平均面部形状划分为子区域。然后，我们均匀采样的人脸形状内的像素作为信号传感器。像素传感器的总数为490。针对给定图像帧

表1.数据集方案概述

Nr.	名称描述	场景
101	自然光自然光无运动	
103	突变 照明	快速和 较小的光变化
104	慢慢改变 照明	场景缓慢， 更大的光变化
201	旋转 运动	带旋转的运动场景- 磁头移动
202	缩放 运动	运动场景， 定标头运动
203	平移 运动	运动场景与跨- 旋转头运动
204	文字撰写	运动场景 写作文

在（a）中定义的每个三角形网格中，由面部跟踪器检测到，如图7（b）所示。网格内的每个像素（x, y）可以根据其所在三角形的顶点位置进行跟踪，通过使用分段仿射扭曲[18]，如图7（c）所示。跟踪的像素显示在图7（d）中。

对于人脸图像中的每个跟踪像素，我们通过对局部补丁中的像素强度进行平均来计算原始信号。然后将获得的信号在时间上连接成长度为l的线。对于CHROM、POS和所提出的方法，我们选择线长为1.2s。由于伊卡利用信号的统计特性，并且需要足够的观测变量样本，因此我们为ICA设置了更长的10s迹长。由于定义的基础人脸形状网格也覆盖非皮肤区域，因此我们应用单类支持向量机[19]来修剪非皮肤像素，类似于[20]。

然后通过上述核心rPPG算法处理提取的原始信号。我们根据[14]中提出的框架实现了CHROM和POS。对于伊卡，采用JADE[21]作为核心算法。然后，将由核心rPPG算法提取的提取信号与重叠长度为l-1的长期血脉搏信号相加。

为了比较算法的性能，我们评估了长期血脉搏信号的信号质量。使用30 s的滑动窗口进行评价。滑动步长被选择为1 s，这为每个120 s的视频提供91个信号片段。

4.3.在实验中，我们使用信噪比（SNR）和曲线下面积（AUC）作为评价指标。

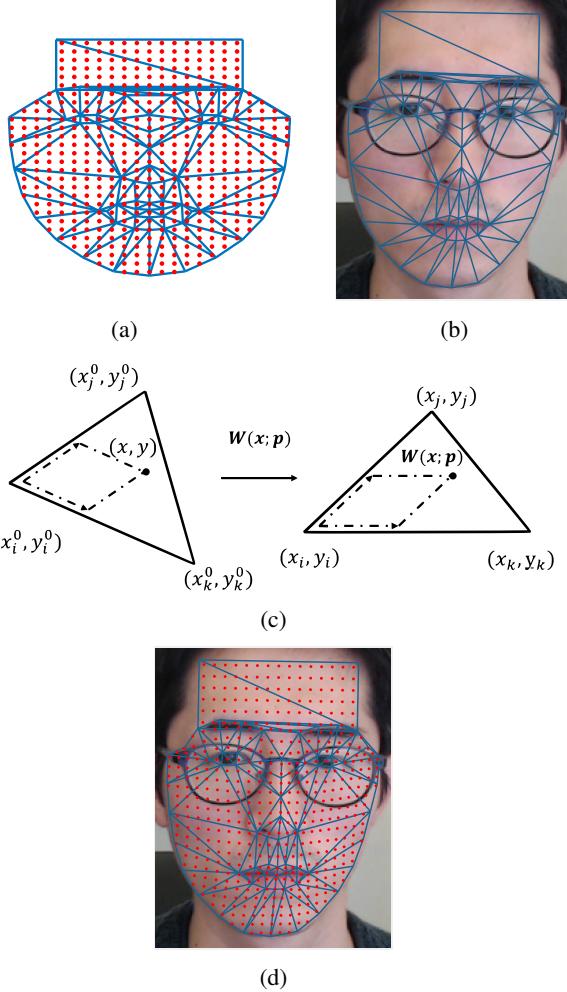


Figure 7. Alignment of pixels in ROI (a) pixels are uniformly sampled in a mean shape; (b) facial landmarks are registered by a face alignment algorithm; (c) pixels in each triangle is registered through affine warping, $(x_{(i,j,k)}^0, y_{(i,j,k)}^0)$ denotes the vertex location in the base mean shape, 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]:

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

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

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

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.

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

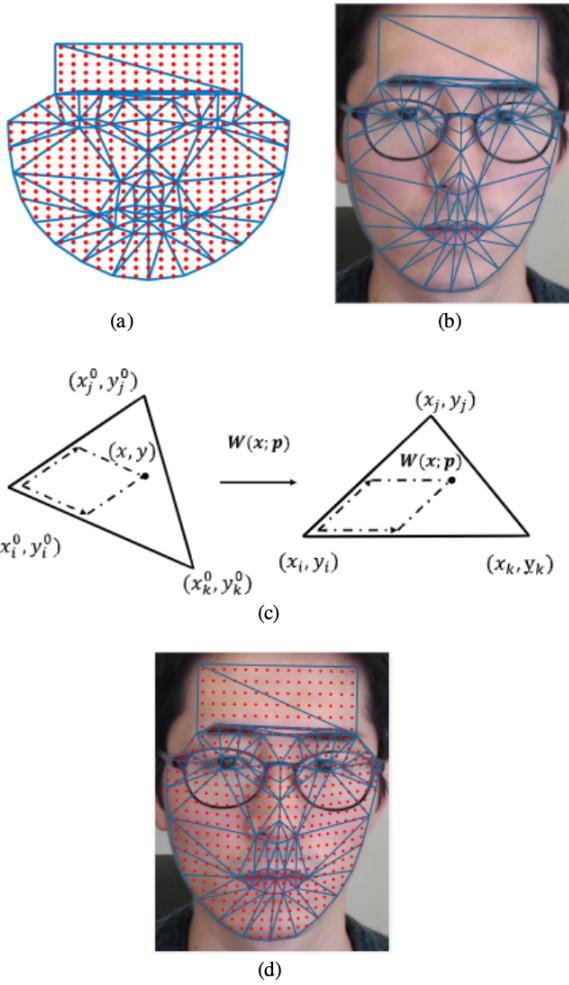


图7. ROI中像素的对准 (a) 以平均形状均匀地采样像素;
(B) 通过面部对准算法配准面部标志; (c) 通过仿射变形配准每个三角形中的像素, (x, y) 表示基本平均形状中的顶点位置, 并且 (x, y) 表示它们在对准的面部图像中的位置。 $W(x, y; p)$ 是图像帧中 (x, y) 的扭曲位置; (d) ROI中的对齐像素。

表2.基准数据集上算法的SNR

表演者:	ICA	CHROM	POS	PSCg
101	2.74	3.08	3.54	2.23
103	0.28	-0.40	-0.49	-0.10
104	2.42	2.33	2.63	2.15
201	-3.17	-3.01	-2.65	-1.40
202	0.91	1.21	1.64	0.80
203	-0.06	-0.33	-0.04	-0.30
204	-1.51	-1.46	-0.79	-1.28

表3.基准数据集上算法的AUC

表演者:	ICA	CHROM	POS	PSCg
101	0.70	0.71	0.71	0.71
103	0.64	0.65	0.64	0.62
104	0.70	0.71	0.71	0.71
201	0.41	0.46	0.47	0.52
202	0.60	0.63	0.62	0.60
203	0.54	0.54	0.56	0.55
204	0.55	0.56	0.60	0.61

在参考的误差容限内。我们通过选择频率峰值从频谱中评估心率。AUC的最大耐受性设定为5 bpm。

4.4结果和讨论表2和表3总结了在每个场景中获得的rPPG算法的SNR和AUC值, 其中红色和蓝色条目表示通过相应方法获得的最佳(红色)和次佳(蓝色)结果。

如表中所示, PSCg在所有情况下均未优于其他最先进的方法, 但提供了相当的结果。这验证了利用光谱特性设计投影矢量的可行性。PSCc算法将谱特征和约束平面相结合, 在大多数情况下取得了更好的效果, 优于其他算法。

通过跨场景比较, 我们可以发现所有算法在场景101中都表现良好。POS和PSCg分别显示最佳的SNR和AUC。

在场景103中, 伊卡给出了最高的SNR值, CHROM显示了最高的AUC。PSCg和PSCc没有带来明显的性能改善。通过快速开关环境光源来模拟该场景中的光照突变干扰。环境光源引起的干扰在每个颜色通道中具有不同的波形, 导致在频域中不同的频率分布。这使得跨越频域的分离方向不均匀, 从而基于HF分量的补偿方向不能抵消干扰分量,

SNR由以下公式估计, 也在[13]中提出:

$$10 - 12 - 2016 \text{ 刘晓} \frac{\text{信号}}{E}) \quad (14)$$

基线所提取的血液脉搏信号计算的参考信号的峰值频率及其频谱中的第一谐波频率周围的能量, 而E包括频谱中的剩余能量。

AUC表示估计心率的百分比

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. Benetzeth, 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. Verkruyse, 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

因此基于HF分量的补偿方向不能抵消低频范围内的干扰分量。场景104的结果表明，缓慢和适度的照明干扰对所有方法的性能影响有限。在这种情况下，PSCc实现最佳SNR和AUC。在存在运动扰动的情况下，所有算法的性能都会下降。在用于缩放头部运动的场景202中，CHROM显示出比其他方法稍好的心率估计准确度，并且POS实现了最佳SNR。在场景201、203和204中，PSCc优于其他方法。特别是在场景201中，涉及最复杂的皮肤-光交互，PSCg和PSCc都实现了其他算法的性能提升。这可以通过以下事实来解释：当头部运动发生时，照明和长期分量在这种情况下产生强烈的变化，这破坏了[14]中提出的模型的假设线性，并使相反的相位对准特性对于CHROM和POS无效。

总之，我们可以说，如果干扰在每个颜色通道中具有相同的波形，则频谱特性可以用于定义用于rPPG信号提取的投影向量。结合约束平面，该方法提供了最佳的整体性能。

5.结论和局限性在本文中，我们提出了一种新的方法PSCc，它从脉动信号定义的投影向量的基础上提取频谱特性的rPPG。受POS算法的启发，我们通过将投影向量限制在一个约束平面上来改进该方法。

在包括7个场景的基准数据集上对所提出的方法进行了评估。评估结果表明，所提出的方法执行以及或更好的状态的方法在静止的情况下或情况下，干扰具有相同的波形，在所有的颜色通道。

所提出的算法的一个局限性是，所提出的方法分别考虑平稳和扰动的情况下，导致增加的复杂性的参数设置的算法。此外，该方法通过求解优化问题来确定最优投影向量，这比POS和CHROM使用的第一步 α 调整慢。由于优化问题的最大维数为3，并且可以解析求解，因此所提出的方法仍然可以实现实时操作。

[2]D. J. McDuff, S. Gontarek 和 R. W. Picard, “通过心率变异性远程测量认知压力”，

会议记录：. IEEE 医学与生物工程学会年会。IEEE 医学与生物工程学会。Annual Conference, vol. 2014, pp. 2957-2960, 2014。1

[3]T. Blush, J. Schneider, M. Schinle 和 W. Stork, “An online ppgi approach for camera based heart rate monitoring using beat-to-beat detection,” in 2017 IEEE Sensors Appli.

阴离子研讨会 (SAS) , 页。2017年1-6月。1

[4]B- F. 吴玉- W. Chu, P.- W. 黄文 L. Chung 和 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. 的布伦, P. G. Davis 和 C. O. F. Kamlin, “监测婴儿和新生儿心率的新技术的系统综述”，《儿科学报》，第106卷，第5期，第107 - 108页。710-720, 2017年。1 [6] P. Pham 和 J. Wang, “通过生理信号感知和面部表情分析理解对移动的视频广告的情绪反应”，

第22届智能用户交互国际会议
Faces, pp. 67-78, 2017年。1

[7]A. 达塞尔河 Graaff, W. Zijlstra 和 J. Aarnoudse, “新生儿前额的反射脉搏血氧仪：探头上不同压力的影响”，《临床监测和计算杂志》，第12卷，第6期，第13页。421-428, 1996 1

[8]D. J. McDuff, S. Gontarek 和 R. W. Picard, “使用五个波段的远程心肺测量的改进

数字相机, "IEEE 生物医学工程学报-
ing, 第61卷, 第10期, 第10页。2593-2601, 2014 一、二

[9]M.- Z. Poh, D. McDuff 和 R. W. Picard, “非接触式多参数生理测量的进展

使用网络摄像头, "IEEE 生物医学工程学报-
neering, vol. 58, no. 1, pp. 2011年7月11日。一、二

[10]R. Macwan, Y. Benzeth 和 A. 心率-
使用具有多目标优化的远程光电体积描记术进行估计, ”
Biomedical Signal Processing and Control, 第49卷, pp. 2019年24-
33日。一、二

[11]W. 韦尔克昌伊斯湖 O. Svaasand 和 J.S. 纳尔逊, “远程
使用环境光的体积描记成像。”Optics Express, vol. 16, no. 26,
pp. 21434-21445, 2008. 1

[12]G. R. 和 Z. 李, “论另类色彩的好处
使用标准红-绿-蓝相机进行非接触式心率测量的空间”, Journal
of Biomedical Optics, 第20卷, 第4期, 第120 - 125页。48002-
48002, 2015. 1

[13]G. de Haan 和 V. Jeanne, “从
基于色度的rppg, ”IEEE Transactions on Biomedical
Engineering, 第60卷, 第10期, 第10页。2878-2886, 2013 二、
四、七

[14]W. WANG, A. C. 的 Brinker, S. 这是有趣的, 和 G. de Haan ,
“远程ppg的物理原理”, IEEE 生物医学工程学报, 第64卷, 第7
期, 第110页。1479-1491, 2017年。二、六、八

参考文献

[1]R.- Y. Huang 和 L.- R. Dung, “一种使用色度和自适应文件的运动鲁棒非接触式光电体积描记术,

2015年IEEE Biomedical Circuits and Systems Con.

(BioCAS) , pp. 2015年1-4月。1

- [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

- [15]W。WANG, A。的Brinker, S。这是有趣的, 和G。de Haan, “用于远程ppg的幅度选择性滤波”, 《生物医学光学快报》, 第8卷, 第3期, 第113页。1965-1980年, 2017年。3
- [16]X. Xiong和F. D. la Torre, “监督下降法 and its applications to face alignment”, 2013年IEEE计算机视觉和模式识别会议, 第2013卷, 第111 - 112页。532-539, 2013年。6
- [17]p n. Belhumeur, D。W。雅各布, D。J。Kriegman, 以及N。谁—Mar., “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年。6
- [18]I.马修斯和S.贝克, “积极的外观模型重新-visited, ”International Journal of Computer Vision, vol. 60, 号第2页。135-164, 2004. 6
- [19]D. M. J. Tax和R. P. W. Duin, “支持向量数据描述-第54卷, 第1期, 第100页。45-66, 2004年。6
- [20]W. Wang, S. Stuijk和G. de哈安, “利用空间资源, 用于运动鲁棒ppg的图像传感器的设计”, IEEE生物医学工程学报, 第62卷, 第2期, 第110 - 119页。415-425, 2015年。6
- [21]J. - F. Cardoso, “独立分量的高阶对比, Nent analysis, ”Neural Computation, vol. 11, no. 1, pp. 157-192, 1999年。6