Digital Signal Processing ••• (••••) ••••••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Robust algorithm for remote photoplethysmography in realistic conditions

Mikhail Artemyev *, Marina Churikova, Mikhail Grinenko, Olga Perepelkina

Neurodata Lab LLC, Miami, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Remote photoplethysmography
Facial imaging
Plane-Orthogonal-to-Skin
Vital signs monitoring
Remote photoplythesmography dataset

ABSTRACT

Over the last decade remote photoplethysmography (rPPG) algorithms have been developed extensively. As a result, pulse rate can now be accurately estimated by video data for still subjects. However, in realistic conditions both the accuracy of these algorithms and benchmark datasets are far from perfect. In this paper we propose a new rPPG method which enables heart rate detection by video from a standard webcam. The algorithm is robust with respect to such factors as illumination changes or the subject's movements, and can track fast pulse rate changes. To do that, the algorithm determines the approximate value of the pulse rate and then specifies it with high time resolution. In order to comprehensively study the proposed method, we collected a new dataset consisting of videos recorded in various challenging conditions of several categories as well as reference photoplethysmograms recorded synchronously with a contact pulse oximeter. The proposed method showed high performance under all conditions including blinking illumination, speech and large-amplitude movements. We tested two simplified versions of the algorithm, which provided competitive scores as well. However, with human movement videos the full method showed better results than its simplified versions ($p<0.001$). The proposed algorithm was tested on the existing UBFC-RPPG database and compared with previous methods. Our method showed high results (2.10 MAE, 3.43 RMSE).

1. Introduction

Photoplethysmography (PPG) is a simple optical method of skin light reflectance or transmission measurement that can be used for pulse detection. Contact PPG relies upon the fact that blood volume, blood vessel wall movement and the orientation of red blood cells affect the amount of light measured by the detector and, consequently, the photoplethysmographic signal. Thus, the usual PPG requires a light source to illuminate the skin and a photodetector to measure the light intensity changes. This technology is widely used in medicine, sports and other fields [1]. However, most of the popular PPG devices demand tight contact with the subject's skin, so the long-term use of such devices may cause discomfort. Furthermore, the PPG signal may be corrupted by motion artifacts, which makes this technology unsuitable for those cases when the subject is moving [1].

Remote PPG (rPPG), on the other hand, is used for non-contact obtaining of heart rate data. An important advantage is that rPPG requires only ambient light and a digital camera to acquire the person's vital signals, i.e., to detect skin color changes that reflect light absorption by hemoglobin [2]. Moreover, the camera may be placed at a certain distance from the subject, which is more convenient as compared with the contact PPG. Thus, rPPG can potentially be used to perform cardiovascular assessment in realistic settings. Taking into account all the above, this technology is well-suited for both business and everyday application [3]. The existing rPPG algorithms allow for performing an accurate heart rate monitoring based on videos recorded with common web-cameras, which makes them potentially useful for long-term health assessment. Nevertheless, most of the rPPG methods can only be applied to static subjects in laboratory surroundings with minimal changes [4]. This is done to eliminate various factors that may considerably affect the performance of rPPG, such as non-stationary conditions, human movements, illumination variations, high pulse value and high pulse variability [5]. Therefore, such algorithms have a very limited application in real conditions.

The main factor affecting the accuracy of rPPG is the change in face brightness. It may occur due to light changes that make RGB values change synchronously in all pixels (possibly by different amounts), or because of various head movements which cause significant changes in the illumination of some face parts as well as their appearance and disappearance in the video [6]. Other fac-

* Corresponding author.
E-mail address: m.artemyev@neurodatalab.com (M. Artemyev).

https://doi.org/10.1016/j.dsp.2020.102737
1051-2004/

tors contributing to signal artifacts include color temperature, the quantity and quality of light sources, and light direction [7,8].

If the subject is under physical stress, additional difficulties may arise. Firstly, the algorithms should be able to process a wide range of pulse frequencies, since the heart rate value in this case is 1.5-2 times higher than the resting one [9]. Secondly, while increasing during physical stress, the heart rate tends to gradually decrease to its standard values afterwards [10]. Such quick pulse changes are difficult to monitor in realistic settings. Therefore, there is a need for an improved method of heart rate monitoring that would minimize the influence of the said artifacts. We propose a novel video processing method for accurate pulse rate estimation in different naturalistic conditions, including various illumination types, physical stress and the subject's movements. To validate the method we collected a **Mo**tion and **Li**ght **p**hoto**p**lethysmo**g**raphy (MoLi-ppg) dataset using webcams and contact PPG data as a reference in different settings.

2. Related work

Since webcams are cheap and widely used, it is highly desirable for the algorithms to process such video data correctly. However, there is often artifact noise in webcam videos. It may be said that an accurate estimation of the pulse signal from such videos that involve head movements, speech, and illumination variations is one of the main challenges for rPPG methods.

A number of methods have been proposed to solve this problem. Most popular are different modifications of a region of interest (ROI) [6,11–14]. For example, Kumar et al. [11] presented a method that combined skin-color change signals from a number of patches of the face by using a weighted average with weights depending on blood perfusion and incident light intensity in the patches. Although this method can enhance the quality of the final pulse signal to some extent, there is still room for improvement when it comes to sudden facial changes in the video [11]. Tulyakov et al. [12] proposed a self-adaptive matrix completion approach which dynamically selected the most relevant face ROI for robust pulse estimation. The main drawback of this algorithm was a large number of hyperparameters to be tuned. Liu et al. [13] used self-adaptive signal separation to separate the noiseless block of facial region with a weight-based scheme. This noiseless signal containing vital information was used to obtain the holistic pulse signal, based on which the average pulse was computed by means of wavelet transform and data filter. The proposed method was shown to outperform the methods of Kumar et al. and Tulyakov et al. in realistic conditions [11–13].

Finally, Yang et al. [14] suggested a novel method similar to the one described above, which is a patch-based fusion framework for accurate pulse estimation in moving subjects. Wavelet time-frequency analysis is applied to a raw signal to select less contaminated patches. Next, a weighted fusion formula is constructed to get the final precise pulse signal based on frequency and gradient information. According to the authors, their method outperforms the methods of Liu et al., Kumar et al. and Tulyakov et al.

Now, another way to obtain a noiseless pulse signal is to process the color channels from the ROI. A common approach to extracting the rPPG data from the ROI is based on color channel combination. For this purpose blind source separation (BSS) may be applied. The approach was introduced by Poh et al. [15]. Two typical methods of BSS involve Independent Component Analysis (ICA) and Principal Component Analysis (PCA). ICA decomposes an RGB channel signal into component signals based on the assumption that the input signals corresponding to different sources are statistically independent [16], while PCA maximizes the variance of original points' projection onto components, whereby the source signals are assumed to be uncorrelated. The obtained signal may be used for pulse detection, but it still contains moving artifacts, which makes it difficult to analyze naturalistic data [17].

While BSS-based methods are suitable for demixing the signal into source signals for pulse extraction without any prior information, another group of approaches incorporates model-based methods that rely upon the knowledge of different components' color vectors in the demixing procedure. The model-based methods typically refer to methods based on the chrominance model (CHROM), methods using blood volume pulse signature (PBV) to distinguish pulse signals from motion distortions, and methods based on a "Plane-Orthogonal-to-Skin" (POS) [3]. Unakafov et al. compared different methods using the DEAP dataset [18], and the greatest assessment precision was achieved when pulse rate was estimated based on rPPG extracted with the POS method [19]. However, it should be noted that the limited number of movements in the DEAP dataset videos prevents its usage for testing the motion artifact removal methods. The POS algorithm was proposed by Wang et al. [20] as a novel method for remote pulse extraction from video. The method demixes the signal determined by light intensity, specular reflection and pulse-induced temporal color variations, based on optical/physiological considerations and specific characteristics of skin reflection. According to Wang et al., POS demonstrated the overall best performance among other methods, such as ICA, PCA, CHROM, and PBV. It was shown to be especially advantageous in fitness challenges where the skin-mask was noisy.

Thus, we have described two main groups of denoising approaches. The methods included in the first group are based on selecting a facial ROI from which a clean PPG signal can be extracted. The second group is formed by various methods of PPG signal filtering.

3. Proposed method

Here we provide a novel robust method for accurate pulse rate estimation from video in different conditions. In order to validate our method, we collected a MoLi-ppg dataset. The proposed algorithm (*cPR+fine*) implies evaluating the approximate pulse rate (central pulse rate, or cPR) with a robust but not precise algorithm and then estimating pulse rate values more accurately at each time point taking into account the central heart rate found.

We estimated the accuracy of the proposed algorithm (*cPR+fine*) in comparison with two of its simplified versions:

cPR: central pulse rate without refinement;

fine: pulse determination algorithm with no central pulse detection step.

Our algorithm is presented in Fig. 1.

3.1. Video stream preprocessing

The first step of our algorithm is to detect the face and ROI in the input video stream. The ROI includes the face except the eye and mouth areas, since these areas cause artifacts by moving. 68 facial landmarks [21] are detected using the dlib C++ library (dlib.net). This way, background does not affect the pulse rate estimation process.

Then we average the red (R), green (G), and blue (B) channels from the ROI in each frame. Further, we apply cubic interpolation to the R, G, B channels to resample our data to 256 frames per second (fps). This step is especially important if no face was detected in some frames (so the R, G, B channels had missing values before the interpolation) or if the input video stream had variable frame rate.

Please cite this article in press as: M. Artemyev et al., Robust algorithm for remote photoplethysmography in realistic conditions, Digit. Signal Process. (2020), https://doi.org/10.1016/j.dsp.2020.102737

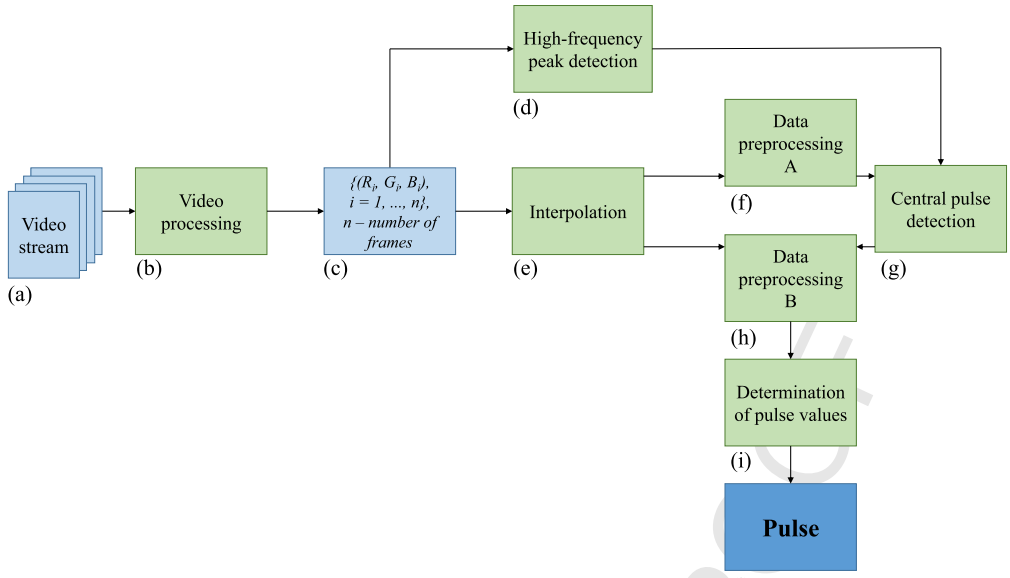


Fig. 1. Overview of the proposed method (*cPR+fine*). It involves four main steps: (1) video stream preprocessing, (2) extracting the high-frequency peak, (3) detecting the approximate pulse rate value (central pulse or cPR), and (4) determining pulse values at each time point. *cPR* method involves steps (1), (2), and (3). *fine* method involves steps (1) and (4).

3.2. Extracting the high frequency peak

This step of the proposed method is necessary to remove the light source artifacts, such as bulb flicker at the double alternating current (AC) frequency. This flicker can noticeably affect the spectrogram of the PPG signal as well as the final pulse rate estimation results. Our experiments show that if a video is captured in a room lit by bulbs connected to an AC network with a frequency of 50 Hz, then the RGB channels averaged over the ROI will contain a signal with a large amplitude and a frequency of $\frac{50\text{Hz}}{24} = 125\text{bpm}$, and this frequency may easily be mistaken for a pulse rate of any person in the video. To reduce the impact of light flicker, we filter out the frequencies that are likely to be corrupted.

The subharmonics at $\frac{50\text{Hz}}{2}$ and $\frac{50\text{Hz}}{3}$ usually have a higher amplitude, so it is much easier to detect them than the one at $\frac{50\text{Hz}}{24}$. That is why we first need to find a subharmonic of the high-frequency flicker that is easy to detect – let us call it the high-frequency peak (HFP).

In order to find the HFP, we first apply the POS method, then get the spectrogram of the POS output, and average its columns. In the averaged spectrum we look for a frequency $f_{HFP} > 7.0$ Hz with the highest amplitude. If the amplitude of f_{HFP} is at least twice as large as the median amplitude of the nearby frequencies, f_{HFP} is considered the HFP. Otherwise, we state that there is no HFP in the video.

3.3. Data filtering

After detecting the HFP, we apply some filtering procedures to the data.

First, we implement a bandpass filter which rejects frequencies beyond the 0.75-3 Hz range. The rPPG signal can then be extracted using any existing method. In the present work we obtained a single filtered signal with the POS method (see Section 2).

3.4. Central pulse detection

To make allowance for the pulse change, we select 36-second fragments with a 3-second step in the recording and then calculate the central pulse in each fragment. The central pulse rate calculation consists of the following steps:

1. The short-time Fourier transform (STFT) is applied, and the mean amplitude is determined for each frequency by averaging absolute values of the STFT result over the time domain. We use window length of 3840 samples (15 seconds) and hop size of 256 samples (1 second) in the STFT.
2. The resulting amplitude vectors are normalized.
3. The amplitudes expected to be unrelated to the pulse are subtracted from the normalized spectrum. These expected amplitudes are estimated using videos from the RAMAS database (see [22]). For each frequency f the expected amplitude is calculated by averaging over various videos with people whose pulse was different from f. This step appears necessary since the low-frequency part of the signal usually has a larger amplitude than the high-frequency part, so if the described step is skipped, the spectrum will have a negative trend and the heart rate estimation will be biased downward. But after the mean L2-normalized spectrum (presumably, unrelated to the pulse) is subtracted, the resulting spectrum has no well-defined trend. The L2-normalized spectrum is calculated by dividing all amplitudes by the root of the sum of the squares of all amplitudes in the spectrum.
4. If the HFP is detected in the record as described in Section 3.2, and it is a subharmonic of 50Hz (standard AC frequency in Europe), then it is necessary to filter out the amplitudes of the frequencies that could have been affected by the light flicker. We found out that 125 bpm is the only frequency that belongs to the (45 bpm, 180 bpm) interval and at the same time can be significantly affected by a 50Hz flicker. The amplitude of this frequency is replaced with the average amplitude of the neighbouring frequencies ($\pm$ 4 bpm neighbourhood in our case).
5. The frequency that has the maximum amplitude after preprocessing is considered to be the central frequency in the given 36-second fragment.

For each timestamp t of the initial recording, the central frequency is defined as the average of the central frequencies of all fragments containing t.

3.5. Pulse rate determination

The interpolated RGB signal is processed by a bandpass filter ($\pm$ 25 bpm from the "central pulse curve") and the POS algorithm. After that, spectrograms are plotted for the POS time series, and the most relevant pulse track is determined. Each track can be defined with a sequence of frequencies (f) and a sequence of their amplitudes (p). The quality of each track (f, p) is measured as (1), where T is the number of frames (video duration$\times$256), c_k is the central frequency at the timestamp k, $\alpha = 1$, $\beta = 25$, and $\gamma = 50$.

We shall call a track (f, p) locally optimal if for each $k \leq T$, (f_k, p_k) there is a local maximum of the spectrum at the moment k. The track (f, p) with the highest $\varphi(f, p)$ value is considered the most relevant of all locally optimal tracks and can be found with a dynamic programming algorithm. The sequence (f) of this track will be the output of the proposed method.

$$\varphi(f, p) = \alpha \sum_{k=1}^{T} (p_k) - \beta \sum_{k=1}^{T-1} \left| f_k - f_{k+1} \right| - \gamma \sum_{k=1}^{T} (f_k - c_k)^2 \quad (1)$$

The values of hyperparameters α, β, γ were selected on several laboratory samples, so that each of the three terms had approximately equal contributions in the formula (1). These samples were not included to the validation sets. We did not use hyperparameter grid search to avoid overfitting the validation dataset conditions. However, the variations of the proposed method with $\gamma = 0$ and with $\alpha = \beta = 0$ are described (see sections 3.6 and 3.7 respectively) and validated (see section 5).

3.6. Proposed method (cPR)

Video processing, HFP extraction and central pulse evaluation in the *cPR* method are performed the same way as in the *cPR+fine* method. The only difference is that in the *cPR* method the obtained central pulse rate is the end result of the algorithm.

3.7. Proposed method (fine)

With the *fine* method, video processing is performed the same way as with the *cPR+fine* method. Next, the pulse rate tracks are constructed (see Section 3.5). Since in this method the central pulse rate is not evaluated, the bandpass filtering of $\pm$ 25 bpm from the "central pulse curve" is substituted with a 0.75-3 Hz filtering, and as the third parameter in (1) is not defined in this case, the following constants are used: α = 1, β = 25, and γ = 0.

4. Algorithm comparison

We used two datasets to validate the proposed rPPG algorithm: our new dataset and the existing dataset. The first dataset is a **Mo**tion and **Li**ght **p**hoto**p**lethysmo**g**raphy (MoLi-ppg) dataset, and the second dataset is a UBFC-RPPG [23].

4.1. MoLi-ppg dataset

The MoLi-ppg dataset containing 229 video sequences was created. The videos were recorded with the following webcams: Logitech C920, Logitech C270, and an HD video camera Canon LEGRIA HFG40. The ground-truth contact PPG (cPPG) was obtained with a contact optical pulse sensor Shimmer3 GSR+ (http://www.shimmersensing.com) attached to the subject's finger (sampling rate = 256 Hz), and the data was synced with the video recording. The videos from the webcams were in uncompressed bitmap format with either a 800$\times$600 or a 1280$\times$720 pixel resolution, and 25 fps. The HD camera videos were in uncompressed bitmap format with a 1920x1080 pixel resolution and 50 fps. A total of 21 subjects aged 23-33 who identified as Caucasians took part in the experiments. Unless mentioned otherwise, the subject was lit by fluorescent ceiling lamps and sat in front of the cameras at a distance of about 1 m. We included four different categories to investigate them independently. These categories involved various close-to-natural conditions described below.

1. **Baseline.** We recorded the subjects in standard conditions while they were sitting naturally in front of the webcam.
2. **Illumination.** We defined two illumination settings, namely *bright* and *blinking*. The *bright light* setting included an additional spotlight directed at the subject's face while he or she was watching a six-minute video "Patti Smith Interview: Advice to the Young" (http://www.youtube.com/watch?v=L2EO3aXTWwg). Naturally, the light of the spotlight was brighter than that of the monitor. The *blinking light* setting involved the screen light falling on the subject's face without any additional spotlight. In this case the subjects had to watch a five-minute cartoon "Lifted" by Pixar, where the screen brightness changed rapidly during the movie, and also the video with Patti Smith.
3. **Movements.** We investigated three cases of head motion in standard conditions, which included *large* and *small* movements as well as *speech*. In the first two cases the subjects were instructed to perform various types of head movements: left-right, up-down and round. The amplitude of these movements (measured from the straight head position) had to be no more than 45 degrees in the task with *small* head movements and 80 degrees in the one with *large* head movements. As for the *speech* subcategory, the participants were asked to sit facing the cameras and read a text without any head movements.
4. **Recovery after physical stress.** To test the accuracy of our algorithm in relation to pulse trend detection, we recorded a series of videos that allowed us to analyze the heart rate recovery after the squat exercise. Each subject had to do 20-30 squats and sit down for recording immediately after that.

Informed consent for data collecting was obtained from each subject. Fig. 2 a-d shows the snapshots of some videos from the MoLi-ppg dataset. We intend to release the dataset for research purposes.

4.2. UBFC-RPPG dataset

The public dataset UBFC-RPPG [23] is used to verify the performance of our algorithm. The UBFC-RPPG is specifically designed for remote pulse rate measurement task. It contains 42 videos from 42 different subjects. The videos were recorded by a Logitech C920HD Pro camera with a resolution of 640x480 in an uncompressed 8-bit RGB format.

The participants were asked to play a time-sensitive mathematical game to keep their heart rates varied. The video records natural movements of subjects, including different motions.

4.3. Evaluation metrics and statistics

To evaluate the performance of our three pulse-determining rPPG algorithms we used the following metrics:

- **Mean Absolute Error (MAE)** in beats per minute (bpm) is calculated as the mean between the pulse obtained from

Table 1
Conditions and technical parameters of the MoLi-ppg dataset videos (PS - video with Patti Smith, C - cartoon).

Settings	Norm		Movements			Illumination		
	Baseline	Rec. after physical stress	Small	Large	Speech	Bright light	Blinking light-PS	light-C
N subjects	20 (13M, 7F)	11 (5M, 6F)	9 (4M, 5F)	11 (3M, 8F)	10 (2M, 8F)	10 (6M, 4F)		10 (6M, 4F)
Cameras (sync.)	Logitech C920	Logitech C920	Logitech C920			Canon LEGRIA HFG40		Logitech C920
			Logitech C270	Canon LEGRIA HFG40		Logitech C270		
Time	3 min	3 min	6 min	1 min	3-4 min	6 min	10 min	5 min
Lux	160	170	170	180	180	300	100	170

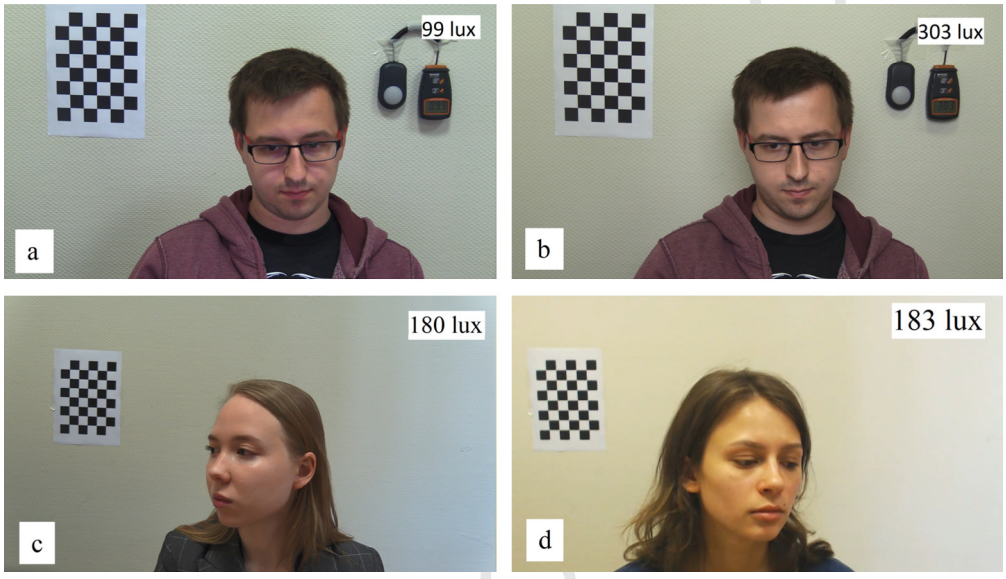


Fig. 2. Snapshots of the MoLi-ppg dataset videos. a - frame from video with "cartoon" settings (HD camera), b - frame from video with "light" settings (HD camera), c - frame from video with large head movements (HD camera), d - frame from video with large head movements (Logitech C920). Informed consent for publication was obtained from the subjects.

rPPG signals and the pulse obtained from cPPG signals with $\frac{\sum_{v\in\text{videos}}\sum_{k=1}^{T_v}|rPPG_{v,k}-cPPG_{v,k}|}{\sum_{v\in\text{videos}}T_v}$, where T_v is the number of frames in the video v.

- **Median Absolute Error (Median AE)** = = $\text{median}(\{|rPPG_{v,k} - cPPG_{v,k}| : v \in \text{videos}, 1 \le k \le T_v\})$.
- **Root mean square error (RMSE)** = $\sqrt{\frac{\sum_{v\in\text{videos}}\sum_{k=1}^{T_v}(rPPG_{v,k}-cPPG_{v,k})^2}{\sum_{v\in\text{videos}}T_v}}$.
- **Precision at 2.5 and 10 bpm.** This metric represents the percentage of estimations with the absolute error below the threshold (2.5 or 10 bpm).
- **Pearson correlation coefficient (r)** is the correlation between pulse estimated from the rPPG signal and the reference pulse estimated from the cPPG.

To compare the performance of algorithms in different conditions we constructed linear mixed models. ANOVA was used for the models testing. Since absolute errors were not normally distributed, as the depended variable logarithmic absolute errors were used.

5. Results

All recordings of MoLi-ppg and UBFC-RPPG datasets were processed by three pulse-determining rPPG algorithms: *fine, cPR* and *cPR+fine*. The overall absolute errors were calculated as the difference between the ground truth (cPPG) and the output of the rPPG algorithms for all videos.

5.1. Comparison of algorithms in different experimental conditions of MoLi-ppg dataset

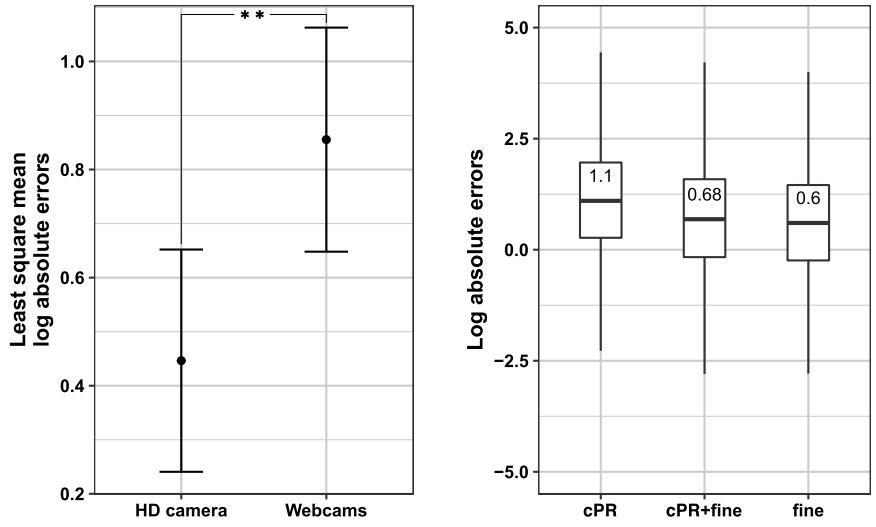
In each condition of MoLi-ppg dataset the videos were recorded synchronously on two out of three cameras (see Table 1). The data from the HD camera and the webcams were analyzed separately. The effect of different conditions was studied using the data from only one camera (the lower-quality webcam) in order to avoid video duplication. The videos with *baseline* conditions and *recovery after physical stress* were united into one category – *norm*, and analyzed together, since the illumination was the same and the subjects were sitting still facing the cameras in both cases.

Fig. 3a illustrates the effect of camera type. It can be seen that the webcams have significantly higher values of logarithmic absolute errors than the HD camera for the *cPR+fine* rPPG algorithm ($Sumsq = 15.066$, $meansq = 15.066$, $NumDF = 1$, $DenDF = 126.93$, $F = 9.8117$, $Pr(>F) = 0.002$). Fig. 3b shows the boxplots of logarithmic absolute errors for the three rPPG algorithms.

Table 2 represents the metrics on the MoLi-ppg and UBFC-RPPG datasets for these algorithms.

Table 2
Validation results on the UBFC-RPPG and MoLi-ppg databases. The best results are highlighted in bold.

algorithm	dataset	MAE ↓	median AE ↓	RMSE ↓	P@2.5 ↑	P@10 ↑
cPR	UBFC-RPPG	3.58	2.41	5.2	0.52	0.94
fine		2.23	1.37	3.83	**0.74**	0.97
cPR + fine		**2.10**	**1.36**	**3.43**	**0.74**	**0.98**
cPR	MoLi-ppg	6.98	3.0	**12.45**	0.44	0.81
fine	(all webcam samples)	**6.02**	**1.82**	13.21	**0.61**	**0.85**
cPR + fine		6.13	1.98	12.46	0.58	0.84
cPR	MoLi-ppg	5.64	2.93	**9.0**	0.45	**0.83**
fine	(webcam *movements*)	7.27	2.31	13.92	0.52	0.79
cPR + fine		**5.45**	**2.2**	9.48	**0.54**	**0.83**



(a) The effect of camera type (b) rPPG algorithms

Fig. 3. (a) Log absolute errors for the *cPR+fine* rPPG algorithm for two types of cameras (boxes represent the least square mean, error bars indicate the 95% confidence interval of the least square mean). (b) Boxplots of absolute error logarithms for three rPPG algorithms (*fine*, *cPR*, *cPR+fine*) in all experimental settings of the MoLi-ppg dataset.

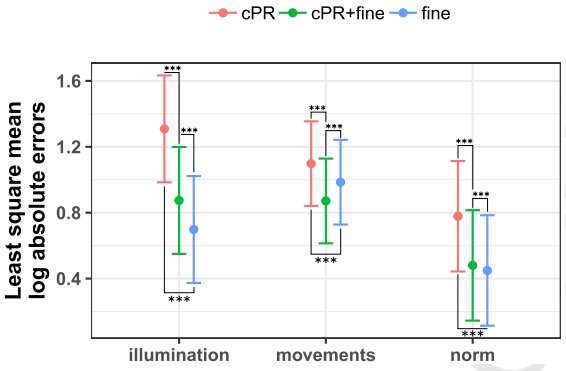


Fig. 4. Log absolute errors for rPPG algorithms in different conditions of the MoLi-ppg dataset (boxes represent the least square mean, error bars indicate the 95% confidence interval of the least square mean). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

A linear mixed model was constructed with logarithmic absolute error rate as the dependent variable, the algorithm type and condition type as predictors, and a random intercept for the video id. The analysis of this model (ANOVA) demonstrated that all predictors were significant ($F(2,4) = 43817.6$ for algorithms, $F(2,4) = 34273.7$ for conditions, $P \leq 0.001$), as well as their interaction ($F(2,4) = 2744.5$, $P \leq 0.001$).

It was found out that the *fine* algorithm was significantly more accurate than the *cPR+fine* algorithm in normal conditions with no subject movements and in baseline light conditions ($z = 5.8$, $p < 0.0001$). As for the central frequency algorithm (*cPR*), it determined the pulse significantly worse than the others (Fig. 4).

Under different illumination conditions the *cPR+fine* algorithm showed much better results than the average frequency *cPR* algorithm ($z = 134.9$, $p < 0.0001$), but its performance was considerably poorer as compared with the *fine* algorithm with no central frequency evaluation ($z = 54.7$, $p < 0.0001$) (Fig. 4).

With moving subjects, the *cPR+fine* method was far more accurate than the other two (*cPR*: $z = 44.4$, $p < 0.0001$; *fine*: $z = -22.3$, $p < 0.0001$).

There were no significant differences between the results obtained with the *cPR+fine* algorithm under various conditions (Fig. 5). With the *cPR* algorithm, however, the differences between the normal condition and the conditions with different illumination types were significant (the performance was better under the normal one). As for the *fine* algorithm, significant were the differences between the normal condition and the conditions that involved movements (the latter were more challenging).

A more detailed comparison of the algorithms' accuracy in various condition subcategories is displayed in Fig. 6. After physical stress the heart rate has a downward trend, decreasing to its initial level. With videos recorded in *baseline* conditions (i.e., not after the exercise), the *fine* algorithm outperformed the others. In conditions *after physical stress* the *cPR* algorithm was the least effective, while no significant differences were observed between *fine* and *cPR+fine*. The *cPR+fine* algorithm was significantly better than the

Please cite this article in press as: M. Artemyev et al., Robust algorithm for remote photoplethysmography in realistic conditions, Digit. Signal Process. (2020), https://doi.org/10.1016/j.dsp.2020.102737

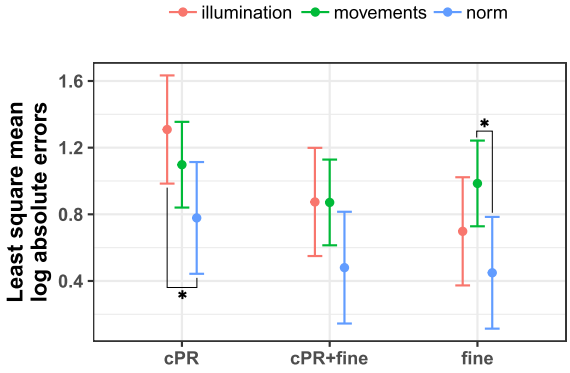


Fig. 5. Log absolute errors for the algorithms in different conditions of MoLi-ppg dataset (boxes represent the least square mean, error bars indicate the 95% confidence interval of the least square mean).

Table 3
Result metrics of different rPPG methods on the UBFC-RPPG dataset (CK - chrominance based on kernel density independent component analysis [24,23]). The best results are highlighted in bold.

Method	MAE	RMSE	r	P@2.5
ICA [24]	3.507	8.635	0.908	–
CHROM [24]	3.435	4.614	0.968	–
POS [24]	2.436	6.608	0.936	–
CK [24]	2.292	3.803	0.981	–
TSP + CHROM [23]	–	**2.388**	0.961	**0.826**
cPR + fine	**2.10**	3.43	**–**	0.74

rest in conditions with all types of movements, while the *fine* algorithm was the most effective one with regard to *bright* and *blinking light*.

5.2. Comparison of algorithms on UBFC-RPPG dataset

To evaluate the performance of the proposed method *cPR+fine*, we compare metrics of our algorithm with some other rPPG algorithms on UBFC-RPPG dataset (see Table 3). We take as state of the art paper by Bobbia et al. [23], and also we compare results from a recent article by Song et al. [24]. Our method outperforms all algorithms except one by Bobbia et al., that is TSP - Temporal SuperPixel chrominance-based method (TSP + CHROM) [23]. It is necessary to take it into account for further development.

6. Discussions

The *cPR+fine* algorithm works best on subjects who were exposed to physical stress prior to examination *after physical stress*, and with all types of movements. Movements make remote heart rate monitoring more challenging [6,7], so our proposed method is well suited for use in real life conditions when a user is sitting still. The most suitable conditions for our system is when a person sits still, unobstructed by large movements or blinking light. The most difficult conditions are the recordings that feature large head movements and speech that may be caused by a shift of the tracked ROI. In such situations, landmarks corresponding to a half of the face get lost and the tracked ROI location may be erroneous, thus degrading the accuracy of HR measurement. In our future work, the ROI tracking could be improved to tolerate more significant head movements.

However, our algorithm has some limitations. One of them is the distance between the camera and the person as we tested our method at 1-1.5 meters distance only. Another limitation of this rPPG method is the need for adequate lighting. Low accuracy was demonstrated in conditions with blinking light. This case is challenging for remote PPG methods, which suggests a hypothesis that the lighting level plays an important role in correct pulse detection [7,8]. Moreover, we tested the algorithm using a webcams and Canon HD camera only, while we are going to test it in on built-in smartphone cameras for business applications.

The proposed pulse rate estimator achieved high accuracy when tested on the UBFC-RPPG database (Table 3). This is probably due to the fact that the UBFC-RPPG dataset was recorded in relatively simplified laboratory conditions [23]. This dataset was recorded using a low cost webcam; the subjects were required to play a mathematical game that emulates the scenario of a normal user activity in front of a computer. The authors proposed the TSP+CHROM method that showed the best RMSE and P@2.5 results. This approach was tested on UBFC-RPPG dataset only, therefore, it is unknown how said approach will work on more heterogeneous videos such as those in MoLi-ppg dataset. In our future work, we intend to test the basic algorithms such as ICA, POS, and CHROM on MoLi-ppg dataset since it was recorded in more complex and challenging conditions than UBFC-RPPG dataset. We intend do that to provide baseline accuracy for out dataset and compare it with results received on UBFC-RPPG to test this hypothesis.

We intend to further investigate this issue in more detail, including testing our algorithm in various more complex conditions such as different distance between the camera and the subject, various color temperature, quantity, quality, and position of light sources. Furthermore, in addition to external conditions, in the future it is necessary to study the quality of our methods work on people of different races and ages. Our robust algorithm for the remote heart rate detection may be used in various healthcare and sport application for monitoring cardiac activity and potential prevention of cardiovascular disorders.

7. Conclusions

In this paper, we proposed a new MoLi-ppg dataset recorded in a variety complex conditions that is useful for the development of remote PPG algorithms. The current rPPG method combines the advantages of detection of central pulse with unit of pulse rate determination. The algorithm was proven to be robust with respect to such conditions as blinking light and subjects' movements and speech in MoLi-ppg dataset. Moreover, our method shows high results comparable to the state-of-the-art results on public UBFC-RPPG database (2.10 MAE, 3.43 RMSE). This robustness is guaranteed by the central pulse rate evaluation procedure that allows for estimation of approximate pulse rate in long video fragments. At the same time, high precision and time-domain resolution, which are essential for handling high pulse rate variability, are obtained through a refinement procedure. The full algorithm that involves both of these steps demonstrates significant improvement (that is, the decrease of logarithmic absolute error rate) in processing videos of moving subjects as compared to the algorithm where each of these two steps applied separately.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dsp.2020.102737.

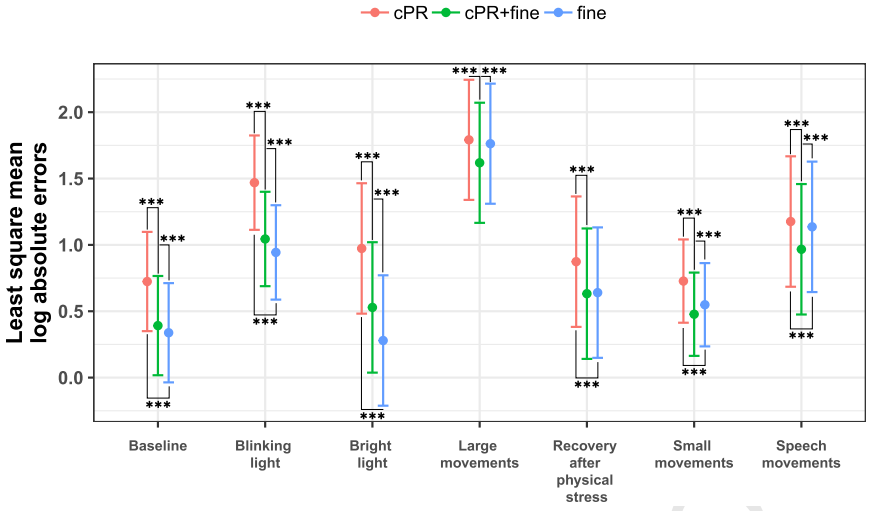


Fig. 6. Log absolute errors for the algorithms in all subcategories of the MoLi-ppg dataset (boxes represent the least square mean, error bars indicate the 95% confidence interval of the least square mean).

References

[1] J. Allen, Photoplethysmography and its application in clinical physiological measurement, Physiol. Meas. 28 (3) (2007) R1.
[2] L. Feng, L.-M. Po, X. Xu, Y. Li, C.-H. Cheung, K.-W. Cheung, F. Yuan, Dynamic roi based on k-means for remote photoplethysmography, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2015, pp. 1310–1314.
[3] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, Z.J. Wang, Video-based heart rate measurement: recent advances and future prospects, IEEE Trans. Instrum. Meas. (2018).
[4] Y. Sun, S. Hu, V. Azorin Peris, S. Greenwald, J. Chambers, Y. Zhu, Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise, J. Biomed. Opt. 16 (2011) 077010, https://doi.org/10.1117/1.3602852.
[5] S. Zaunseder, A. Trumpp, D. Wedekind, H. Malberg, Cardiovascular assessment by imaging photoplethysmography–a review, Biomed. Eng./Biomed. Technik 63 (5) (2018) 617–634.
[6] X. Li, J. Chen, G. Zhao, M. Pietikainen, Remote heart rate measurement from face videos under realistic situations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4264–4271.
[7] R. Amelard, C. Scharfenberger, F. Kazemzadeh, K.J. Pfisterer, B.S. Lin, D.A. Clausi, A. Wong, Feasibility of long-distance heart rate monitoring using transmittance photoplethysmographic imaging (ppgi), Sci. Rep. 5 (2015) 14637.
[8] J. Przybyło, E. Kańtoch, M. Jabłoński, P. Augustyniak, Distant measurement of plethysmographic signal in various lighting conditions using configurable frame-rate camera, Metrol. Meas. Syst. 23 (4) (2016) 579–592.
[9] M.P. Tulppo, T.H. Makikallio, T. Seppaänen, R.T. Laukkanen, H.V. Huikuri, Vagal modulation of heart rate during exercise: effects of age and physical fitness, Am. J. Physiol., Heart Circ. Physiol. 274 (2) (1998) H424–H429.
[10] M. Javorka, I. Zila, T. Balharek, K. Javorka, Heart rate recovery after exercise: relations to heart rate variability and complexity, Braz. J. Med. Biol. Res. 35 (8) (2002) 991–1000.
[11] M. Kumar, A. Veeraraghavan, A. Sabharwal, Distanceppg: robust non-contact vital signs monitoring using a camera, Biomed. Opt. Express 6 (5) (2015) 1565–1588.
[12] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J.F. Cohn, N. Sebe, Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2396–2404.
[13] X. Liu, X. Yang, J. Jin, J. Li, Self-adaptive signal separation for non-contact heart rate estimation from facial video in realistic environments, Physiol. Meas. 39 (06) (2018), https://doi.org/10.1088/1361-6579/aaca83.
[14] Z. Yang, X. Yang, J. Jin, X. Wu, Motion-resistant heart rate measurement from face videos using patch-based fusion, Signal Image Video Process. (2019) 1–8.
[15] M.-Z. Poh, D.J. McDuff, R.W. Picard, Non-contact, automated cardiac pulse measurements using video imaging and blind source separation, Opt. Express 18 (10) (2010) 10762–10774.
[16] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (4–5) (2000) 411–430.
[17] M. Lewandowska, J. Rumiński, T. Kocejko, J. Nowak, Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity, in: 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2011, pp. 405–410.
[18] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: a database for emotion analysis; using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2011) 18–31.
[19] A.M. Unakafov, Pulse rate estimation using imaging photoplethysmography: generic framework and comparison of methods on a publicly available dataset, Biomed. Phys. Eng. Express 4 (4) (2018) 045001.
[20] W. Wang, A.C. den Brinker, S. Stuijk, G. de Haan, Algorithmic principles of remote ppg, IEEE Trans. Biomed. Eng. 64 (7) (2016) 1479–1491.
[21] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 faces in-the-wild challenge: the first facial landmark localization challenge, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 397–403.
[22] O. Perepelkina, E. Kazimirova, M. Konstantinova, Ramas: Russian multimodal corpus of dyadic interaction for affective computing, in: International Conference on Speech and Computer, Springer, 2018, pp. 501–510.
[23] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognit. Lett. (2017).
[24] R. Song, S. Zhang, J. Cheng, C. Li, X. Chen, New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method, Comput. Biol. Med. 116 (2019) 103535.

Mikhail Artemyev was born in Moscow, Russia, on January 23, 1995. He received the B.S. degree in Mathematics, and the M.S. degree in Applied Mathematics from Higher School of Economics, Moscow, Russia in 2016 and 2018, respectively. He is a machine learning specialist at Technical Department of Neurodata Lab LLC. His research interests include machine learning, deep learning, computer vision, and audio signal processing.

Marina Churikova was born in Moscow, Russia, on April 6, 1994. She received the B.S. degree in Biology in 2016, and the M.S. degree in Neurobiology in 2018 from Lomonosov Moscow State University, Moscow, Russia. She is a research scientist at Research & Development Department of Neurodata Lab LLC. Her research interests include signal processing, machine learning and their applications in physiology, health care technologies and cognitive neuroscience.

Mikhail Grinenko is a scientific consultant at Research & Development Department of Neurodata Lab LLC. He obtained Ph.D. degree and then Doctor of Physical and Mathematical Sciences degree at Steklov Mathematical Institute of Russian Academy of Sciences. M. Grinenko has his main research interests in various domains of Fundamental Mathematics, Computational Mathematics, Machine Learning, etc.

Olga Perepelkina is a Chief research scientist at Research & Development Department of Neurodata Lab LLC and a Ph.D. candidate at the Neuro- and Pathopsychology Department of Lomonosov Moscow State

University. Her research interests include affective computing, multisensory processing, machine learning and computational methods in cognitive neuroscience and clinical psychology. Perepelkina received a M.S. degree in Clinical psychology from Lomonosov Moscow State University. She has received a nomination "Fundamental research" in the Graduate research competition of Lomonosov Moscow State University. She is a member of International Organization of Psychophysiology and European Society for Cognitive and Affective Neuroscience (ESCAN).

Please cite this article in press as: M. Artemyev et al., Robust algorithm for remote photoplethysmography in realistic conditions, Digit. Signal Process. (2020), https://doi.org/10.1016/j.dsp.2020.102737