

Video-Based Pulse Rate Variability Measurement Using Periodic Variance Maximization and Adaptive Two-Window Peak Detection

Peixi Li, Yannick Benezeth, Richard Macwan, Keisuke Nakamura, Randy Gomez, Chao Li, Fan Yang

▶ **To cite this version:**

Peixi Li, Yannick Benezeth, Richard Macwan, Keisuke Nakamura, Randy Gomez, et al.. Video-Based Pulse Rate Variability Measurement Using Periodic Variance Maximization and Adaptive Two-Window Peak Detection. Sensors, 2020, 20 (10), pp.2752. 10.3390/s20102752 . hal-02740220

HAL Id: hal-02740220

https://u-bourgogne.hal.science/hal-02740220

Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

sensors

MDPI

Article

Video-Based Pulse Rate Variability Measurement Using Periodic Variance Maximization and Adaptive Two-Window Peak Detection

Peixi Li [1,*], Yannick Benezeth [1], Richard Macwan [2], Keisuke Nakamura [3], Randy Gomez [3] and Chao Li [4] and Fan Yang [1]

1 ImViA-EA7535, Univ. Bourgogne Franche-Comté, 21000 Dijon, France; yannick.benezeth@u-bourgogne.fr (Y.B.); fanyang@u-bourgogne.fr (F.Y.)
2 LPL, CNRS-UMR7538, Univ. Paris 13, 93430 Villetaneuse, France; Richard.macwan@univ-paris13.fr
3 Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako-shi, Saitama 351-0114, Japan; keisuke@jp.honda-ri.com (K.N.); r.gomez@jp.honda-ri.com (R.G.)
4 State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; lichao8601@hotmail.com
* Correspondence: Peixi_Li@etu.u-bourgogne.fr

Received: 14 March 2020; Accepted: 8 May 2020; Published: 12 May 2020



Abstract: Many previous studies have shown that the remote photoplethysmography (rPPG) can measure the Heart Rate (HR) signal with very high accuracy. The remote measurement of the Pulse Rate Variability (PRV) signal is also possible, but this is much more complicated because it is then necessary to detect the peaks on the temporal rPPG signal, which is usually quite noisy and has a lower temporal resolution than PPG signals obtained by contact equipment. Since the PRV signal is vital for various applications such as remote recognition of stress and emotion, the improvement of PRV measurement by rPPG is a critical task. Contact based PRV measurement has already been investigated, but the research on remotely measured PRV is very limited. In this paper, we propose to use the Periodic Variance Maximization (PVM) method to extract the rPPG signal and event-related Two-Window algorithm to improve the peak detection for PRV measurement. We have made several contributions. Firstly, we show that the newly proposed PVM method and Two-Window algorithm can be used for PRV measurement in the non-contact scenario. Secondly, we propose a method to adaptively determine the parameters of the Two-Window method. Thirdly, we compare the algorithm with other attempts for improving the non-contact PRV measurement such as the Slope Sum Function (SSF) method and the Local Maximum method. We calculated several features and compared the accuracy based on the ground truth provided by contact equipment. Our experiments showed that this algorithm performed the best of all the algorithms.

Keywords: remote photoplethysmography (rPPG); video analysis; pulse rate variability (PRV); periodic variance maximization (PVM); peak detection

1. Introduction

The contact photoplethysmography (PPG) devices used by hospitals and research labs are small photoelectric sensors that measure the blood volume pulse (BVP) through fingers or ears by detecting the change of light intensity which passes through the tissue. This technique is widely used in medical applications for its benefits of being low cost and high convenience compared with an electrocardiogram (ECG). The remote photoplethysmography (rPPG) has recently attracted significant attention because it is a non-contact method to measure physiological parameters and can be used in long-term and non-interruptive monitoring. The basic principle behind rPPG derives from that of

contact PPG where instead of using a photoelectric sensor, rPPG uses an RGB camera. The variation in the light intensity reflected by the human face can be used to measure BVP, which in turn can be used to estimate physiological parameters such as Heart Rate (HR) and Pulse Rate Variability (PRV). Heart Rate Variability (HRV) is defined as the variation of the inter-beat intervals measured by the distance of the R-peaks in the ECG signal [1]. HRV has been studied intensively by biomedical researchers as a sign of cardiac health [2] and Autonomic Nervous System (ANS) [3] which can be used to reflect human stress and emotion [4]. According to the previous research, the Pulse Rate Variability (PRV) measured by PPG techniques can be a surrogate measurement of HRV in some conditions [5]. And Poh et al. showed that high degrees of agreement were achieved between the PRV measured by a contact sensor and a Webcam [6]. This means that PRV measured by rPPG has great potential for multimedia applications such as remote assessment of pain, stress and emotion, although it cannot replace HRV in critical medical analysis due to the low frame rate, the noise and errors of rPPG method and completely different experimental conditions.

According to existing studies, the accuracy of remote HR measurement can be higher than 90% [7]. However, PRV measurement with rPPG is more complicated, because it requires precise peak detection of the BVP signal, which is prone to noise emanating from sensors, light illumination variation, movements, and so forth. Although some methods exist to improve the BVP peak detection and PRV measurement for contact PPG [8–10], most of them have not been applied in rPPG applications. Some works used complex equipment such as thermal cameras and 5-band cameras to improve the performance of remote PRV measurement [11,12], but such equipment is either too expensive or not widely available in everyday life. Another issue is that the existing remote PRV research works usually assess the accuracy of measured PRV through stress detection or emotion recognition instead of evaluating the performance with respect to the ground truth PRV signal obtained from a contact sensor. For instance, Macduff et al. [13] and Mitsuhashi et al. [14] used rPPG methods for stress/emotion recognition where the results showed an accuracy of 60%–85% for emotion prediction via remote PRV. But the errors between the remotely measured PRV and the ground truth were not studied and discussed. None of the work which adopted the low-cost RGB camera has been able to recognize the emotion states with the accuracy higher than 90%. Therefore, improving the remote PRV measurement and assessing it with quantitative ground truth is an important task.

In this paper, we use the dataset generated by a low-cost 3-band web camera and adopted the newly proposed Periodic Variance Maximization (PVM) method [15] to extract the BVP signal. Then, we used an event-related Two-Window [16] algorithm for BVP peak detection to increase the accuracy of PRV measurement via rPPG. To dynamically adapt to the observed data, we propose a method to adaptively set the parameters for the Two-Window peak detection. We compare our algorithm with other peak detection methods using several assessment metrics and show that this method performs the best of all. Section 2 describes related work, followed by proposed method in Section 3. Sections 4 and 5 describe the experiments and results, followed by the conclusion in Section 6.

2. Related Works

In this section, the related works are described in three research directions: (1) Stress and emotion detection with rPPG framework; (2) Remote PRV measurement improvement with novel cameras; (3) Improvement of BVP peak detection and HRV/PRV measurement with contact equipment.

2.1. Stress and Emotion Detection with rPPG Framework

As explained, the rPPG can be used for remote detection of stress and emotion. Macduff et al. [13] proposed a framework that adopted facial landmarks to get the region of interest (ROI) from the face, and used ICA to process and combine the RGB signals to get the BVP signal, and then obtained PRV based on the extracted BVP signal. They then utilized several frequency features of PRV such as High Frequency (HF) power, Low Frequency (LF) power and LF/HF to train the model on the dataset which included two emotion states, relaxed and stressed states. It turned out that PRV alone could achieve an

accuracy of 70% to distinguish between these two states. Mitsuhashi et al. [14] proposed a method that combined the two-layer (melanin and hemoglobin layers) model and singular valued decomposition in RGB space for pulse signal measurement from videos. They defined three different stress levels by the difficulty of the tasks. For instance, easy (mental arithmetic of 5*6), middle (mental arithmetic of 13*16), and difficult (mental arithmetic of 114*123). They used the KNN method to classify the stress modes. The results showed that the accuracy of the classification is around 66%–83% for different stress levels. Belaiche et al. [17] used both micro-expressions and PRV to predict three emotion states, namely, happiness, disgust and anger. This study found that although the accuracy of the PRV based method is higher than the micro-expressions based method in emotion states recognition but the average accuracy is usually not higher than 60% for the dataset that contains sudden emotion change.

2.2. Improving the Remote PRV Measurement with Novel Cameras.

An approach to get more precise PRV measurement is to make use of novel experimental equipment and materials to achieve better performance. Gupta et al. [18] proposed a system that used a thermal camera, a monochrome camera with a color filter and an RGB camera to extract the BVP signal. This novel system was proved quite effective to reduce the noise caused by motion and light illumination variation. With this system, one can monitor the HR and PRV and visualize the data in a real-time scenario. The RGB cameras are the most widely used cameras available in everyday life, however, some researchers suggested that the 5-band RGBCO cameras could work better since more information could be used in the ICA and PCA methods to combine the signals. Mcduff et al. [11] presented a work that adopted the novel 5-band camera and found that the cyan, green, and orange (CGO) bands performed better than RGB bands in measuring the PRV in the frequency domain. The correlation between the signals measured by the contact sensor and the camera was over 90%. Their further work [12] utilized the CGO bands and conducted an experiment which included two randomized-order tasks. The results demonstrated that the PRV features extracted by CGO bands could distinguish between the relaxed and stressed modes with an accuracy of 70%–80%. Additionally, they showed that the value of PRV features could capture the changes of stress for individuals since the two tasks in the experiments caused different stress levels for the participants and they were detected with correlated value change. In the realistic applications, the remote PRV measurement may suffer from missing observations caused by subject movement and subject getting obscured by an object, they addressed the issue by proposing an algorithm to fuse partial camera signals generated from an array of cameras and they improved the PRV measurement in the scenarios where significant amount of data is lost [19]. Fukunishi et al. [20] proposed to use a new ROI detection, a new rule-based method and a new filtering algorithm to improve the performance of the 5-band camera in PRV measurement and they successfully reduced the noise of PRV features in the frequency domain.

2.3. Improvement of BVP Peak Detection and HRV/PRV Measurement with Contact Equipment

The remote measurement of PRV can be improved with novel algorithms for BVP peak detection. This improvement has been the focus of biomedical research, especially on signals generated by Electrocardiography (ECG) and contact PPG sensor. Béres et al. [21] gave a comprehensive study of the adequate sampling frequency for contact PPG measurement and the detailed instructions for interpolation and sampling. Zong et al. [22] proposed to use Slope Sum Function (SSF) to enhance the rising part of the ECG signal and reduce the falling part so that the shape of the signal is simplified and more clear to conduct peak detection. Jang et al. [9] adopted this method and used it in the contact PPG. It turned out the SSF could work well on contact PPG signal with some pre-processing and post-processing methods. Computer vision researchers Li et al. [23] further showed that this algorithm could improve the performance of remote PRV measurement despite completely different experimental conditions in the non-contact scenario. The contact PPG measurement under tropical conditions is difficult especially after exercise. Elgendi et al. [16] used an event-related method to solve this problem. This algorithm utilized the property of the PPG signal that the average height of the

systolic peak period should be higher than the beat period and the systolic peaks are the highest points inside the peak periods. Thus the method detected the BVP peaks and addressed the non-stationary effects caused by severe exercise conditions in hot and humid environment. The results showed that this method detected the peaks with the sensitivity of nearly 100%. However, the parameters of this algorithm have to be optimized by a brute force search, so it is very time-consuming.

2.4. Summary of the Related Works

According to these related works, it can be concluded that: (1) the basic framework of rPPG measurement has been intensively studied and widely adopted. (2) Some computer vision researchers used more complex cameras to improve the rPPG performance, however, the equipment is either expensive or not widely available in everyday life. (3) The majority of the rPPG works do not focus on improving BVP peak detection, which is critical for PRV measurement. (4) Biomedical and signal processing researchers used some novel algorithms to improve the contact PPG measurement, and some of the algorithms have not been adopted by computer vision researchers, possibly due to completely different experimental conditions. (5) Some existing algorithms [16] for peak detection are time-consuming.

In this paper, we adopt a dataset that was recorded by a widely available low-cost camera and utilized the algorithms that could be specifically used in the BVP extraction and peak detection in remote conditions and we show that our framework indeed improved the precision of remotely measured PRV.

3. Method

3.1. RPPG Signal Extraction with PVM Method

Periodic Variance Maximization (PVM) with Generalized Eigenvalue Decomposition (GEVD) [15] was adopted to process RGB signals and obtain the BVP signal. This algorithm combines PCA with periodicity maximization to extract the quasi-periodic component with unknown period. PVM uses generalized eigenvalue decomposition to obtain a periodicity maximizing basis at a given frequency. This process is then iterated over the human heart rate range to obtain the frequency exhibiting the highest global periodicity. The GEVD step is applied on the pair of covariance and lagged covariance matrices which encapsulate the idea of periodicity. Intuitively, a periodic signal shall exhibit high similarity to its lagged version, if this lag is close to its effective period. This intuition is quantified using a periodicity metric, ρ, described below. Let $\mathbf{x}$(i) be the temporal RGB signal at time i after centering and detrending. The covariance matrices and lagged covariance matrices are defined as:

$$\mathbf{C_x} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(i)\mathbf{x}(i)^T, \mathbf{P_x} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(i)\mathbf{x}(i+\tau)^T, \quad (1)$$

respectively, where $\mathbf{x}(i+\tau)$ is the signal that lagged by τ seconds. Then GEVD is applied on $\mathbf{P_x}$ and $\mathbf{C_x}$ and the generalized eigenvector corresponding to the highest generalized eigenvalue $\mathbf{w}$ is used to obtain the signal with highest periodic variance at a given lag τ using:

$$y(i) = \mathbf{w}^T\mathbf{x}(i). \quad (2)$$

In practice, the GEVD is applied on symmetrized versions of $\mathbf{P_x}$ and $\mathbf{C_x}$ to ensure positive generalized eigenvalues, using $(\mathbf{P_x} + \mathbf{P_x^T})/2$, which effectively represents the two-way variance among the RGB channels. GEVD aims to estimate the matrix of generalized eigenvectors $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$ and the generalized eigenvalues $\mathbf{D} = diag(\lambda_1, \lambda_2, \lambda_3)$ that satisfy:

$$\mathbf{W}^T\mathbf{P_x}\mathbf{W} = \mathbf{D}, \mathbf{W}^T\mathbf{C_x}\mathbf{W} = \mathbf{I}. \quad (3)$$

Denoting $\mathbf{w}$ as the generalized eigenvector corresponding to the highest generalized eigenvalue in $\mathbf{D}$, the extent of periodic information in $y(i)$ can be quantified as:

$$\rho(\tau, \mathbf{w}) = \frac{\sum_{i=1}^{N} y(i)y(i+\tau)}{\sum_{i=1}^{N} y(i)^2} = \frac{\mathbf{w}^T \mathbf{P}_\mathbf{x} \mathbf{w}}{\mathbf{w}^T \mathbf{C}_\mathbf{x} \mathbf{w}}, \tag{4}$$

where ρ is the periodicity metric. If the component signals were fully periodic, ρ would be 1 given that $\mathbf{C}_\mathbf{x}$ and $\mathbf{P}_\mathbf{x}$ were equal. This periodicity metric is then optimized over the human heart rate range to obtain τ^*, the period corresponding to the BVP, where at each τ the optimum weighting matrix $\mathbf{w}$ is obtained by GEVD.

We have extensively evaluated this algorithm and compared it with other methods such as Green, PCA, ICA and Chrominace and the results showed that this algorithm performed the best for extracting the rPPG signal [15].

3.2. Adaptive Two-Window Peak Detection

The BVP signal is considered to have two important time windows. The first one is the "beat period", which is the entire period of one heart beat. And the other is the "systolic peak period", which is the period where a systolic peak appears. A systolic peak period is inside a beat period, so the time length of the systolic peak period is smaller than the beat period. The two periods have two important physical properties: firstly, the average signal amplitude of the systolic peak period is usually higher than the average amplitude of the beat period, and secondly, the systolic peaks are supposed to be the highest points within the systolic peak periods. With the definitions of the two periods and the properties, an event-related algorithm was proposed by Elendi et al. [16] to reduce the noise of systolic peak detection with contact PPG in tropical conditions after exercise. We consider this method useful to reduce the noise caused by the light illumination variation, movement, sensors, and so forth, in the rPPG method.

However, there are two disadvantages of the original event-related method: (1) It adopted too many parameters. (2) Brute force search was used to set all the parameters, namely the frequency band, the duration of the beat and peak periods, and the offset between the peak area's amplitude and beat area's amplitude. Therefore the method is very time-consuming and not practical in realistic applications. To address this issue, we propose a fast version of the event-related algorithm that uses two adaptively determined parameters and this algorithm was named as "Adaptive Two-Window Peak Detection":

Let W_b be the window size of the "beat period", and the moving average of beat period is defined as:

$$MA_b(i) = \frac{1}{W_b}\Big(y(i - W_b/2) + ... + y(i) + ... + y(i + W_b/2)\Big), \tag{5}$$

where $y(i)$ is the BVP value at time i. W_b can be the time length of the window, or the number of time stamps of the window if the signal is discrete. Similarly, let the W_p be the window size of "systolic peak period", and the moving average of the systolic peak period is defined as:

$$MA_p(i) = \frac{1}{W_p}\Big(y(i - W_p/2) + ... + y(i) + ... + y(i + W_p/2)\Big) \tag{6}$$

and two thresholds are defined as:

$$THR_1 = MA_b(i). \tag{7}$$

$$THR_2 = |\overline{W}_p|, \tag{8}$$

where $|\overline{W}_p|$ is the average window size of the peak periods within a certain range. As discussed, the average signal amplitude of peak period is usually higher than that of beat period and the systolic

peak is usually the highest value within the peak period. So the systolic peaks can be detected with such conditions:

- $y(i_1), y(i_2), y(i_3)$ ... are considered as the block of interest if $MA_p(i_N)$ is larger than THR_1
- The block of interest is discarded, if the width of the block is smaller than THR_2. The THR_2 is calculated as $\left(W_p(i_1) + (W_p(i_2) + (W_p(i_3) + ... + (W_p(i_N)\right)/N$.
- The peaks are the maximum values in the blocks of interest.

Since the BVP signal has been very well filtered by PVM method, the only parameters that have to be determined are the window sizes of the beat periods and peak periods. We applied Fast Fourier Transform (FFT) on the BVP signal over a 10 s' window and get the frequency $F_b(i)$ for each point in time and then the $W_b(i)$ is calculated as $1/F_b(i)$. For the detrended BVP signal, the peaks usually appear in the positive part that is approximately half of the signal. The peak period should be within the beat period, therefore we can set the window size of the peak period as half of the positive part of the beat period which is 0.25 times of the beat period so the value of $W_p(i)$ is $0.25 \times W_b(i)$ for each point.

Figure 1 shows an example of the peak detection of the BVP signal with the Two-Window method. The black curve is the BVP signal generated from the MMSE dataset [24]. The blue curve is the moving average of the systolic peak area (MA_p). The green curve is the threshold (THR_1). It can be seen that in this case, the false peaks between the first and second beat periods are eliminated because the MA_p is not higher than the THR_1. So this part is considered noise.

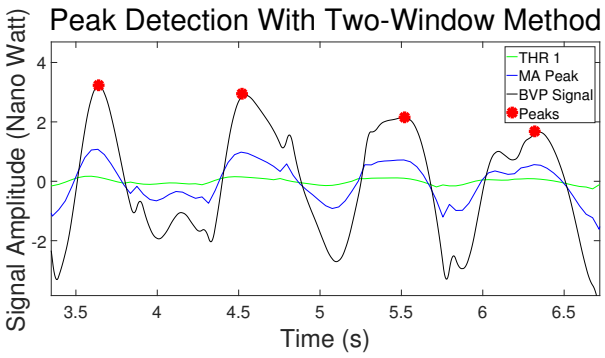


Figure 1. An example of peak detection with Two-Window method.

4. Experiments

The new algorithm was tested and compared with the state-of-the-art peak detection methods in the framework of remote PRV measurement with MMSE dataset created by Zhang et al. [24].

4.1. MMSE Dataset

For the experiments, we used the MMSE dataset which consists of about 100 videos recorded by RGB cameras with the frame rate of 25 Hz. This dataset was chosen for the following reasons: first of all, the dataset is large enough for training and testing. Secondly, the dataset includes different peoples, such as Europeans, Middle Easterners, South Asians, South Americans and East Asians which potentially makes it more challenging for the RGB signal processing. Thirdly, the contact PPG sensor was used to record the pulse signal as the ground truth which was synchronized with the videos so that it can be used to quantitatively assess the results. Lastly, with emotion elicitation in the experiments, the volunteers showed some movement of faces and heads which make the dataset more complicated and closer to the realistic scenarios. Similar to all the other rPPG experiments, the volunteers were asked to sit at a fixed distance from the web camera with a background board. A simplified description of the experiment set up and several sample images are shown in Figure 2.

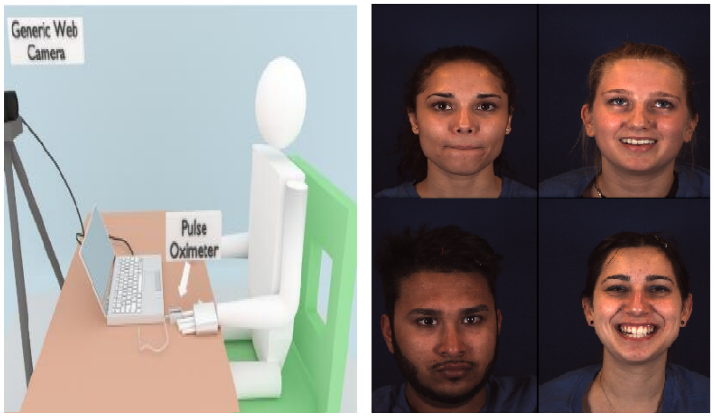


Figure 2. Experimental set up and some sample images from the MMSE database.

4.2. The State-of-the-Art Methods for BVP Peak Detection

In the experiments, we assessed two state-of-the-art BVP peak detection methods as reference, namely the Local Maximum method and the SSF method.

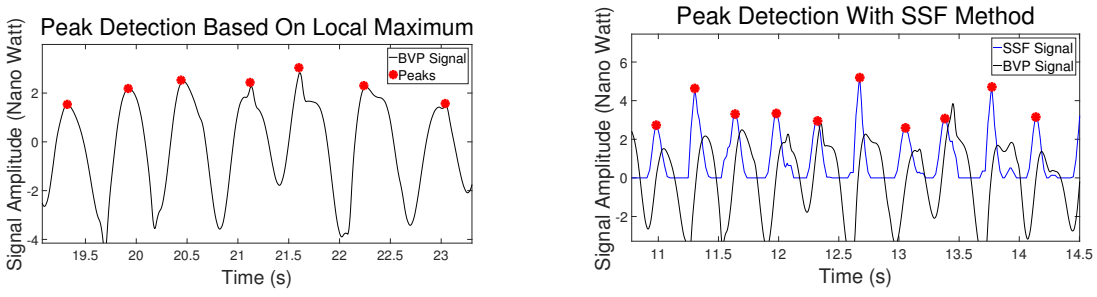
4.2.1. Local Maximum

The Local Maximum detection with rules [20] is the most straightforward method for peak detection of BVP signals. Fortunately, the MATLAB function *findpeaks* has provided us with several useful features:

- *MinPeakHeight*: the minimum height of detected peaks.
- *MinPeakDistance*: the minimum distance between detected peaks.
- *MinPeakProminence*: the minimum height of the peaks relative to the lowest bottom line within a certain range.

To detect the peaks of the remotely measured BVP signal with *findpeaks*, we used the brute force search to determine the values of the parameters. The ground truth of peak locations are given by the contact PPG. We used the number of correctly detected peaks with regards to the ground truth, the number of false and missing peaks with regards to the ground truth, and the average errors of peak locations to optimize the parameters. With the the MMSE dataset, 37 videos were used to get the parameters, and the other 54 videos were used to test the methods. As a result, the *MinPeakHeight* was set as $0.75 * |\overline{y}|$ and the *MinPeakProminence* was $0.3 * |\overline{y}|$, where $|\overline{y}|$ is the average absolute value of the detrended BVP signal. The *MinPeakDistance* was set as 0.24 s.

Figure 3a shows an example of the peak detection based on Local Maximum. It can be seen that it functioned effectively in this case.



(**a**) (**b**)

Figure 3. (**a**) An example of Local Maximum method. (**b**) An example of the Slope Sum Function (SSF) method. The original BVP signal is black and the SSF signal is blue.

4.2.2. Slope Sum Function (SSF)

Zong et al. [22] proposed to use slope sum function (SSF) to detect the onset of Arterial Blood Pressure Pulse by enhancing the rising part of the signal and reducing the descent part. Since remotely measured BVP signals are periodicity exhibiting signals, it is reasonable to adopt this idea to reduce the noise of the signal and make it clearer to detect the peaks [9,23]. The slope sum function is expressed as:

$$SSF(i) = \sum_{i-w}^{i} \Delta y_i \text{ and } i = w+1, w+2, ..., N \tag{9}$$

and Δy_k is expressed as:

$$\Delta y_i = \begin{cases} y(i) - y(i-1) & \text{if } y(i) - y(i-1) > 0 \\ 0 & \text{if } y(i) - y(i-1) \leq 0. \end{cases} \tag{10}$$

Equations (9) and (10) show the calculation of the new signal $SSF(i)$ transformed from the original signal, where y is the original signal, i represents the time index of the signal and w is the window size.

To maximize SSF, the window size w should be approximately the same with the length of the rising phase of the original signal. Similar to the Local Maximum method described in Section 4.2.1, this window size and the minimum height were determined by the brute force search. The exact value of w was 0.4 s. The threshold height was 0.175 * $|\overline{y}|_{ssf}$, where $|\overline{y}|_{ssf}$ is the average absolute value of the detrended new signal. The minimum peak distance was 0.24 s.

Figure 3b shows an example of the pre-filtered BVP signal transformed by SSF. The black curve is the BVP signal generated from the MMSE dataset. The blue curve is the new signal after it got processed by SSF. It can be seen that after using the SSF, the upslope part is enhanced and it becomes sharper and more straightforward for peak detection.

4.3. System Framework

Our system framework of remote PRV measurement is presented in Figure 4. To get the temporal RGB signals from the video frames, firstly the Viola-Jones face detector and the Kanade-Lucas-Tomasi tracker provided by the OpenCV toolbox were used to get stable facial regions. Then the facial landmarks implemented by Dlib C++ Library [25] were applied to the regions to get the coordinates of corners. The region of interest was cropped based on these corners. Then we used Conaire's method [26] to detect the skin pixels and discard the non-skin pixels. These skin pixels were spatially averaged to obtain the 3-band RGB signals. The one-dimensional Blood Volume Pulse (BVP) signal was obtained using the newly proposed PVM method as explained in Section 3.1. Then the peak detection methods such as Slope Sum Function (SSF), Adaptive Two-Window method and Local Maximum were used to get the peaks of BVP. With the peaks, the inter-beat intervals (IBI) can be calculated. According to Malik et al. [1], the PRV can be represented in two different ways. It is either calculated as a peak interval series versus number of progressive peaks, or a peak interval series versus time, which is obtained as a signal of a function of time by interpolating the discrete event series (DES). We chose the latter, because in video analysis the FFT is usually utilized to extract the frequency features, and the FFT can only be used with evenly sampled data. In our case, the PRV was obtained by interpolating the IBI signal with the frame rate of 200 Hz.

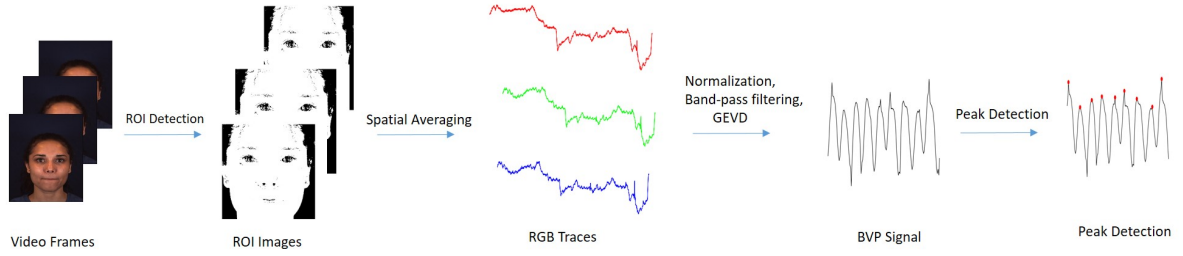


Figure 4. System Framework. (**a**) The original video frames. (**b**) The detected skins. The white part is the detected skin pixels and the black part is the non-skin pixels. (**c**) Spatially averaged RGB signals. (**d**) BVP signal. (**e**) Peak detection.

5. Results

5.1. Evaluation Metrics

The ground truth of the BVP signal and peak locations were given by the contact sensor, so the precision of the peak detection and PRV measurement can be assessed by the errors between the rPPG signals and the ground truth. The evaluation metrics of this paper are shown in Table 1, which are classified as three groups of metrics:

(1) Peak Detection Errors. This group of metrics is used to evaluate the accuracy of peak detection.

- Peak Location Errors (**PLE(s)**). It is calculated as the average absolute difference between the peak locations detected on the rPPG signal and the annotated peaks of the ground truth.
- Proportion of correctly/incorrectly detected peaks and missing peaks (**%CP**,**%IP** and **%MP**). Since the ground truth is provided by the contact sensor measured from fingers and the rPPG was measured from faces, there is a time difference between the peaks on the rPPG signal and the contact PPG signal which is possibly caused by the different distance from the heart and the recording sensor. As a result, the search range for the correctly detected peaks was set to 0.2 s. If there is more than one peak in the search range, then the extra peaks are considered as incorrectly detected peaks. If there is no peak, then it is considered as a missing peak. With these conditions, **%CP** is calculated as the number of correctly detected peaks over the number of peaks of ground truth. **%IP** and **%MP** are calculated in the same way.

(2) PRV Errors. This group of metrics is used to evaluate the accuracy of PRV measurement.

- PRV Errors (**$\mathbf{PRV_{er}}$(s)**) and Inter-beat Interval Errors ((**$\mathbf{IBI_{er}}$(s)**). PRV is obtained as the peak interval series over time interpolated with the frame rate of 200 Hz. IBI is the peak interval series versus number of progressive peaks. Both **$\mathbf{PRV_{er}}$(s)** and **$\mathbf{IBI_{er}}$(s)** are calculated as the absolute average difference between the rPPG signal and gound truth contact PPG signal.
- Relative PRV Errors (**%$\mathbf{PRV_{er}}$**). It is calculated as the average value of **$\mathbf{PRV_{er}}$(s)** over the PRV of the ground truth.

(3) PRV Feature Errors. This group of metrics is used to evaluate the accuracy of PRV features.

- Errors of Standard Deviation of IBI Series (**$\mathbf{STD_{er}}$(s)**). This is calculated as the absolute difference between the Standard Deviation (STD) of rPPG measured IBI and the STD of IBI measured by ground truth contact PPG signal.
- Errors of Root Mean Square of Successive Inter-Beat Interval Differences (RMSSD) (**$\mathbf{RMSSD_{er}}$(s)**). As before, this metric is calculated as the absolute difference between the RMSSD measured by rPPG and the RMMSSD measured by the ground truth. The RMSSD was defined as:

$$RMSSD = \sqrt{\frac{1}{N-1}(\sum_{i=1}^{N-1}(IBI_{(i+1)} - IBI_i)^2)} \tag{11}$$

where IBI_i is the ith peak interval value.

The experimental results were the average values of the entire testing dataset (54 videos).

Table 1. Evaluation Metrics.

Category	Metrics	Denotation	Unit
Peak Detection Errors	Peak Location Errors	**PLE(s)**	Seconds (s)
	Proportion of correctly detected peaks	**%CP**	Percentage (%)
	Proportion of incorrectly detected peaks	**%IP**	Percentage (%)
	Proportion of missing peaks	**%MP**	Percentage (%)
PRV Errors	Inter-beat interval Errors	$\mathbf{IBI_{er}(s)}$	Seconds (s)
	PRV Errors	$\mathbf{PRV_{er}(s)}$	Seconds (s)
	Relative PRV Errors	$\mathbf{\%PRV_{er}}$	Percentage (%)
PRV Feature Errors	Errors of Standard Deviation of IBI signal	$\mathbf{STD_{er}(s)}$	Seconds (s)
	Errors of Root Mean Square of Successive Inter-Beat Interval Differences (RMSSD)	$\mathbf{RMSSD_{er}(s)}$	Seconds (s)

5.2. Results and Discussion

Table 2 shows the errors of the peak detection from the three methods. The first column is average peak location errors (**PLE(s)**). It can be seen that the Two-Window method performed better than Local Maximum with much smaller error. The SSF has no result in this column because there is a shift between the new signal transformed by SSF and the original signal due to shape change, as can be seen in Figures 3b, 5b and 6a. Therefore, it is not fair to compare SSF with other methods with location errors. The second, third and fourth columns are the proportion of correctly detected peaks/missing peaks/incorrectly detected peaks. These columns show that the Adaptive Two-Window method gives more peaks correctly detected than the other two methods.

The Local Maximum method worked very well in the majority of the cases after we used the brute force search to determine the parameters, as shown in Figure 3a. However, the results of Table 2 do indicate that the Local Maximum method has a higher probability to fail. Figure 5 shows a typical example where the Local Maximum failed but SSF and Two-Window methods were effective. In this specific case, the Local Maximum did not detect the peak between the 18th second and the 19th second because the height of the beat period is significantly lower than the average height of the entire signal which makes the peak value lower than the parameter "*MinPeakHeight*". This case shows the weakness of this method. For all such rule-based Local Maximum methods, the physical properties of the BVP curve itself are ignored, and the parameters of the rules are either set manually or using optimization approaches, and it is effective in most of the cases, but it could fail in a few specific cases even if the parameters are fully optimized. On the other hand, the Two-Window method utilizes the physical property that the average height of the systolic peak period is higher than that of the beat period, so a sudden decline of the curve does not affect the performance of the algorithm.

Figure 5b shows why the SSF cannot perform the precise location detection. The SSF is effective because it transforms the signal with an enhanced increasing trend, and thus it avoids losing the fourth systolic peak in this case. However, with the transformation, the shape of the original signal was changed, so the detected locations are slightly more distant from the original peaks. We used this method because the PRV calculation may not be affected if the shift is close to a constant for every peak.

Table 2. The average peak detection errors.

Methods	PLE(s)	%CP	%MP	%IP
Local Maximum	0.1423	87.84%	3.770%	8.390%
SSF	X	90.53%	4.030%	6.310%
Two-Window	0.1221	94.02%	1.960%	4.020%

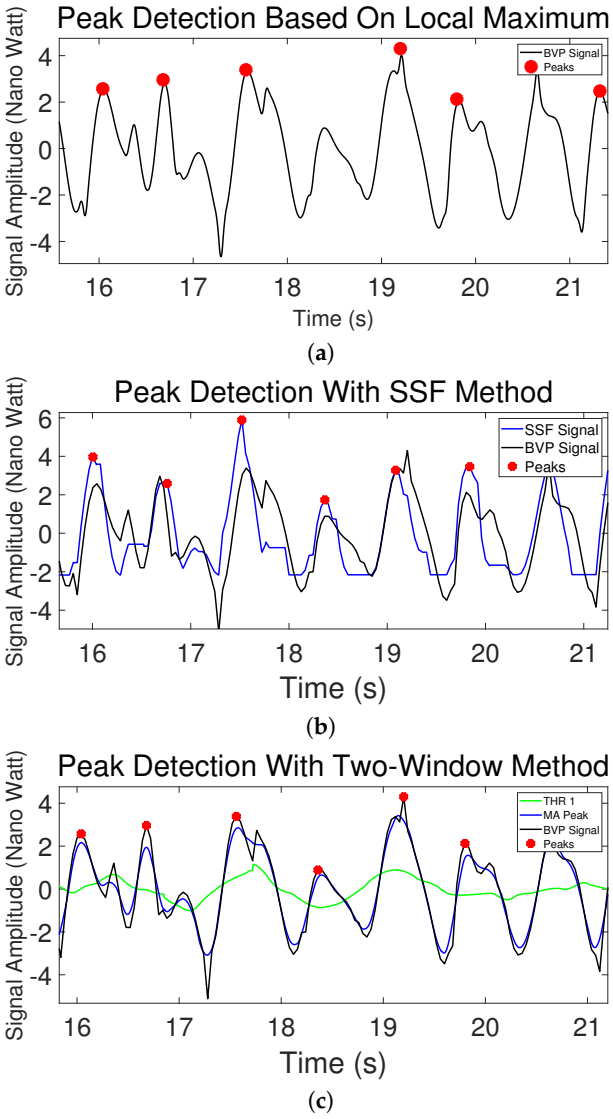


Figure 5. Peak detection on rPPG signal (BVP) with (**a**) Local Maximum, (**b**) SSF and (**c**) Two-Window methods.

Table 3 shows the results of PRV measurement errors. The IBI is the peak intervals versus number of progressive peaks and the PRV is the peak intervals versus time and is obtained by interpolating the interval series with the time stamps of 200 Hz. According to the table, the Adaptive Two-Window method generated better results for all the three PRV metrics than the other two methods with smaller errors. And the SSF is better than the Local Maximum method. As explained, the PRV measurement is more difficult than the HR measurement and is the key work of this paper. Thus the better results in PRV measurement are significant achievements and prove the advantages of this new algorithm. The table shows that the results of the PRV obtained by the SSF method are worse than the Two-Window method although it performed better than the Local Maximum, and it means the shift caused by SSF transformation is not constant and the location errors are not perfectly reduced in the decrement calculation of the locations.

Table 3. The average PRV errors.

Methods	IBI_{er}(s)	PRV_{er}(s)	$\%PRV_{er}$
Local Maximum	0.1718	0.1574	21.74%
SSF	0.1510	0.1413	21.56%
Two-Window	0.1407	0.1185	17.03%

Figure 6a shows an example where the IBI calculation of SSF is incorrect. It can be seen that the peaks detected between the first and the fourth second on the SSF signal are at the left side of the peaks of the original BVP signal. And the distance between the first peak on the SSF signal and the first peak on the original BVP signal is much larger than the distance between the second pair of the peaks on the two signals. As a result, the IBI calculated by the first peak location and the second peak location with the SSF method is larger than the real value. On the other hand, the peak detection within the same period performed by Adaptive Two-Window method does not have this problem.

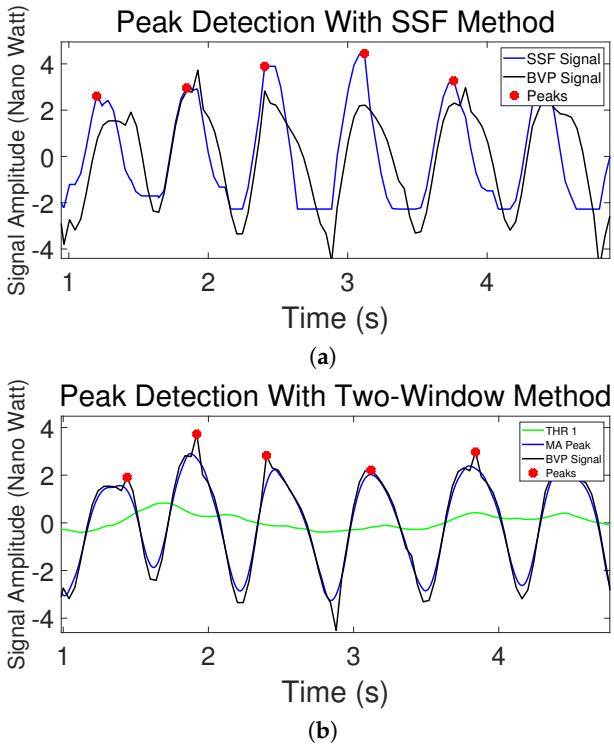


Figure 6. (**a**) SSF in peak detection. (**b**) Two-Window method in peak detection.

The features of the PRV are used in some applications such as emotion recognition. We calculated the errors of the two time domain features ($\mathbf{STD_{er}(s)}$ and $\mathbf{RMSSD_{er}(s)}$). The results are shown in Table 4 and the 95% Confidence Interval of the PRV features' values are shown in Table 5. According to the tables, the Adaptive Two-Window performed the best in both metrics with smaller errors. The Confidence Interval of Adaptive Two-Window method is closest to the ground truth compared with the other two methods. And the Local Maximum performed the worst. It should be noted that although our algorithm has significantly improved the remote PRV measurement compared with the other methods, the errors are still large. This is possibly caused by the low frame rate and the noise of the environment, cameras, and so forth. It means the remotely measured PRV cannot replace HRV in critical medical analysis.

Table 4. The average errors of Pulse Rate Variability (PRV) features.

Methods	STD_{er}(s)	$RMSSD_{er}$(s)
Local Maximum	0.0938	0.1072
SSF	0.0781	0.0718
Two-Window	0.0511	0.0664

Table 5. The 95% Confidence Interval of the PRV features' values.

Methods	STD (95% Confidence Interval) (s)	RMMSD (95% Confidence Interval) (s)
Ground Truth	0.0631 ± 0.0067	0.0670 ± 0.0097
Local Maximum	0.1570 ± 0.0140	0.1742 ± 0.0173
SSF	0.1412 ± 0.0181	0.1388 ± 0.0229
Two-Window	0.1142 ± 0.0133	0.1334 ± 0.0185

The measured PRV in the frequency domain can be evaluated by calculating the Power Spectral Density (PSD) and compared it with the ground truth. Figure 7 shows an example. The black curve is the PSD curve of the ground truth measured by contact PPG. The red/green/blue curves are the PSD signals measured by the Adaptive Two-Window method, SSF method and Local Maximum method respectively. The area within a certain frequency range represents the power of the frequency band. For instance, the area within 0.04 HZ to 0.15 HZ represents the power of the Low Frequency (LF) part which symbolizes the stress state, and the area within 0.15 HZ to 0.4 HZ represents the power of High Frequency (HF) part which symbolizes the relaxed state. Thus the ideal rPPG measured PRV signal should match perfectly with the ground truth signal in the frequency domain. The duration of the videos of the MMSE dataset is very short and some of them are only 20 s so it would be difficult to evaluate the LF part quantitatively. But the results can still be assessed visually. According to the figure, the Adaptive Two-Window method performed better than the other two methods in this case. The SSF method is worse than the Two-Window method with bigger errors in both LF part and HF part and the Local Maximum method performed the worst with very large error in LF part.

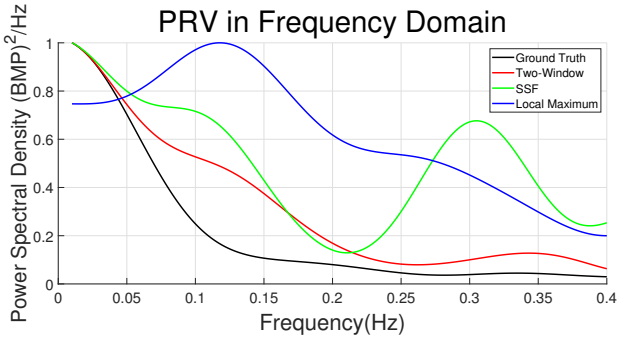


Figure 7. Comparison of PRV measurement with Local Maximum, SSF and Two-Window methods in the frequency domain. The black curve is the ground truth. The red curve is the Two-Window measured signal. The green curve is the SSF measured signal. The blue curve is the Local Maximum measured signal.

6. Conclusions and Future Work

Biomedical researchers have shown that the Heart Rate Variability (HRV) and Pulse Rate Variability (PRV) are vital physiological signals for many applications such as stress detection, emotion recognition and recovery monitoring. Computer Vision researchers have shown that the Heart Rate (HR) can be remotely measured based on remote photoplethysmography (rPPG) technique with very

high accuracy. However, the precision of remote PRV measurement has not been studied intensively and needs further improvement. In this paper, we proposed to use the Periodic Variance Maximization (PVM) method to extract the remotely measured Blood Volume Pulse (BVP) signal and the Adaptive Two-Window Peak Detection to improve the precision of the peak detection of BVP so that we improved the remote PRV measurement. We tested the algorithm with the MMSE dataset which is recorded with the low-cost RGB camera and quantitatively compared this method with the Local Maximum method and Slope Sum Function method with ten metrics. The results showed that the Adaptive Two-Window method performed better than the other two methods so that our new algorithm can be potentially used in the realistic PRV applications.

However, the results also showed that the relative precision of the PRV measurement is not higher than 90%. A possible approach to address the issues in the future is to study the physical properties of the noise, combine all the methods and propose detailed global rules for the BVP filtering and peak detection for different dataset so that the achievements can be used in realistic applications such as long-term health monitoring, lie detection, and so forth. Another issue is that the PRV measured by rPPG cannot yet replace HRV in critical medical applications, therefore it is important to produce a dataset with the ground truth provided by ECG and improve the remotely measured PRV based on this ground truth.

Author Contributions: Conceptualization, P.L. and Y.B.; methodology, P.L. and R.M.; software, P.L., R.M. and Y.B.; validation, P.L.; formal analysis, P.L.; investigation, P.L.; resources, Y.B.; data curation, Y.B.; writing–original draft preparation, P.L.; writing–review and editing, P.L., Y.B., R.M., C.L., and F.Y.; visualization, P.L.; supervision, C.L., Y.B. and F.Y.; project administration, F.Y.; funding acquisition, K.N. and R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Honda Research Institute Japan Co., Ltd and China Scholarship Council.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Malik, M. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. *Ann. Noninvasive Electrocardiol.* **1996**, *1*, 151–181. [CrossRef]
2. Acharya, U.R.; Joseph, K.P.; Kannathal, N.; Lim, C.M.; Suri, J.S. Heart rate variability: A review. *Med. Biol. Eng. Comput.* **2006**, *44*, 1031–1051. [CrossRef] [PubMed]
3. Kim, H.G; Cheon, E.J.; Bai, D.S. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investig.* **2018**, *15*, 235. [CrossRef] [PubMed]
4. Levenson, R.W. The autonomic nervous system and emotion. *Emot. Rev.* **2014**, *6*, 100–112. [CrossRef]
5. Gil, E.; Orini, M.; Bailón, R. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiol. Meas.* **2010**, *31*, 1271. [CrossRef] [PubMed]
6. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 7–11. [CrossRef] [PubMed]
7. Li, P.; Benezeth, Y.; Nakamura, K. Comparison of Region of Interest Segmentation Methods for Video-Based Heart Rate Measurements. In Proceedings of the 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, Taiwan, 29–31 October 2018; pp. 143–146.
8. Lázaro, J.; Gil, E.; Vergara, J.M. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE J. Biomed. Health Inform.* **2013**, *18*, 240–246. [CrossRef] [PubMed]
9. Jang, D.G.; Park, S.; Hahn, M. A real-time pulse peak detection algorithm for the photoplethysmogram. *Int. J. Electron. Electr. Eng.* **2014**, *2*, 45–49. [CrossRef]
10. Akar, S.A; Kara, S.; Latifoğlu, F. Spectral analysis of photoplethysmographic signals: The importance of preprocessing. *Biomed. Signal Process. Control.* **2013**, *8*, 16–22. [CrossRef]
11. McDuff, D.; Gontarek, S.; Picard, R. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2593–2601. [CrossRef] [PubMed]

12. McDuff, D.; Hernandez, J.; Gontarek, S. Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 4000–4004.
13. McDuff, D.; Gontarek, S.; Picard, R. Remote measurement of cognitive stress via heart rate variability. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 2957–2960.
14. Mitsuhashi, R.; Iuchi, K.; Goto, T. Video-Based Stress Level Measurement Using Imaging Photoplethysmography. In Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 90–95.
15. Macwan, R.; Bobbia, S.; Benezeth, Y. Periodic variance maximization using generalized eigenvalue decomposition applied to remote photoplethysmography estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1332–1340.
16. Elgendi, M.; Norton, I.; Brearley, M. Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. *PLoS ONE* **2013**, *8*, e76585. [CrossRef] [PubMed]
17. Belaiche, R.; Sabour, R.M.; Migniot, C. Emotional State Recognition with Micro-expressions and Pulse Rate Variability. In Proceedings of the Image Analysis and Processing—ICIAP 2019, Trento, Italy, 9–13 September 2019; pp. 26–35.
18. Gupta, O.; McDuff, D.; Raskar, R. Real-time physiological measurement and visualization using a synchronized multi-camera system. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 312–319.
19. McDuff, D.; Blackford, E.; Estepp, J. Fusing partial camera signals for noncontact pulse rate variability measurement. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 1725–1739. [CrossRef] [PubMed]
20. Fukunishi, M.; Mcduff, D.; Tsumura, N. Improvements in remote video based estimation of heart rate variability using the Welch FFT method. *Artif. Life Robot.* **2018**, *23*, 15–22. [CrossRef]
21. Béres, S.; Holczer, L.; Hejjel, L. On the Minimal Adequate Sampling Frequency of the Photoplethysmogram for Pulse Rate Monitoring and Heart Rate Variability Analysis in Mobile and Wearable Technology. *Meas. Sci. Rev.* **2019**, *19*, 232–240. [CrossRef]
22. Zong, W.; Heldt, T.; Moody, G.B. An open-source algorithm to detect onset of arterial blood pressure pulses. *Comput. Cardiol.* **2003**, 259–262. [CrossRef]
23. Li, P.; Benezeth, Y.; Nakamura, K. An Improvement for Video-based Heart Rate Variability Measurement. In Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19 July 2019; pp. 435–439.
24. Zhang, Z.; Girard, J.M.; Wu, Y. Multimodal spontaneous emotion corpus for human behavior analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3438–3446.
25. Davis, E. Dlib-ml: A Machine Learning Toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
26. Conaire, C.O; O'Connor, N.E.; Smeaton, A.F. Detector adaptation by maximising agreement between independent data sources. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–6.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).