Heart Rate Estimation Using Remote Photoplethysmography

Ali Azak, Murat Sahin

***Abstract*—Remote photoplethysmography (rPPG) methods are receiving significant attention due to their potential use cases for remote heart rate measurement. This work presents a pipeline consisting of several key steps to implement this technique. The methodology was tested on a popular dataset in the field, and the results were comparable to those obtained using classical methods.**

***Index Terms*—PPG, rPPG, remote photoplethysmography, heart rate estimation, non-invasive measurement**

I. Introduction

HEART rate is an important indicator of cardiac health and a widely used tool for clinical diagnosis. Historically, heart rate was measured through manual pulse checking and in modern times, EKG and PPG methods are popular. PPG works by employing light beams through sensors attached to thin body parts. Sensors emit light onto the skin and measure reflected light resulting from the absorption of blood vessels. This method is used in smartwatches, fitness bands, and other commercial products.

In cases involving sensitive skin, the previous methods may be difficult to use. This motivates the use of remote PPG (rPPG) methods. Blood volume changes during each cardiac cycle, leading to continuous changes in light reflectance. Optical sensors, mostly cameras, can observe blood volume changes which are otherwise invisible to the human eye, to derive a PPG signal. In our work, our goal is to estimate subjects' heart rates using this technique, and we have constructed a pipeline consisting of several key steps to achieve our goal. Our results were evaluated on the UBFC-rPPG [2] dataset.

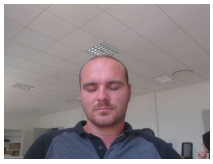
II. Dataset

UBFC-rPPG dataset consists of video recordings captured by a low-cost webcam shooting at 30 fps with a resolution of 640x480. Ground truth heart rates captured by an oximeter are provided. Experiments are conducted indoors with varying illumination scenarios. The dataset splits into two subsets. In the first set, participants are asked to stand still and do nothing, and in the other, they are asked to play a time sensitive mathematical game. Sample frames can be seen from the Figure 1a and 1b.

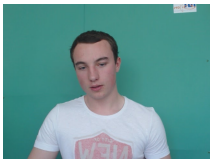
III. Methodology

A. Region of Interest (ROI) Extraction

1) Initial approach: Initially, frames of a video were read and Viola Jones was used for face detection to identify and extract the subject's face from each frame. Figure 2 demonstrates this process.



(a) First set



(b) Second set

Fig. 1. Comparison of dataset samples from different sets

(a) Example frame

(b) Extracted face

Fig. 2. Face extraction demonstration

Several issues were encountered with this approach. Firstly, the outcome was significantly influenced by whether the person's eyes or mouth were open or closed (see Figure 3b for an example where the eyes are open). Also, face extraction can lead to different coverage of the face each frame and affect the results, see Figure 3a.

2) Second approach: To mitigate these problems, we considered locating the face, extracting a small area on the forehead and use this predetermined area for spatial averaging. While this approach addresses the aforementioned issues, it introduces another problem. Subjects have varying hair styles and forehead sizes, as shown in Figure 4. Therefore, we had to define the forehead region for each subject individually, which was not feasible nor realistic.

3) Final approach: To address the issue, facial landmark points were utilized. The 81 point facial landmark predictor [3] was used with dlib library [5] to extract these points. Figure 5a displays the extracted landmarks for an example subject. Following this operation, various regions were chosen using landmarks. To account for varying lighting conditions and improve color observation, four regions were selected from different areas of the face, including the forehead, cheeks,

(a) Bigger background

(b) Open eyes

Fig. 3. Some problematic cases for Viola Jones approach

(a) Case I

(b) Case II

Fig. 4. Problem demonstration for fixed location approach
The subject's hair is blocking the region in Case II

(a) Visualization of landmarks

(b) Visualization of the regions formed by landmarks

Fig. 5. Demonstrations of landmark approach

and chin. Figure 5b provides a visual representation of these regions of interest on the same subject.

These regions are then individually masked for further processing. The information gathered from different ROIs differed for different cases. Initially, we attempted to average over them and continue with a single pass, but this approach performed poorly. Therefore, we process them individually, and later in this section, we explain how we merge their information. This approach provides robustness.

B. Constructing Initial Signal

The initial signals from the entire video are obtained by averaging the pixel values of selected regions of interest, namely mean signals. In the case of RGB color space, there are three different color channels. The most common approach is to average the values of each color channel individually. Figure 6 shows an example of signals obtained using this method.

However, in the literature, different methods are employed. Verkruysse et al. [14] found that the green channel in RGB produces the strongest HR signal, while Pal et al. [10] state that the red channel in RGB gives better results. Due to these conflicting results in the literature, we conducted our own experiments. Figure 7 shows the results of an experiment that compares using only red and green channels to using all RGB channels. It is important to note that this comparison is not completely reflective since the component analysis step cannot be performed with a single channel, and its effect greatly

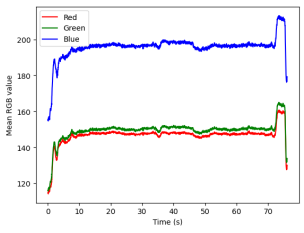
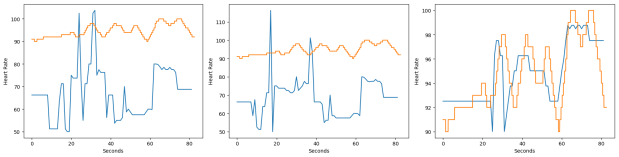
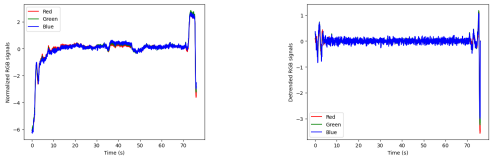


Fig. 6. Mean pixel values of a ROI



(a) Red Channel (b) Green Channel (c) RGB Channels

Fig. 7. Color channel experiments (GT: Orange, Pred.: Blue)



(a) Normalized signal (b) Detrended signal

Fig. 8. Different operations over signal

impacts the results. We have concluded that RGB channels combined give the best results and continue our experiments using this combination.

C. Normalizing and Detrending

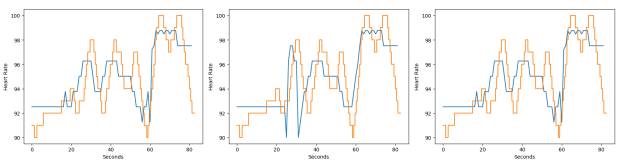
The extracted signals contain pixel intensity values, which are related to the heart rate of the person. However, environmental artifacts can also affect the intensity values, such as changes in illumination, shadows, and camera noise. To mitigate these effects, as much as possible, we have applied normalizing and detrending methods.

The signals are normalized by subtracting the mean and dividing by the standard deviation. Next, a detrending method commonly used in this field [13] is applied to eliminate irrelevant fluctuations in the signals, resulting in a cleaner and more accurate representation of the periodic heart rate variations. The resulting normalized and detrended signal can be seen in Figure 8a and Figure 8b, respectively.

The detrending method assumes that the signal has two components: stationary and trend. It then employs a regularized least squares method to find the trend. It is important to note that the regularization parameter is changeable. We tested different regularization rates (λ = 100, 120, and 140) and found that 120 gave the best result. Refer to Figure 9.

D. Moving Average Filter

The literature commonly uses a moving average filter to smooth signals, facilitating further processing in later pipeline



(a) $\lambda = 100$ (b) $\lambda = 120$ (c) $\lambda = 140$

Fig. 9. Regularization rate experiments for detrending
(GT: Orange, Pred.: Blue)

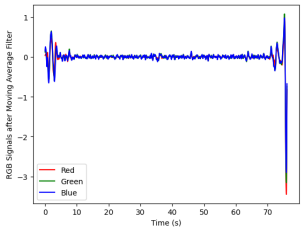
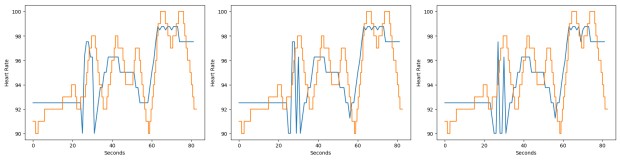


Fig. 10. Filtered signal



(a) Window size = 5 (b) Window size = 7 (c) Window size = 9

Fig. 11. Window size experiments for filtering signal (GT: Orange, Pred.: Blue)

steps, see Figure 10. Various works propose different window sizes, and we experimented with sizes of 5, 7, and 9. The heart rate estimations resulting from the experiments are shown in Figure 11. It is important to note that the window sizes are given in terms of frames, and we have found that a window size of 5 frames (about 0.18 seconds) works best.

E. Source Separation

Source separation methods, such as PCA or ICA, are beneficial for obtaining a more robust signal. These methods separate mixed signals into their primary sources, allowing for a better understanding and analysis of individual components. In this scenario, the extracted components can be interpreted as distinguishing the main signal from the noise, making them quite useful for noise reduction. An example of transforming RGB channels into three independent components using ICA can be seen in Figure 12, it is observable that the noise is clearly separated.

There are varying approaches in the literature regarding the use of these methods. Poh et al. [11] found that ICA improved their results, while Kwon et al. [6] reported that ICA slightly degraded their results. Balakrishnan et al. [1] utilized PCA to create more robust representations. We attempted to use both PCA and ICA, as well as neither of them. The heart rate estimations obtained from these different approaches can be seen in Figure 13.

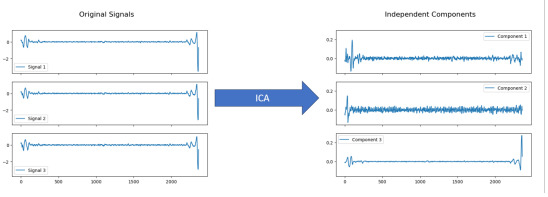
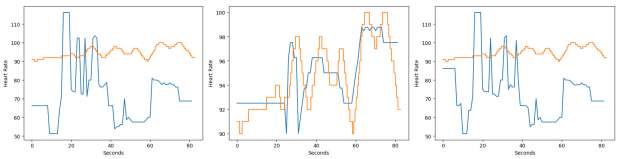


Fig. 12. Independent Component Analysis over RGB signals



(a) Original Signals (b) ICA Components (c) PCA Components

Fig. 13. Source separation experiments (GT: Orange, Pred.: Blue)

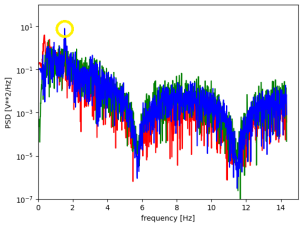


Fig. 14. Power Spectral Density analysis

F. Selecting The Final Signal

Now, we have several signals to choose from and need to determine which one contains the heart rate value. The common approach is to use Power Spectral Density (PSD), which measures how the power of a signal is distributed across different frequencies. Figure 14 shows the power distribution of various independent components. The signal with the highest power at a specific frequency (presumably the heart rate frequency) was selected as the final signal.

G. Bandpass Filtering

A butterworth bandpass filter was used to process the signal and focus on the frequency range associated with heart rate. Bounds for bandpass filter range from 0.5 Hz to 4 Hz in the literature. We have chosen the interval to be in the 0.8 Hz to 2 Hz frequency range, this is determined from the heart rate range in our dataset (0.8 to 2.1 for second dataset). In this step, we are incorporating our prior knowledge of human nature with computer vision. The result of this process is shown in Figure 15.

H. Revisiting ROI selection

It is important to note that we arrived at this point using four different signals obtained from four different ROIs. After the bandpass filtering step, we have four different final signals available for selection. Similar to the component selection step, we again utilize PSD analysis to choose the signal with the most power. This approach makes our predictions more robust against occlusions. Even if some regions of the face are not

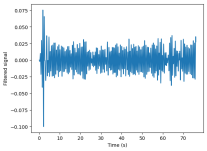


Fig. 15. Bandpass filtered signal

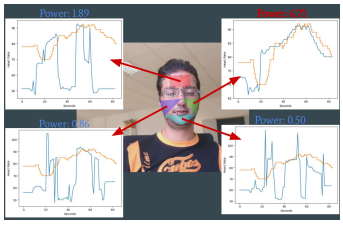


Fig. 16. Heart rate signals obtained with different ROI

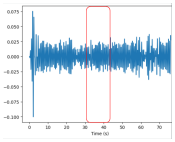


Fig. 17. 12 seconds windowing operation on signal

available, the model can still provide accurate results. Refer to Figure 16 for the various heart rate signals obtained using different ROIs, only the signal with the highest power contains the correct information.

I. PSD

After performing the aforementioned operations, PSD analysis is conducted once again. However, this time, the focus is on identifying the most dominant frequency in a single channel. This frequency is then selected as the heart rate frequency and multiplied by 60 to obtain the final BPM result.

IV. Evaluation

We first calculated the differences between the mean of the ground truth and the estimation for the whole video, but this is not always meaningful. To evaluate the performance of the estimation, we utilized several metrics commonly used in the literature, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Pearson correlation.

For each video in the dataset, the heart rate value is calculated every second using 12-second windows, see Figure 17. Different window sizes was also tested, see Table III for those experiments. The corresponding metric is then calculated using this value and the ground truth value at that second. The mean of these calculations is reported as the final value. Table I displays the individual results for the first set of samples (7 in total) and their mean. Table II presents the results for the combined first and second sets, comprising approximately 50 samples. The comparison between different works in the literature can be observed from the same table. It is important to note that the second dataset is considerably more challenging than the first dataset, as mentioned in the Dataset section. Therefore, calculating for the combined set increases the error as expected.

Table I shows that some subjects performed worse than others. For example, our method had difficulty capturing the heart rate signal for Subject 8, as shown in Figure 18a. However, it performed well in the case of Subject 11, as shown in Figure 18b. This is mainly caused by the challenging

Subject	Mean Difference	MAE	RMSE	MAPE	Pearson Corr.
10-gt	0.521	2.661	3.427	3.654	-0.215
11-gt	0.764	1.151	1.493	1.474	0.768
12-gt	0.486	1.743	2.267	1.838	0.574
15-gt	0.617	1.754	2.771	2.365	0.764
6-gt	0.824	3.763	5.126	4.833	0.752
7-gt	0.121	1.449	1.902	1.530	0.739
8-gt	0.030	6.224	7.117	8.955	0.505
Average	1.337947	2.678	3.443	3.521	0.555

TABLE I
Evaluation on the first dataset

Method	MAE	RMSE	Pearson Corr.
POS (Classical) [16]	8.35	10.00	0.24
CHROM (Classical) [4]	8.20	9.92	0.27
Ours	6.98	10.11	0.40
Green (Classical) [15]	6.01	7.87	0.29
META-rPPG (CNN-LSTM) [7]	5.97	7.42	0.53
SynRhythm (CNN) [9]	5.59	6.82	0.72
PulseGAN (GAN) [12]	1.19	2.10	0.98
Dual-GAN (GAN) [8]	0.44	0.67	0.99

TABLE II
Comparison with other works on the combined dataset

illumination and movement cases, and they are more common in second dataset. You may also look at Figure 19 for some example successful estimations done for the samples from both datasets.

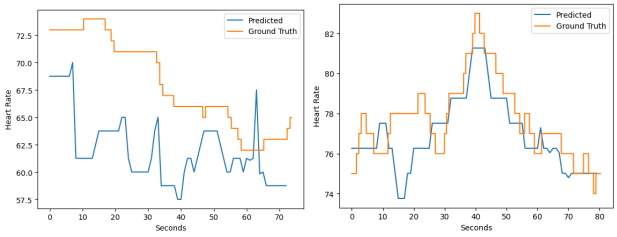
By looking at the Table II, our method works well and it finds a place among the other works that use classical approaches, also our results are more correlated in general. However, it is important to note that deep learning based approaches generally outperform the classical approaches.

V. Conclusion

In conclusion, remote photoplethysmography shows great potential for non-invasive heart rate measurement. Our methodology performs well compared to classical methods, but falls short when compared to those that utilize deep learning techniques. The biggest challenges in this task are subject movement and changes in illumination, and despite our efforts to mitigate their effects, they still persist.

Window Size	Mean Difference	MAE	RMSE
6 frames	4.802	7.312	10.608
8 frames	4.906	7.025	10.173
12 frames	5.065	6.982	10.110

TABLE III
Evaluations for different window sizes



(a) Subject 8 (b) Subject 11

Fig. 18. Example results

REFERENCES

[1] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3430–3437, 2013.

[2] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit. Lett.*, 124:82–90, 2017.

[3] codeniko. shape_predictor_81_face_landmarks. https://github.com/codeniko/shape_predictor_81_face_landmarks, 2019.

[4] Gerard de Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[5] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[6] Sungjun Kwon, Jeehoon Kim, Dongseok Lee, and Kwang Suk Park. Roi analysis for remote photoplethysmography on facial video. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4938–4941, 2015.

[7] Eugene Lee, Evan Chen, and Chen-Yi Lee. Meta-rppg: Remote heart rate estimation using a transductive meta-learner. In *European Conference on Computer Vision*, 2020.

[8] Hao Lu, Hu Han, and S. Kevin Zhou. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12399–12408, 2021.

[9] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Synrhythm: Learning a deep heart rate estimator from general to specific. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3580–3585, 2018.

[10] Arpan Pal, Aniruddha Sinha, Anirban Dutta Choudhury, Tanushyam Chattopadyay, and Aishwarya Visvanathan. A robust heart rate detection using smart-phone video. In *Proceedings of the 3rd ACM MobiHoc Workshop on Pervasive Wireless Healthcare*, MobileHealth '13, page 43–48, New York, NY, USA, 2013. Association for Computing Machinery.

[11] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

[12] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 25:1373–1384, 2020.

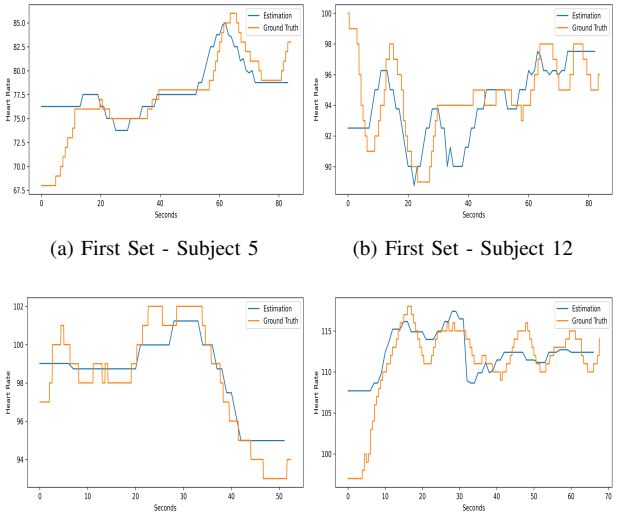
[13] Mika P. Tarvainen, Perttu O. Ranta-aho, and Pasi A. Karjalainen. An advanced detrending method with application to hrv analysis. *IEEE Transactions on Biomedical Engineering*, 49:172–175, 2002.

[14] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.

[15] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, 2008.

[16] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

APPENDIX



(a) First Set - Subject 5 (b) First Set - Subject 12

(c) Second Set - Subject 10 (d) Second Set - Subject 13

Fig. 19. Some successful results from both sets