Students: Adam Grünberg, Leonardo Santiago Benitez Pereira

Introduction

Heart rate monitoring is a crucial component of physiological health assessment, with diverse applications in clinical diagnostics, fitness tracking, and wellness monitoring. Remote Photoplethysmography (rPPG), a technique that extracts pulse signals from subtle color variations in facial skin captured on video, has emerged as a promising contact-free alternative to traditional methods. Recent advancements have imported performance in challenging conditions, like: illumination variations, facial expressions, and makeup effects. Approaches ranging from traditional filtering to deep learning for converting heartbeat signals into spectrograms have further improved the precision of pulse wave extraction and analysis.

This project develops a system for facial rPPG using standard smartphone cameras, aiming to deliver a robust and accurate method for extracting heart rate data from recorded video. The implementation involves facial detection, selection of a stable region of interest (ROI), signal filtering using a Chebyshev bandpass filter, and fiducial point detection for heart rate estimation.

The measurements were validated against a ECG Palmare PM10 device. Additionally, visualizations and quality metrics were used to ensure that the system performs reliably under varying conditions.

In the following section, we will provide a brief overview of the literature and existing solutions, describe how the data was acquired, how the detection pipeline was developed, and validation results. The detection pipeline can be roughly divided 3 sequential steps: selecting the region of interest from the face, filtering the resulting signal to obtain better results, and detecting the fiducial points from the filtered signal.

The implementation of this solution is available in the github repository LeonardoSanBenitez/heart-rate-from-camera, and the recorded data is available in this google drive folder.

Literature review

For the implementation of the project, we took a number of literature as a starting point.

Various approaches have been developed for extracting pulse signals from human face videos, ranging from traditional signal processing techniques to modern deep learning-based methods (Waqar, M., Zwiggelaar, R., & Tiddeman, B. (2021). Classical approaches typically involve preprocessing steps such as region of interest (ROI) extraction, tracking, and skin color estimation, followed by algorithms designed to analyze subtle color changes in the skin. Key methods for signal extraction include PCA (Principal Component Analysis), ICA (Independent Component Analysis), and frequency domain analysis, with PCA being widely used due to its effectiveness in separating PPG signal from noise. Preprocessing techniques such as adaptive filtering and wavelet transforms are key to addressing challenges such as motion artifacts, lighting changes, and facial expressions.

These methods, however, often face challenges with motion artifacts and lighting variations, limiting their robustness. Deep learning techniques aim to address these limitations by leveraging large datasets to automatically learn features for pulse extraction, but they often require significant computational resources (Waqar, M., Zwiggelaar, R., & Tiddeman, B. (2021).

The work of Wang et al. (2020) highlights techniques that could enhance the signal extraction pipeline in our project, particularly with respect to preprocessing methods that preserve waveform fidelity and reduce environmental interference.

Several open-source projects implement softwares and libraries for rPPG. For example, ubicomplab/rppg-web is a web-based solution for real-time heart rate monitoring using a webcam. More oriented towards developers, the repository prouast/heartbeat provides a command-line tool written in C++ from either a prerecorded video or a live feed.

A python solution is provided by habom2310/Heart-rate-measurement-using-camera, with a QT graphical interface. However, this solution performs only minimal filtering of the acquired signal, and estimates the heart rate by simply taking the frequency of the highest peak after the FFT transform. No validation or quality analysis was performed in this tool.

Data acquisition

Face video and ECG were acquired simultaneously, in 40 sessions of 30 seconds each. All recording were performed at PPKE room 204, under similar light conditions, by two subjects (one male and one female, ages between 20 and 30 years old). The videos were taken using two different smartphone cameras: an Iphone 11 Pro (resolution 1920 × 1080, 1x back camera, 30 fps, HD settings, codec H.264 in format mov) and a Galaxy S20 (resolution 474 × 850, 30 fps, codec H.264 in format mp4). The ECG recordings were performed with a Palmare PM10 device (Figure 1).

Figure 1 - ECG device used.

To test the solution under different conditions, 5 recordings were performed for each of the following conditions: neutral facial expressions and looking in the camera direction, varying the illumination in the subject’s face, doing various facial expressions and/or talking, and moving the head without facial expressions. These recordings were repeated with the heartbeat accelerated (after performing some type of physical activity).

The recordings are summarized in the Table 1.

N	Condition	Subject	Recording	HR
1	Neutral	Adam	15:50:04	100
2	Neutral	Adam	15:53:33	101
3	Neutral	Regina	16:45:12	115
4	Neutral	Regina	16:47:16	83
5	Neutral	Regina	16:53:17	93
6	Illumination variation	Adam	15:55:30	129
7	Illumination variation	Adam	15:58:54	128
8	Illumination variation	Regina	16:56:20	89
9	Illumination variation	Regina	17:01:01	84
10	Illumination variation	Regina	17:02:30	89
11	Facial expressions	Adam	16:00:19	140
12	Facial expressions	Adam	16:01:10	143
13	Facial expressions	Regina	17:04:34	116
14	Facial expressions	Regina	17:05:36	86
15	Facial expressions	Regina	17:06:37	93
16	Head movements	Adam	16:02:15	116
17	Head movements	Adam	16:03:30	127
18	Head movements	Regina	17:09:50	82
19	Head movements	Regina	17:11:16	93
20	Head movements	Regina	17:12:06	89
21	Neutral + accelerated	Adam	16:06:11	137
22	Neutral + accelerated	Adam	16:07:00	128
23	Neutral + accelerated	Adam	16:18:44	107
24	Neutral + accelerated	Regina	17:17:49	128
25	Neutral + accelerated	Regina	17:18:49	93
26	Illumination variation + accelerated	Adam	16:08:00	134
27	Illumination variation + accelerated	Adam	16:09:04	111
28	Illumination variation + accelerated	Adam	16:10:00	105

29	Illumination variation + accelerated	Regina	17:20:02	90
30	Illumination variation + accelerated	Regina	17:22:37	113
31	Facial expression changes + accelerated	Adam	16:12:56	126
32	Facial expression changes + accelerated	Adam	16:13:41	116
33	Facial expression changes + accelerated	Adam	16:14:39	145
34	Facial expression changes + accelerated	Regina	17:22:37	113
35	Facial expression changes + accelerated	Regina	17:23:49	110
36	Head movements + accelerated	Adam	16:15:32	116
37	Head movements + accelerated	Adam	16:16:40	110
38	Head movements + accelerated	Adam	16:17:32	107
39	Head movements + accelerated	Regina	17:24:49	97
40	Head movements + accelerated	Regina	17:25:49	97

Table 1 - Recordings

The drive folder contains data organized into the following key components:

1. CSV File:

This file contains the measured data, including:

- Heart rate
- Predicted heart rate
- Systolic peak (SP) points
- Dicrotic notch (DN) points

These data points are essential for analyzing and comparing extracted PPG signals with reference measurements.

2. Recorded Video Folder:

This folder contains video recordings categorized by experimental conditions. The conditions include:

- Neutral
- Illumination variation
- Facial expressions
- Head movements
- Neutral + accelerated
- Illumination variation + accelerated
- Facial expression changes + accelerated
- Head movements + accelerated

These videos were captured under varying environmental and subject conditions to assess the robustness of the PPG extraction method.

3. Reference ECG Signals Folder:

This folder contains reference ECG signals in .dat format. These signals serve as the ground truth for validation, allowing comparisons between the extracted PPG signals and standard ECG measurements.

This organizational structure ensures that all measured data, experimental recordings, and reference signals are easily accessible for analysis and validation purposes.

Detection pipeline

Face detection

For each frame of the video, we detect the presence of a face using the function get_frontal_face_detector of the library dlib (Figure 2). That implementation uses HOG for feature extraction and a Linear SVM for classification, and it is reasonably accurate while still very fast.

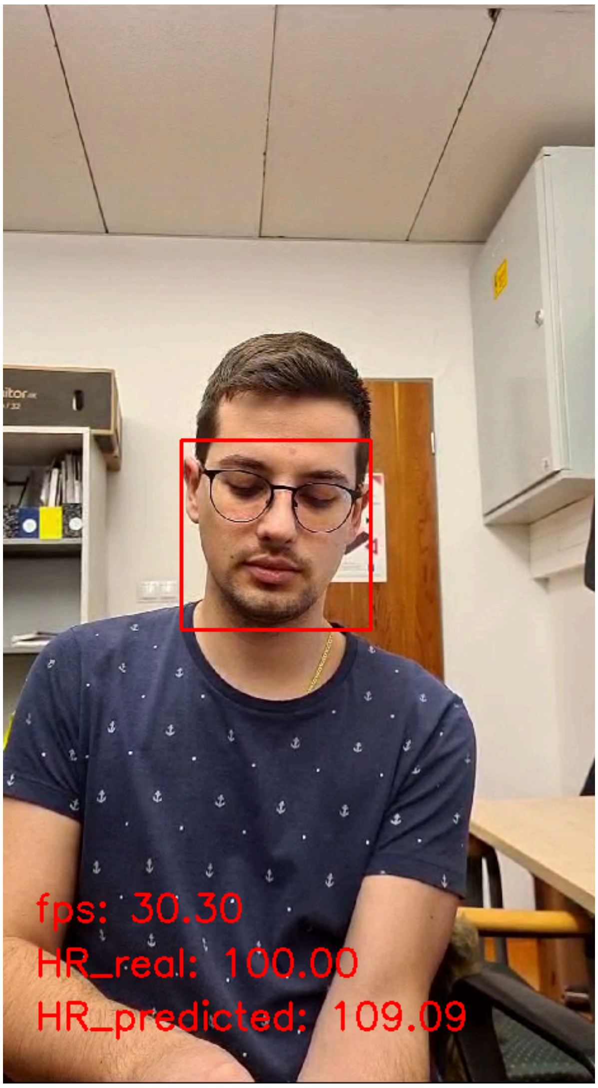


Figure 2 - Face detection using dlib. The recording FPS, ground truth heart rate, and predicted heart rate are superimposed in the image.

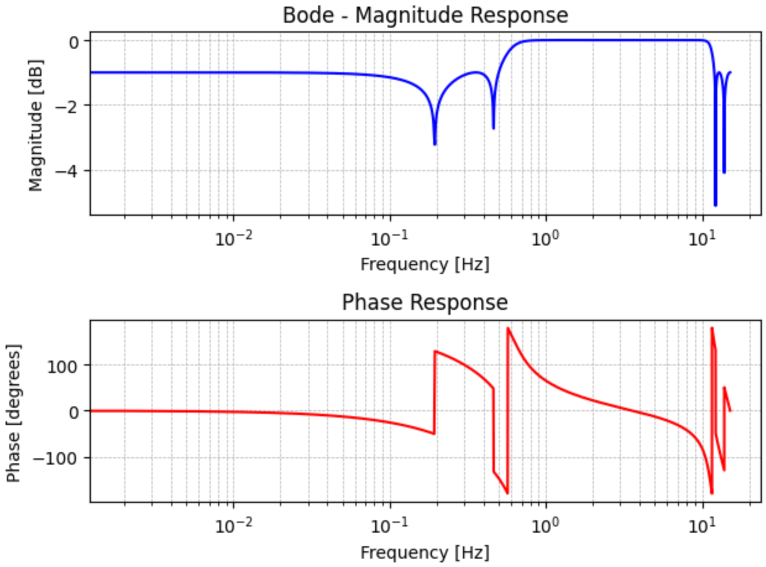
For the image patch containing the face, we use dlib's shape_predictor to estimate 68 landmarks of the face, so we can then extract one "stable region of interest" that does not change considerably with facial expressions or small head movements. The chosen region was the upper part of the cheek, highlighted in green in Figure 3.

Figure 3 - Facial landmark detection and region of interest selection. The detection is robust to differences in gender, use of glasses, facial expressions and head positions.

The face detection pipeline was strongly inspired by the work of https://github.com/habom2310/Heart-rate-measurement-using-camera, and due credit was also provided in the code for the functions used with little further modification.

Data filtering

For each 2D region of interest, we calculate the average value of the green channel. The timeseries of those values is then filtered using a bandpass Chebychev filter of 4th order, with lower and upper cutoff frequency of 0.5 Hz and 12 Hz respectively. The obtained coefficients are `a = [1, -1.45, -0.98, 1.1585, 1.49, -0.70, -0.83, 0.12, 0.20]` and `b = [0.44, -0.18, -1.39, 0.14, 1.98, 0.14, -1.39, -0.18, 0.44]`, and the bode plot and pole-zero plot can be seen in Figure 4. Such implementation was strongly inspired by filtering used in the PyPPG library.



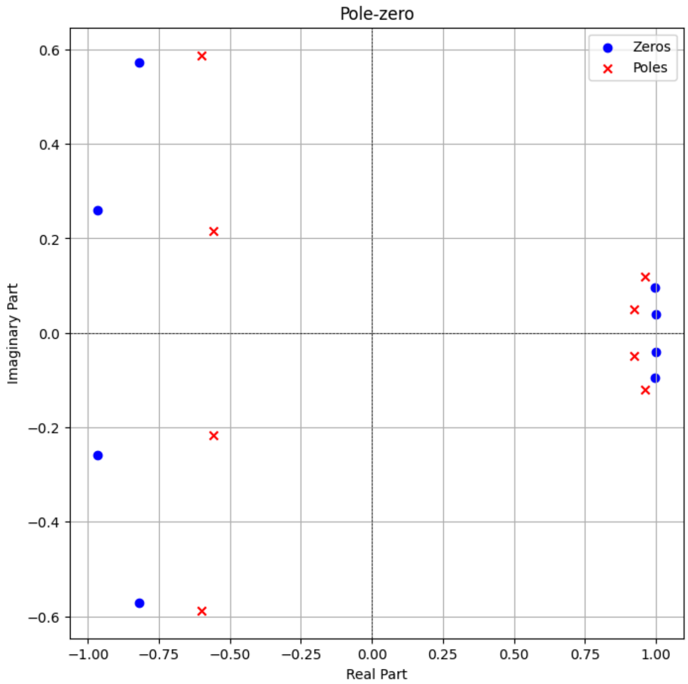


Figure 4 - Frequency and Pole-Zero analysis of the filter. The magnitude plot (blue) demonstrates the filter's selective attenuation of certain frequencies, characterized by sharp notches, indicative of the Chebyshev filter's ripple in the passband and steep cutoff. The phase response (red) reveals rapid phase changes, particularly near the cutoff frequencies. The poles (red crosses) and zeros (blue circles) exhibit the filter's stability and frequency response characteristics, with the positions of poles contributing to the steep roll-off and zeros affecting attenuation.

The results of the filtering seem to work reasonably well, considering how noisy the original signal was. A comparison of the original and the filtered signal can be seen in Figure 5.

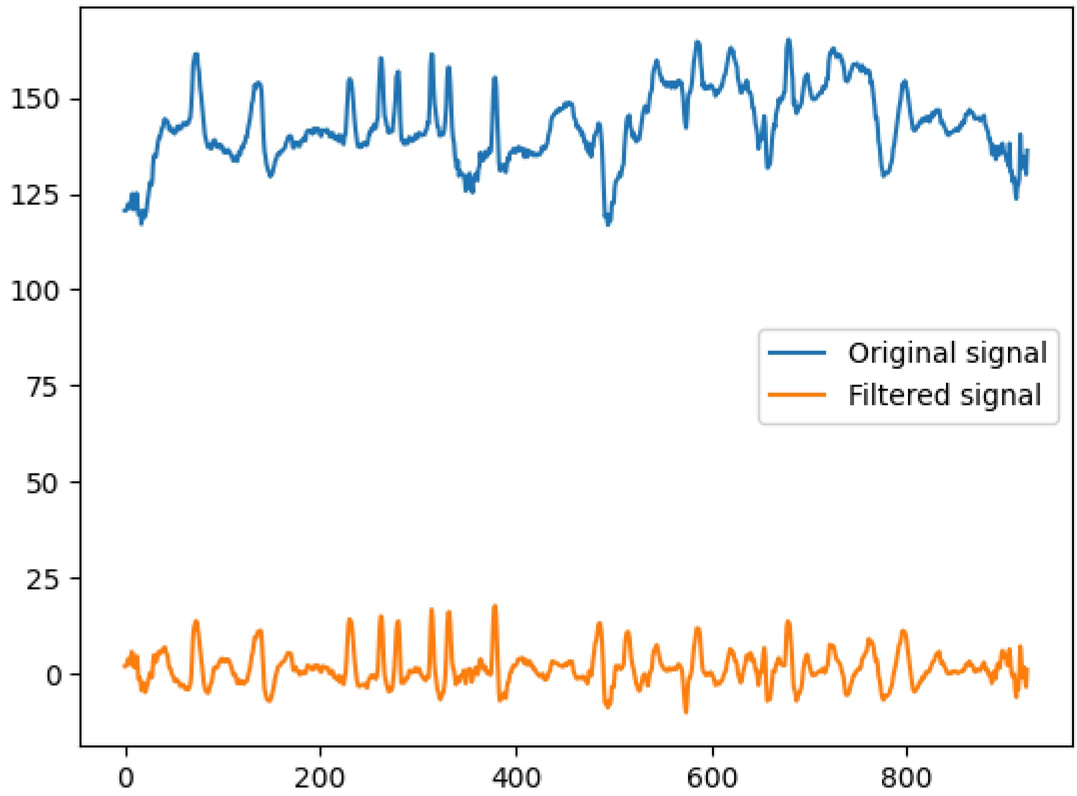


Figure 5 - Comparison of original and filtered signals. the filter effectively attenuates high-frequency noise and large outliers from the original signal, resulting in a smoother and more consistent output.

Heart rate and fiducial points detection

We detected fiducial points on the filtered signal using the PyPPG library, with small modifications due to the low sampling rate of the camera. The heart rate was calculated by the median number of samples between two consecutive systolic peaks. A sample detection if shown in Figure 6.

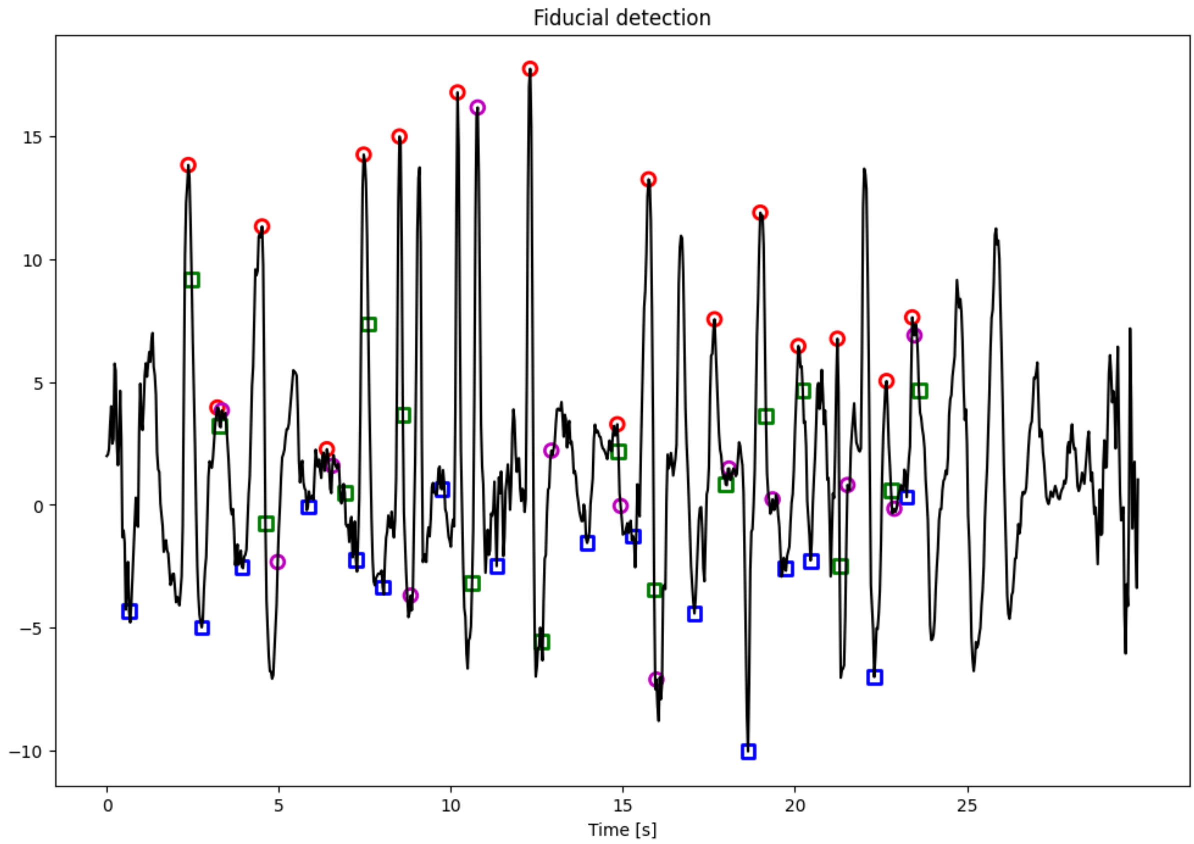
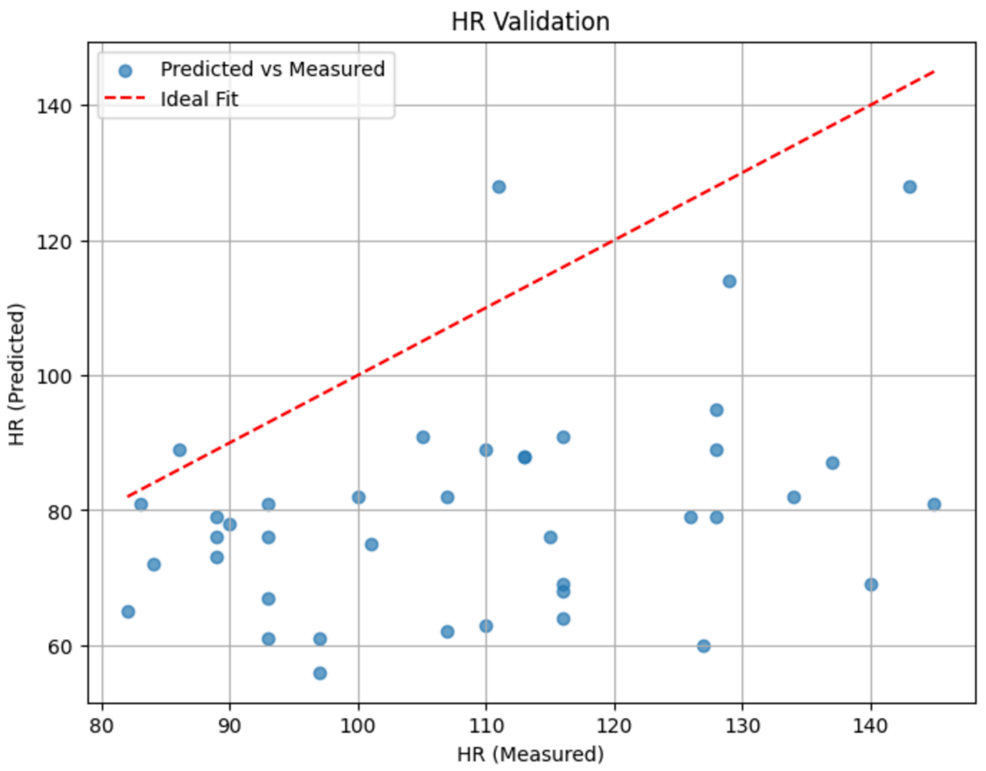


Figure 6 - Fiducial point detection using the library PyPPG. Some beats were not detected, and the final samples were not analyzed.

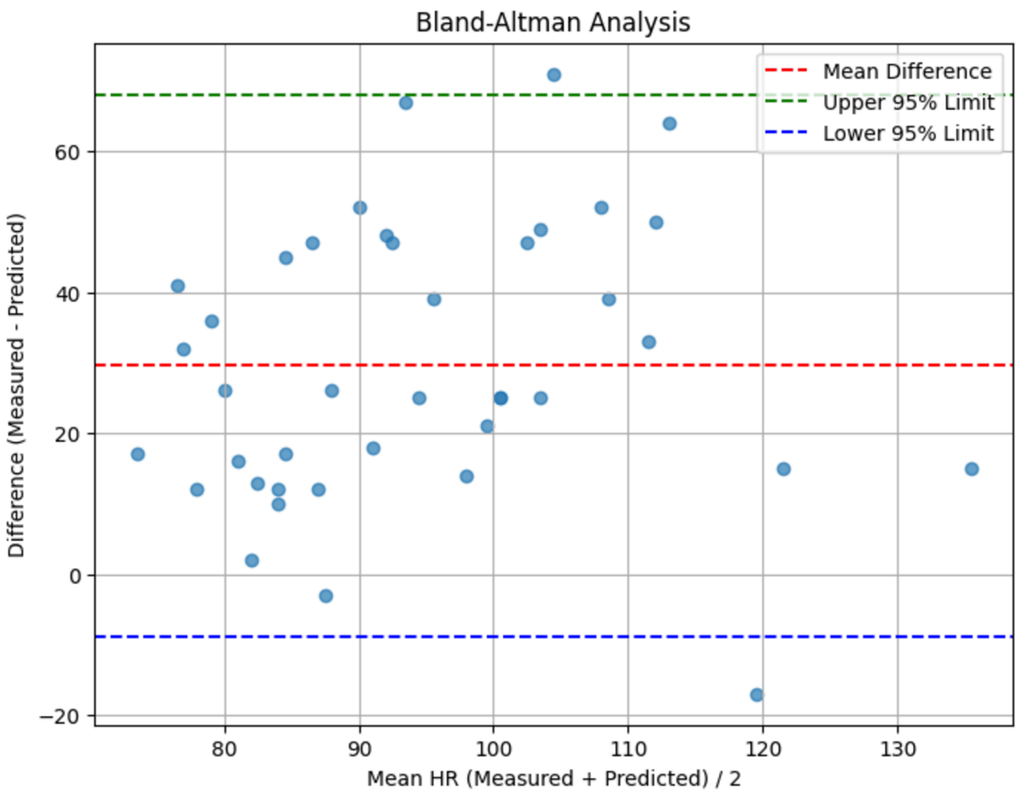
Validation

To validate the extracted PPG signals, comparisons were made against data collected from a Viatom CheckMe Health Monitor's PPG sensor, which serves as a reference standard. The comparison focused on the similarity between the extracted signals and those obtained from the Viatom sensor under identical conditions. Metrics such as signal shape, frequency, and waveform patterns were analyzed to ensure the accuracy and reliability of the smartphone-based PPG extraction method. These graphs are used to validate HR (heart rate) measurement and prediction results.



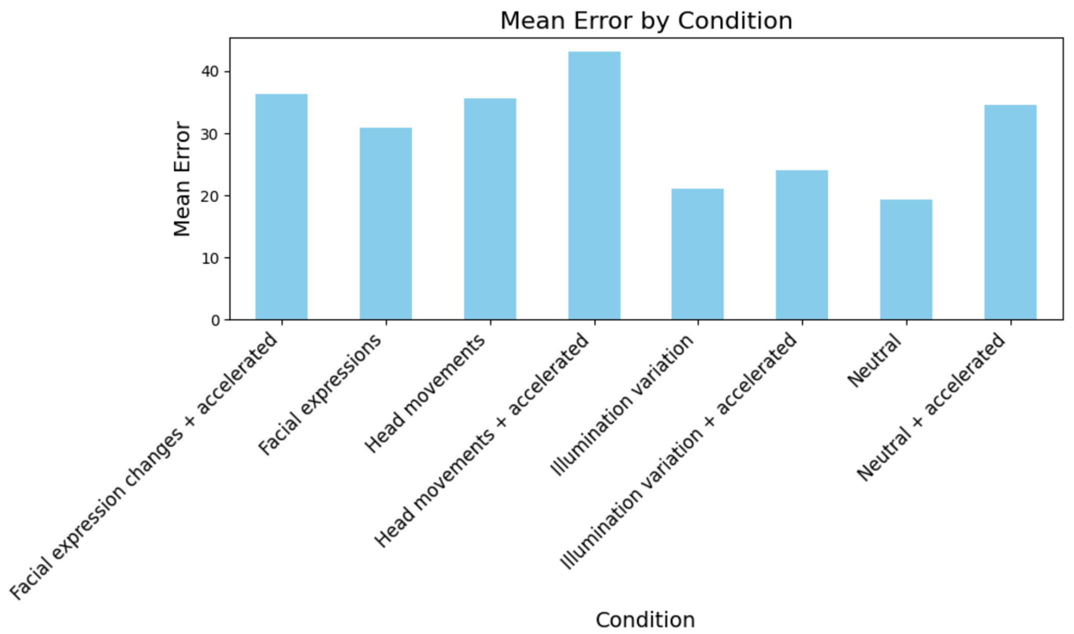
HR Validation Graph:

The first graph compares the measured and predicted HR values. The blue dots show the measured and predicted values. The red dashed line indicates the ideal fit (y=x line).From this graph our observation is that the points have a large variance, meaning the predicted values do not fit the measured data well. Many predicted values significantly underestimate or overestimate the measured HR, especially for higher HRs.



Bland-Altman Analysis Graph:

This graph examines the difference between the measured and predicted HRs (Measured - Predicted) as a function of their mean (Measured + Predicted)/2. The red dashed line shows the mean difference. The green and blue dashed lines indicate the 95% confidence limits. Looking at the graph we can say that the data show a wide spread, suggesting that the predictions are inconsistent. Most of the differences are positive, meaning that the predictions tend to underestimate the observed values. The 95% limits (upper and lower) are wide, suggesting poor prediction accuracy.



Mean error by condition graph:

The next graph shows the average error ("Mean Error") measured under different conditions. The highest error belongs to the "Illumination variation + accelerated" condition. This indicates that the combination of strong light variations and accelerated movement resulted in the highest error. Moving on to the lowest error, which is the "Neutral" condition. This is understandable, since in this case there were probably no distracting factors (e.g. movement, light variation) that would have degraded the performance.

For all "+ accelerated" conditions, we can observe a higher error than for the baseline versions (e.g. "Head movements" < "Head movements + accelerated"). This shows that faster movements significantly increase the error. If we compare the conditions the simple "Facial expressions", "Head movements", and "Illumination variation" conditions show similar levels of error. The largest difference between the simple and "+ accelerated" conditions is observed for "Illumination variation".

From this graph we can conclude that the accelerated movements and illumination variations together significantly degrade the results. This suggests that the model or measurement system cannot handle such interference. The lowest error level indicates that the system works well in a noise-free environment.

Discussion

The results obtained up to the present moment indicate that an PPG-like signal can be obtained from facial videos, but heart rate estimation is far from precise. The error is higher under more challenging conditions such as accelerated movements or significant illumination changes, as was also reported by other works in the literature.

Several improvements can be performed to improve the results, for instance:

- Select another part of the face to be the region of interest
- Improve the filtering to use adaptive filtering or other advanced signal processing techniques
- Use other techniques for identifying fiducial points, like CNNs or LSTMs.

Furthermore, such a system could serve as base for other projects, for example incorporating the solution into an smartphone app or integrating with other physiological monitoring solutions

References

https://link.springer.com/chapter/10.1007/978-3-030-61125-5_10

Dingliang Wang, Xuezhi Yang, Xuenan Liu, Jin Jing, and Shuai Fang, "Detail-preserving pulse wave extraction from facial videos using consumer-level camera," Biomed. Opt. Express 11, 1876-1891 (2020)

ubicomplab/rppg-web

prouast/heartbeat

habom2310/Heart-rate-measurement-using-camera