CS6650 | Smart Sensing for IoT

Smartphone based Photoplethysmography

Assignment 1 Report | CS6650 | Spring 2023

Submitted by

Ashwin A Nayar | NA19B001

Contents

I Introduction 2

II Dataset 2

III Attempt Walkthrough 2

III-A Warmup - Data Collection . . . 2

III-B Sensing Metric . . . 2

III-B1 Assumptions . . . 2

III-B2 Data Pre-processing . . . 2

III-C Temporal Intensity Variation . . . 3

III-C1 Observations . . . 3

III-D Likelihood Distributions . . . 3

III-D1 Observations . . . 4

III-E Threshold Based Detection and ROC Curve . . . 4

III-E1 Preliminaries - The AUC score . . . 4

III-E2 Observations . . . 5

III-F Spatial Correlation Analysis . . . 5

III-F1 Observations . . . 6

I. Introduction

Photoplethysmography (PPG) is a non-invasive optical measurement method that uses a light source and a photo-detector at the surface of the skin to measure the volumetric variations of blood circulation such as Blood Volume Pulse (BVP) signal and Heart Rate (HR). For this study, the blood flow videos at the fingertip is captured using a smartphone camera with the flashlight turned on throughout as the light source. The ground truth was also measured concurrently using a Pulse Oximeter for validation of the technique used for HR estimation. The raw video data is then processed and the BVP signal is extracted. The heart rate is subsequently estimated from the BVP signal followed by some statistical analysis and inferences on the captured data.

II. Dataset

Three different videos are captured for analysis under the following conditions using a smartphone camera at maximum frame rate and resolution available. Each of the videos are approximately 20 seconds long, although data is clipped to first 10 seconds during processing.

- Video ID 1: Resting on bed (70 BPM)
- Video ID 2: After a moderate walk (86 BPM)
- Video ID 3: After a vigorous exercise (140 BPM)

The values in parentheses correspond to the ground truth as measured simultaneously using a pulse oximeter.

III. Attempt Walkthrough

A. Warmup - Data Collection

Frame data from all 3 videos are read and stored along with useful metadata such as frame rate, number of frames read, frame width and frame height. Since duration of videos can be large, the analysis is limited to first N frames, where N corresponds to the number of frames for a 10 second duration.

$$N = 10 \times FPS \tag{1}$$

B. Sensing Metric

The aim is to come up with an aggregate statistical measure for each frame from the individual pixel data (R, G, B values) for a given frame. A time-series of the frame intensity metric is supposed to capture the BVP signal of interest.

1) Assumptions: The raw data as is bound to have perturbations and noise from the environment. Hence noise removal and preprocessing is required to extract a clean BVP signal. We assume that the original signal influences intensity variations in all 3 channels either uniformly or non-uniformly. To deal with the ambiguity in the extent of influence we analyse the projections along the eigenvectors of the data matrix and observe the fact that almost 90% of the energy is captured in the first component itself, this being the highest variance component. We use this projection as our BVP signal henceforth.

2) Data Pre-processing: Each frame is processed sequentially as follows,

- The signal to noise ratio (SNR) is enhanced by spatial average of all the pixels per color channel in each frame. This is followed by application of various filters from Python's scipy.signal library.
- A detrending filter is applied to remove stationary components from the signal.
- Kernel smoothing is carried out using a simple 1x3 kernel per channel.
- Principal Component Analysis (PCA) is applied to map the channel signals to eigenvectors. The projection along the first eigenvector is the component with highest variance and this is chosen as the BVP signal.

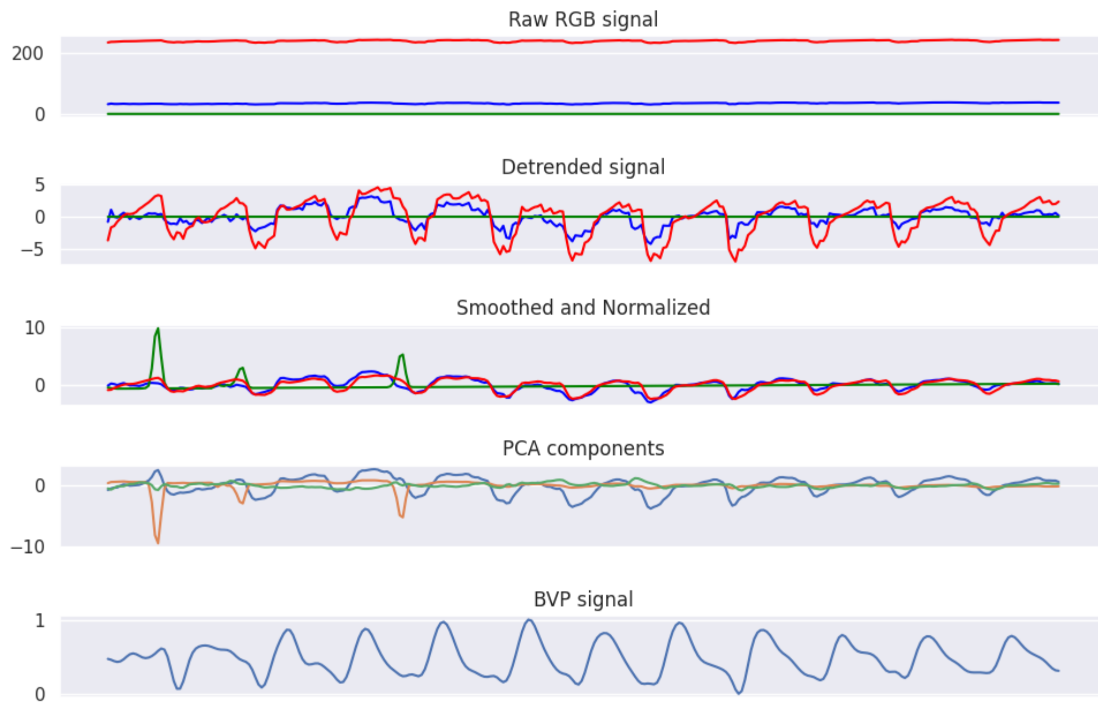


Fig. 1: Preprocessing

- A butterworth bandpass filter is then used to attenuate frequencies outside the desired heartbeat detection range.
- Finally, the signal is normalized to [0, 1] range.

C. Temporal Intensity Variation

The intensity frame metric is plotted against time for the last 5 second chunk of all three captured videos.

1) Observations:

- It is observed that there are noticeable peaks and valleys in the extracted signal. It is possible to estimate the heart rate by manually counting these peaks and the calculated value is in agreement with the ground truth as portrayed in Fig. 2.

$$Heart\ Rate\ (HR) = \frac{no.\ of\ peaks\ in\ sampling\ period \times 60}{Sampling\ period} \tag{2}$$

- Vigorous exercise leads to rapid intensity variations compared to while resting, as is clearely evident from Fig 2

D. Likelihood Distributions

20 frames are chosen in the neighbourhood where the sensing metric is close to the local maximum (Case 1) and another 20 frames are chosen where the sensing metric is close to the local minimum (Case 2). Histograms on all 3 channels (R, G and B) are plotted and the distributions are analysed. The histogram bins are normalized to take into account large frequency values.

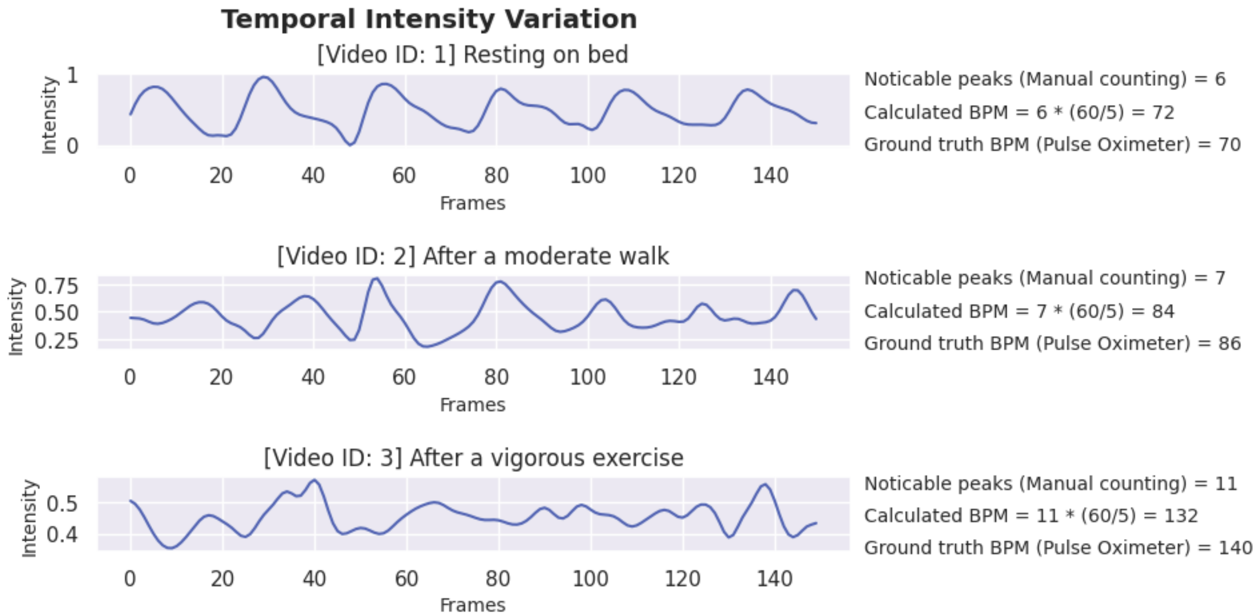


Fig. 2: Temporal Intensity Variation

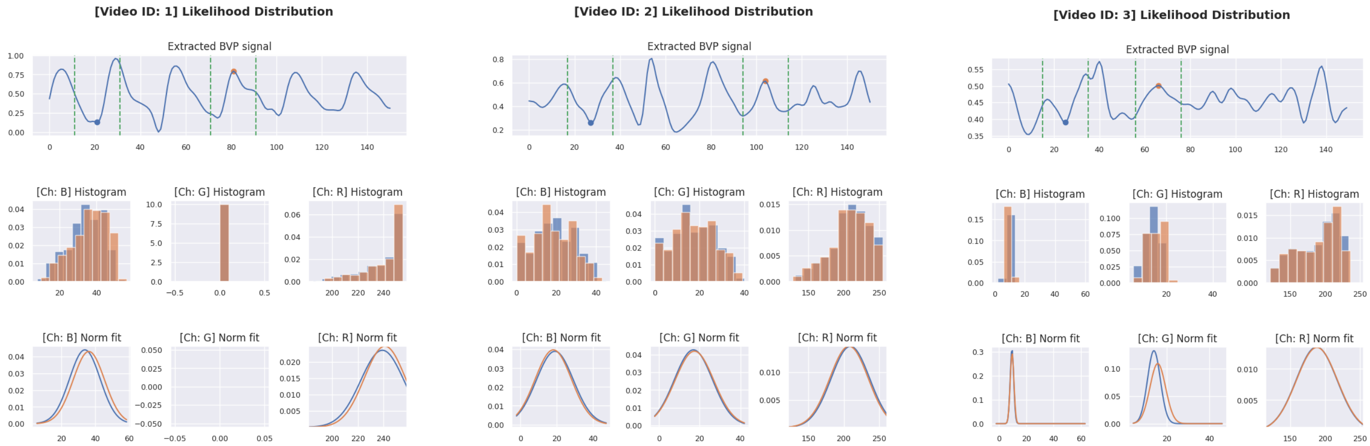
1) Observations:

- For all three datasets and almost all the three channels, there is a significant if not complete overlap between the distributions as seen in Fig 3. This is because even for one frame there is quite significant variation in pixel values in all 3 channels. Thus the information encoded in a single pixel is not enough to capture and classify the frame as a positive/negative event.
- After further analysis, it was found that taking spatial average over an entire frame and plotting histogram instead of considering all the pixels significantly improved the separability of distributions (Fig. 4). This may be because spatial averaging improves SNR in this particular use case, since we can intuitively imagine the average pixel intensity varying with the hearbeat signal proportionately, capturing the signal data with it.
- With regard to separability of distributions it is difficult to conclude because of the overlap in distributions. However in certain cases green channel consistently produced relatively separable distributions compared to red and blue channels (Fig. 3c).

E. Threshold Based Detection and ROC Curve

Only the R-channel is considered and out of each of the 40 (20 + 20) frames, 500 random pixels are chosen and based on the pixels values, a threshold based classification is done to classify the frame as Case 1 (positive event, local maxima) or Case 2 (negative event, local minima). The probability of detection (P_D) and probability of false alarm (P_{FA}) is computed and the ROC curve is plotted for various threshold values.

1) Preliminaries - The AUC score: The area under the Receiver Operating Characteristic (ROC) curve, called the AUC score, gives a statistical measure on the quality of the prediction model. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.



(a) (b) (c)

Fig. 3: (a) Likelihood distributions for Video 1, (b) Likelihood distributions for Video 2, (c) Likelihood distributions for Video 3

(a) (b) (c)

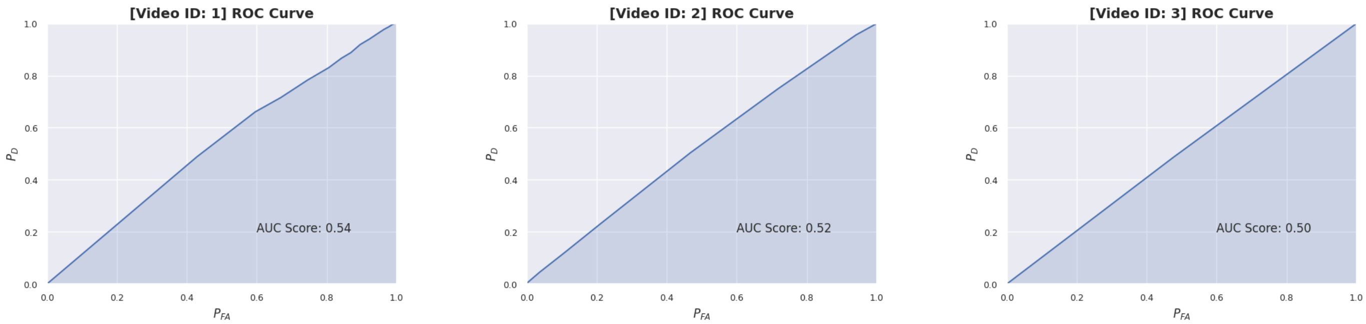
Fig. 4: (a) Likelihood distributions (spatial averaged) for Video 1, (b) Likelihood distributions (spatial averaged) for Video 2, (c) Likelihood distributions (spatial averaged) for Video 3

2) *Observations:*

- For all three datasets, ROC curves are almost entirely straight lines (Fig. 5, with an AUC score close to 0.5. This is expected since there is a significant if not complete overlap of likelihood distributions (Fig. 3).
- This means that the threshold based prediction model used for classification is worthless and no better than a random coin toss and that a single pixel value does not encode sufficient information for reliable classification.

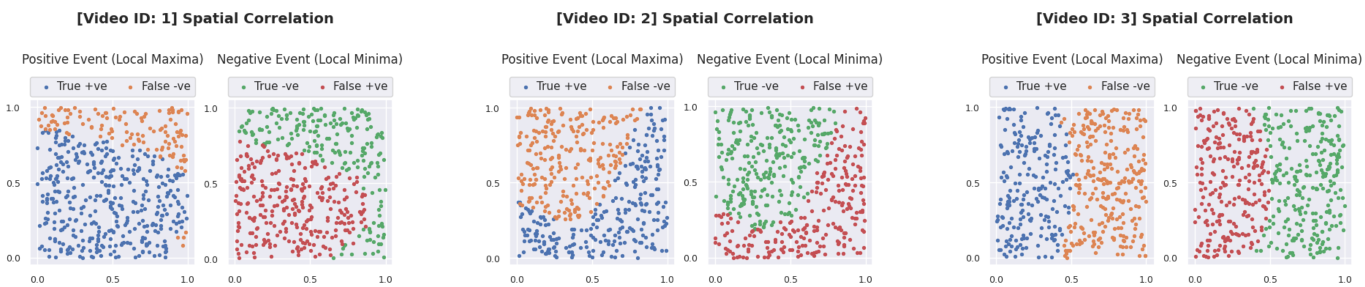
F. Spatial Correlation Analysis

An optimal threshold value is chosen that maximises P_D while minimising P_{FA}. For a given video sample, a random frame from the local maxima neighbourhood (positive event) and one from the local minima neighbourhood (negative event) is chosen. The spatial correlation between the ”good” samples (true positives and true negatives) and ”bad” samples (false positives and false negatives) is analysed.



(a) (b) (c)

Fig. 5: (a) ROC curve (AUC 0.54) for Video 1, (b) ROC curve (AUC 0.52) for Video 2, (c) ROC curve (AUC 0.50) for Video 3



(a) (b) (c)

Fig. 6: Spatial Correlation Analysis

1) Observations:

- There is a huge clustering tendency in the samples, i.e, the true positives and false negatives (in the case of a positive event) and true negatives and false positives (in the case of a negative event) are spatially clustered in certain areas of the frame.
- This may be because of intensity variation within a frame due to flashlight location, i.e, pixels close the flashlight are brighter than the ones farther away from the flashlight. As observed in Fig. 4, average pixel intensity in frame has higher SNR and captures the signal quite well, producing greatly separable distributions. Pixels values close to the average intensity are more likely to be classified correctly and due to the intensity variation within a frame, these ”more likely” pixels will be clustered together due to the non-uniform yet predictable intensity variation due to flashlight positioning.