Report for Assignment 1 of CS6650:JAN-MAY 2023

Name : Siddhanth Ramani Roll Number : CS23Z007

Link to ppgvideos :
https://drive.google.com/drive/folders/1LJfVwv_TfnbxOuFMye-Cw9GKIE-UXrGu?usp=sharing

Standard points across questions

Throughout the code, we used algorithms from Numpy, OpenCV and Scipy.
1. Numpy algorithms in general included mathematical functions.
2. OpenCV algorithms in general included Video/ Image reading and processing.
3. Scipy algorithms in general included mathematical transformations [3].

Specific code explanations - Everything is explained in the notebook
Algorithms used - Everything is explained in the notebook.

Question A

For question A, we created and understood the dataset.

Assumptions -
1. The finger covers the phone completely

Inferences and decisions made -

We understood the following :
1. Each video is stored as a list of frames with the length of list = number of frames. In the example above this is 365, 321 and 347.
2. Each frame is a numpy ndarray with shape as 1920, 1080, 3 (For 3 color channels B, G, R)
3. Each B, G, R pixel value is of type np.uint8 implying they can stores values between 0-255

We could also see that the number of frames captured was 364, 321 and 347. Since the camera captured at a frame rate of 30fps and we needed only a 10s sequence, we took only the first 300 frames into account for any further processing.

Question B

For question B, we analyzed the data to choose a sensing metric.

Assumptions -

1. The average heart beats between 60-100 seconds a minute while resting. This implies the heart beats about 1-1.67 times a second [2].

Inferences and decisions made -

We considered three different image types for choosing a sensing metric :

1. Considering the image as a whole
2. Considering only the B, G, or R components
3. Considering only the greyscale

When we plotted the statistical measures on these cases, we noticed that the information present is approximately the same.

And thus, we opted to go for the greyscale image to apply our sensing metric because
1. When compared to case a, this requires less memory and processing - reduces memory and compute requirement by a factor of 3.
2. When compared to individual B, G, R values - greyscale contains more stable information as it is a mixture of all B, G, R values.

Why is information from the greyscale image more stable?
In the code, we are looking at only one example, and in that example B, G, R seemed to capture equal amounts of information. But if in a scenario where we go with the R component for our predictions and a human somewhere has a green finger (maybe due to a disease), we might not have been able to capture the heartbeat (maybe Green would have been a more able choice for that scenario).

As greyscale is a weighted average of all B, G, R components [1], the information will be less biased and thus more stable.

We also identified that the min and median measures are not consistent across the cases, they seem to be quite random, whereas the information of max and mean appear more consistent. We thus focused only on the max and mean measures as consistent and trustable sensing measures.

To understand which metric to choose between these, we plotted an FFT of both the metrics and checked which one captured heartbeat frequency better.

When we plotted the two FFTs, we could see that
1. both mean and max are able to capture heartbeat well.
2. max is able to distinguish the heartbeat better than mean due to more noise (visual analysis).

But we can still choose to go with mean. This is because post 1Hz(range of human bpm), mean and max are performing equally well.
In that range, median will also return the frequency of heart beat as 1.49Hz.

And between median and max, max is more computationally expensive (requires dealing with memory which requires more compute cycles than an integer addition)

Our final chosen sensing metric was thus - The mean of the greyscaled image.

Question C

For question C, we plotted the intensity variations of the curve and the associated FFTs.

Assumptions - None

Inferences and decisions made -

We can observe from the above plotted intensity vs time graphs that the heartbeat frequency is a lot higher for the third sequence - the sequence which was recorded after a fast run.

We could also observe that the FFT for both the graphs for a 5 second sample is not that accurate. To increase accuracy, we had to look only at frequencies between the 1 (60bpm) - 3.67 (220bpm) - which is the usual heart beat frequency.

We observed then that if the above was the case, the second and third graphs had the most difference in frequency (heart frequency for video 2 will be around 1.2Hz and that for video 3 will be between 2-3bpm). This would be as per expectation.
A 1.2Hz heartbeat will correspond to a 1.2*60 = 72bpm.

Question D

For question D, we created a subset of the data and plotted the required histograms.

Assumptions - None

Inferences and decisions made -

We noticed that the plots for both the hypotheses were close to the same.
This implies that it will be very hard to find a threshold to separate them irrespective of which color dimension we choose for our analysis.

We could also see that for each color dimension in each sequence, the mean and std of the two cases were almost the same.
This implies that they will have highly similar distributions which is also what our plots signify.

We finally concluded that we cannot have a single threshold to effectively split our data.

Question E and F

Assumptions -

I did not understand question E as Question E stated the following,
"For every threshold value, for every frame, choose 500 random pixels. Compute the "Probability of Detection" (P_D) and "Probability of False Alarm""

But this, to me, wass confusing as the PD and PFA values are characterstic to the distributions as a whole.
And when it comes to dealing with individual samples and singular step thresholds, the output is not a probability but a mere 1 or 0, a "yes" or "no".

However looking at Questions E and F together, the only reasonable interpretation I can come up with is the below question

"How does the PFA and PD points (and thus the ROC curve) vary when we compare the distribution as a whole to that of a uniformly randomly sampled subspace? Will the sub-distribution have a better ROC curve? (where better implies threshold points which will result in a high PD/PFA ratio?)

This to me looked like a valid question which ponders upon whether there are some pixel values which are adding noise to the image and producing hard to separate hypothesis distributions.
If we do answer this question, and the answer affirms, we can arrive at that conclusion that maybe we can remove focus on those pixels to increase SNR and improve our accuracy.

This seemed like an appropriate question and I assumed my interpretation of the Question E and Question F were right.

Inferences made -

When we plotted the ROC curve for the “R” dimension but for all the pixels for all the frames, we found it to be quite linear indicating that PD/PFA value is constant. What this implies is that the distributions are overlapping which is what we observed while plotting this histogram.

When we plotted the ROC curve for the “R” dimension but for only 500 randomly chosen pixels, we found it to be linear again indicating that the PD/PFA value is constant.

This looked valid as taking uniformly random samples will be indicative of the same dataset and will not cause a major change in the distribution as whole.

We concluded that we cannot find a good singular threshold for such a plot as a good singular threshold is the one with max PD/PFA ratio and here the ratios are constants.

A final attempt

Hypothesis -
As I was not satisfied that we cannot differentiate between the distributions, I want to make one final attempt at splitting the distributions.
A logical hypothesis that I had is that the center of the image should have the least amount of noise and thus the highest SNR.
This is intuitively valid as the center part of the video seemed to hold the least amount of black pixels and also the location where some part of the finger of users will probably always be present.

Inferences and decisions made -
We noticed that performing this , the histogram distributions were looking a bit distinguishable.
Lanter, we could clearly see that the ROC curve is less linear when compared to all other previous cases. This is especially true for Recording 3.
For Recording 3, a reasonable threshold value "gamma", will thus be around 0.75/0.6 (an inflection point).

We finally inferred and concluded the below :

1. Some experimentations could help us deliver better singular threshold values.
2. A more dynamic thresholding mechanism might be a better way to classify heartbeats.

We also discussed possible next steps :

1. Sample at a higher frame rate - 60fps instead of 30fps
2. Having longer sampling duration - 10-20 seconds instead of 5 seconds
3. Creating a dynamic thresholding mechanism - maybe similar to scipy's find_peak function.

References

1. https://docs.opencv.org/3.4/de/d25/imgproc_color_conversions.html
2. https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979
3. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html