PROJECT REPORT

DS INTERNSHIP 2023

HEART RATE PREDICTION AND ANALYSIS

Mentor – Naman Solanki

BY – Rupsa Mitra

Abstract

Heart rate (HR) monitoring is a critical aspect of assessing an individual's physiological state and overall health. Traditional methods for HR estimation involve the use of wearable devices, which may be inconvenient or uncomfortable for extended monitoring periods. In recent years, there has been growing interest in exploring non-contact and unobtrusive methods to predict heart rate accurately.

This research proposes a novel Eulerian Video Magnification-based approach for heart rate prediction using facial videos. The core idea behind this method is the utilization of facial blood flow variations captured through video sequences as a proxy for heart rate estimation. The method relies on the fact that subtle color changes in facial regions, such as the forehead or cheeks, are related to the cardiac cycle.

This non-invasive and contactless heart rate estimation technique has the potential to revolutionize health monitoring systems, particularly in scenarios where traditional wearable devices are impractical or undesirable. Applications include monitoring patients in medical settings, assessing stress levels in real-time, and enhancing user experience in fitness and wellness applications.

While the proposed method shows promising results, further research is required to investigate its performance in more extensive and diverse datasets and to address any potential privacy concerns related to the use of facial videos for physiological monitoring. Nevertheless, the approach opens up exciting possibilities for non-intrusive heart rate monitoring in a wide range of settings, promising improved healthcare and well-being outcomes.

Research Summary

Introduction

In Heart Rate Estimation and Analysis, we worked on the Facial Video Dataset, Webcam and also Pure Database provided by National Taiwan University of Science and Technology. We have got satisfactory results on the datasets in different videos. The methods and details of our approach is written in the document.

Research

For the given project various principles have been studied such as Photoplethysmography (PPG) through Imaging, Remote Photoplethysmography (rPPG), Eulerian Video Magnification (EVM) and some of the deep learning techniques. Photoplethysmography (PPG) through imaging, specifically remote photoplethysmography (rPPG), gained attention in the early 2000s as researchers began exploring ways to estimate heart rate from video recordings of facial regions. Additionally, with the rise of wearable technologies and smartphones, interest in non-contact heart rate monitoring increased. Smartphone apps that claimed to measure heart rate using the smartphone's camera and flash emerged in the mid-2010s, further driving research in this area.

About the Databases

Webcam Database

1) It has 173 videos of in total 13 subjects
2) Mainly 5 different scenarios are considered in the dataset.
3) Normal Light Static (NS) All fluorescent lights were on. Each subject sit still without body movements.
4) Normal Light Moving (NM) Under the same illumination condition as of the above 1, each subject was directed to move his/her body to the right and left parallel to the camera repeatedly, with frequency kept between 0.2~0.5 Hz.
5) Dark Static (DS) All fluorescent lights were off with ambient lights primarily from windows and computer monitors. Each subject remained still.
6) Dark Moving (DM) Under the same illumination condition as of the above 3, each subject moved in the same way as of the above 2.
7) Normal Light Static Exercise (NSE) Under the same illumination condition as of the above 1, each subject was riding on an exercise bike with a constant speed.

PURE Database

This data set consists of 10 persons performing different, controlled head motions in front of a camera. During these sentences the image sequences of the head as well as reference pulse measurements were recorded. The 10 persons (8 male, 2 female) that were recorded in 6 different setups resulting in a total number of 60 sequences of 1 minute each. The six different setups were as follows:

Steady

The subject was sitting still and looks directly into the camera avoiding head motion.

Talking

Simulated video sequence, where the subjects were asked to talk while avoiding additional head motion. This setup equals a video conference situation in a real robot application.

Slow Translation

These sequences comprise head movements parallel to the camera plane. Therefore, the images recorded by the camera were displayed on screen and shown to the subjects. A moving rectangle of the size of the face was added to the image, and the subjects were asked to keep their face inside. The rectangle was moving horizontally at a controlled speed and with a predefined pattern, thus the sequences of all individuals are repeatable. The average speed was 7% of the face height per second, where the average face height was 100 pixels.

Fast Translation

This dataset has the same setup as slow translation, except twice the speed of the moving target.

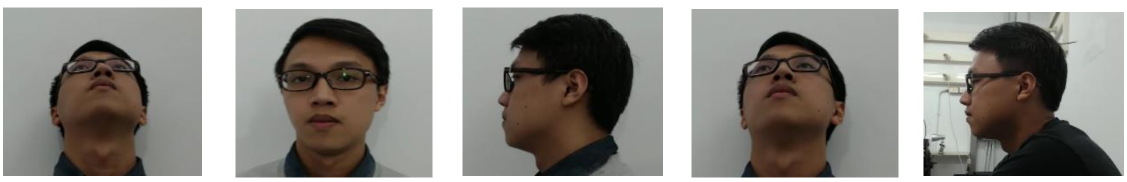
Small Rotation

This setup comprises different targets that were placed at 35 cm around the camera. The subjects were told to look at these targets in a predefined sequence. They were asked to move not only there eyes but orient their head. The one minute sequence of the targets is shown in the little clock in the figure. Random times ensure that the motion artifacts are not periodically. Depending on the distance between the camera and the subject, that roughly varies between 1 m and 1.3 m, the head rotation angles are round about 20°.

Medium Rotation

These sequences had the same setup as for small rotation, but with targets placed 70 cm around the camera resulting in average head angle of 35°.

From the description of the dataset it is very clear that different motions of the subjects were present in the videos.



EVM insights to projects

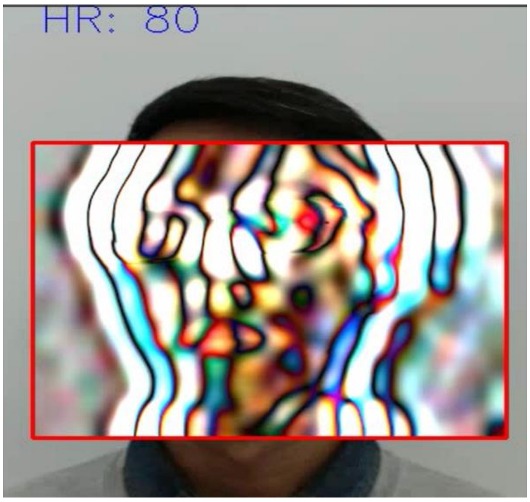
Eulerian Video Magnification is a computer vision technique used for amplifying subtle color and motion changes in videos. Originally introduced by MIT researchers in 2012, EVM can be utilized for various applications, including predicting heart rate from video datasets.

The basic idea behind EVM is to amplify the small, imperceptible color and motion variations in the video that are indicative of physiological processes like blood flow changes associated with the cardiac cycle. By magnifying these subtle changes, it becomes easier to observe and measure physiological signals, such as heart rate, from standard video recordings.

We have tried to develop a system using EVM to predict heart rate from video datasets using a webcam or a video file as input where we also thought of applies signal processing and other video magnification techniques to amplify subtle color variations associated with blood flow changes, which are indicative of the cardiac cycle.

OpenCV (cv2) was used for computer vision tasks and video processing then constructing a Gaussian pyramid for video frames and further reconstructs a frame from the pyramid. The define various parameters related to video dimensions and frame rates. Here we have applied bandpass filtering.

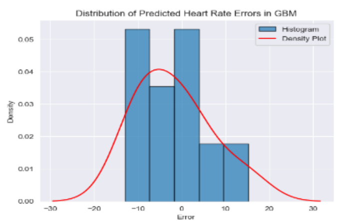
The EVM parameters are the number of levels in the Gaussian pyramid, amplification factor, frequency range for bandpass filtering, and buffer sizes for heart rate calculation. A Gaussian pyramid using the first frame from the video. The Fourier Transform of the Gaussian pyramid is calculated, and a bandpass filter is applied to retain the frequency components corresponding to the desired heart rate range. Further heart rate is calculated by identifying the dominant frequency in the filtered signal and converting it to beats per minute (bpm). The filtered signal is amplified, and the resulting frame is reconstructed from the Gaussian pyramid. The reconstructed frame is then overlayed on the original frame for visualization. The estimated heart rate is displayed on the output video.

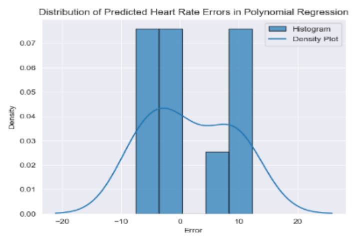


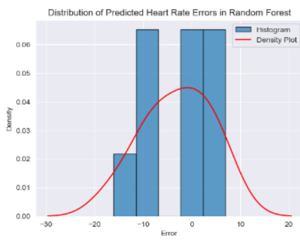
Results

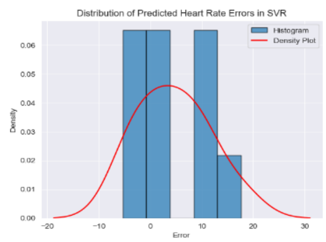
Firstly, we applied the Eulerian Video Magnification on all the facial videos and got an RMSE of around 15.53. Now we had one Ground truth value for each of the videos and hence the from Heart Rate values obtained through EVM were recorded and mean HR , Max HR was calculated for each of the subjects but RMSE for max HR was higher around 19 hence using mean HR into consideration the RMSE as around 15. Further work was to reduce this RMSE to around 4-5. By using the data of ground truth and mean HR the RMSE was reduced to the range of 6 through Linear /Non-linear (also bagging and boosting) models.

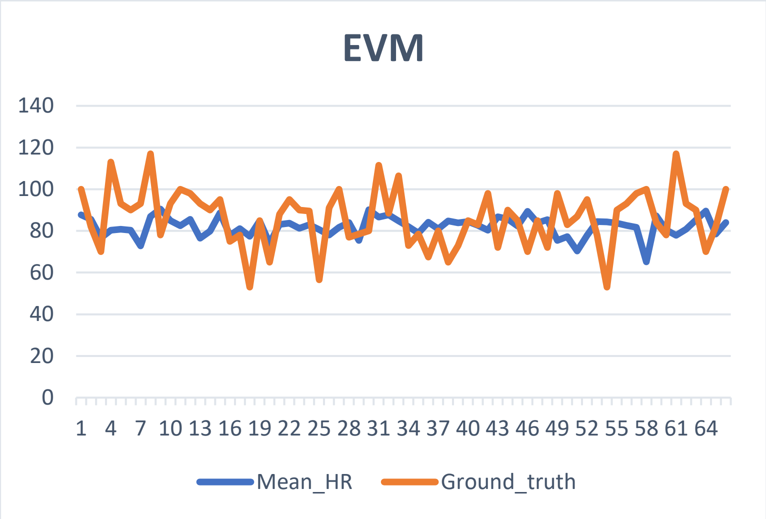
Regression Models	RMSE	MAE	R2 Score
Simple Linear	6.973	5.85	-0.064
Polynomial(2-degree)	7.012	5.835	-0.075
Multiple Linear	7.126	6.025	-0.11
Support Vector Machine	8.04	6.533	-0.41
Random Forest	7.399	5.972	-0.197
K Nearest Neighbours	6.88	5.628	-0.035
Decision Tree	7.507	6.074	-0.233





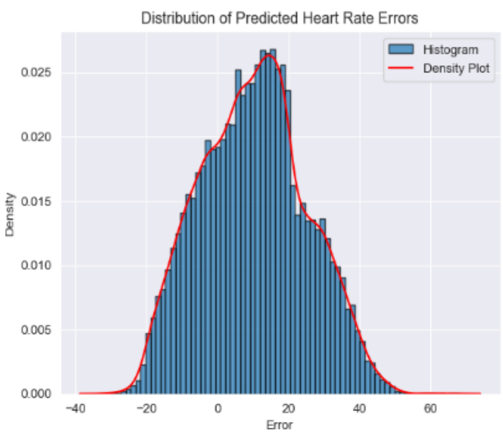


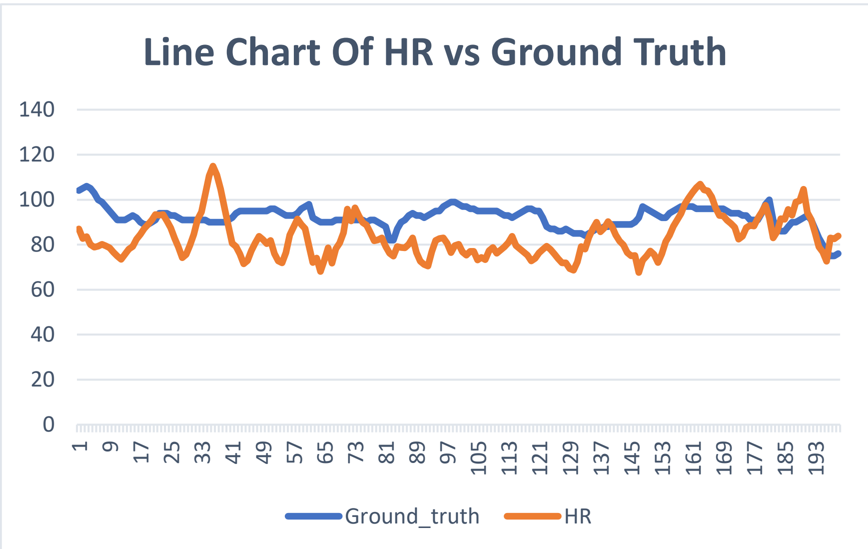




Nextly, we proceed to test the EVM approach on Webcam database for 155+ videos so the RMSE here was around 17.98. The RMSE could be reduced in the range of 13 using the similar above models as mentioned above.

Regression Models	RMSE	MAE	R2 Score
Simple Linear	13.41	11.299	0.019
Decision Tree	13.545	11.416	-0.0005
Polynomial	13.39	11.28	0.022
Support Vector Machine	13.568	11.343	-0.004
Random Forest	13.398	5.972	-0.197
K Nearest Neighbours	13.483	11.334	0.0086
CatBoost	13.418	11.317	0.018





The Deep-Learning Approach

We had PURE Database consisting of image sequences from 13 subjects. So, we have tried to establish a CNN model where we have given the input as the image sequences along with their Ground truth. The model was a 2-D CNN one. There exists many pretrained CNN models for Heart rate prediction using videos in the market but even those models have an RMSE in the range of 8 to 20.Those take facial videos as an input but this approach is purely based on image sequences from videos. To test a new video, we can use OpenCV library to convert the video into an image sequence one and then get the heart rates for those.

Some of the difficulties faced in this approach is that as the input has to be image sequences it takes a lot of time to train the model initially. Also, before training some initial preprocessing is done like normalization and resizing and then histogram equalization and standardization was also performed. The number of epochs was 10. The RMSE was around 10.

```
Root Mean Squared Error (RMSE): 10.741569802658127
Mean Absolute Error (MAE): 9.322350558855677
Coefficient of Determination (R^2): -0.0029436333220462796
```

As we have used only three image sequences here due to memory allocation error if more image sequences are used the model may be more robust. We have also tried to make the model more robust by adding more layers but in this case the loss is not decreasing as such hence we may think of some other approach such as cropping the background and applying EVM on the facial videos again if required. Also we can look up for some pre-trained Deep learning models like RhythmNet etc. which can take directly videos as an inputs. Parameter tuning is not done much hence parameter tuning may result in a more better model.

Conclusions

Through EVM by applying on different datasets with large number of videos we can get RMSE of around 10 to 13 and through Deep learning models also the results remain around 10 even though as we made a target to reduce the RMSE around 4 to 5.

REFERENCES

1. Stricker, R., Müller, S., Gross, H.-M Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot in: Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (Ro-Man 2014), Edinburgh, Scotland, UK, pp. 1056 - 1062, IEEE 2014
2. G. de Haan, V. Jeanne Robust pulse rate from chrominance-based RPPG
3. G. Balakrishnan, F. Durand, J. Guttag Detecting pulse from head motions in video
4. P. Viola, M.J. Jones Robust real-time face detection
5. L. Shan, M. Yu Video-based heart rate measurement using head motion tracking and ICA
6. S. Kwon, H. Kim, K.S. Park Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone
7. Macwan R., Benezeth Y., Mansouri A. Heart rate estimation using remote photoplethysmography with multi-objective optimization
Biomed. Signal Process. Control, 49 (2019), pp. 24-33
8. Niu X., Han H., Shan S., Chen X.VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video Asian Conference on Computer Vision, Springer (2018), pp. 562-576
9. S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J.F. Cohn, N. Sebe, Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2396–2404.
10. Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E.
Deep learning for computer vision: A brief review Comput. Intell. Neurosci., 2018 (2018)
11. Hsu G.-S., Ambikapathi A., Chen M.-S.Deep learning with time-frequency representation for pulse estimation from facial videos 2017 IEEE International Joint Conference on Biometrics, IJCB, IEEE (2017), pp. 383-389
12. Yu Z., Li X., Zhao G. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal network (2019)
13. Niu X., Han H., Shan S., Chen X. Synrhythm: Learning a deep heart rate estimator from general to specific 2018 24th International Conference on Pattern Recognition, ICPR, IEEE (2018), pp. 3580-3585
14. X. Li, J. Chen, G. Zhao, M. Pietikainen, Remote heart rate measurement from face videos under realistic situations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4264–4271.