Contactless Heart Rate Estimation in Humans using Low Cost Face Video

Gandharv Mohan
IIIT-Delhi
gandharv17232@iiitd.ac.in

Jayant Jain
IIIT-Delhi
jayant17155@iiitd.ac.in

Mahika Wason
IIIT-Delhi
mahika16241@iiitd.ac.in

Abstract

In this project, we attempt to measure heart rates in humans using camera-based remote photoplethysmography (RPPG) methods, named after traditional PPG. The fundamental idea is based on capturing minute changes in skin color during a cardiac cycle of the human body, involving the inflow and outflow of blood from the heart to other body parts. We have compared the performance of different methods of Blind Source Separation and face detection which form an integral part in accurately calculating the heart rate.

1. Introduction

Heart Rate is a crucial physiological measure and directly relates to the physical well-being of a person. While HR measurement devices using skin contact are fairly common, there is a growing need for contact-less means for HR measurement in applications like the patient monitoring, new-born HR measurement, driver health, employee well-being among numerous others. It is important that the heart rate is accurately monitored even in the presence of natural motion of the head.

2. Related Work

Heart Rate Estimation from facial videos has been a fairly active research area in the past decade with many researchers presenting various improvements to the previously used methods to make the algorithms more stable and robust in real life conditions with minimal constraints. Blind Source Separation has been used as an integral part of the algorithm since it was demonstrated to significantly improve the performance accuracy [2]. Two algorithms have been widely used to calculate the source signals from the observed signals- Independent Component Analysis(ICA) and Principal Component Analysis(PCA). These two algorithms give similar results and neither is preferred over the other.
Selection of a suitable ROI can lead to an improvement of the estimation accuracy. Traditionally, the work of Viola and Jones [4] has been used for detecting the face of the subject for extracting the signals. This work is also the basis of the Haar cascade classifier in OpenCV. Recently, Convolutional Neural Networks have been used for face or skin segmentation [3]. This has also led to increased resistance to background noise because CNNs give a pixel level mask and try to completely eliminate the background noise which may lead to inaccurate results.

3. Methodology

The methodology consists of 2 main steps. Firstly, the videos in the dataset is run through a face detection model to get the region of interest for heart rate calculation. This is followed by dividing the detected ROI into RBG channels and applying Source Separation(PCA/ICA) algorithm to obtain the source signals. Source signals are converted to frequency domain for filtering and peak detection to obtain heart rate estimates.

3.1. Finding region of interest

It is useful to work on only the skin regions of the video and remove the rest of the background as this increases the SNR(signal to noise ratio). It is also important that the ROI detection algorithm is robust enough to detect the face of the subject in all different orientations during the natural movement of the head. We have explored three techniques to find the ROI.Open CV Haar cascade face detection. [Fig 1], Face segmentation using CNN (LinkNet) [Fig 2] and Faster R-CNN.

3.1.1 Open CV Haarcascade

It is an object detection and classification pre-trained model provided by open CV. It is a Haar feature-based cascade classifier and we use it in our project to detect face of the subject. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in test images.It uses Adaboost which selects the best features and trains the

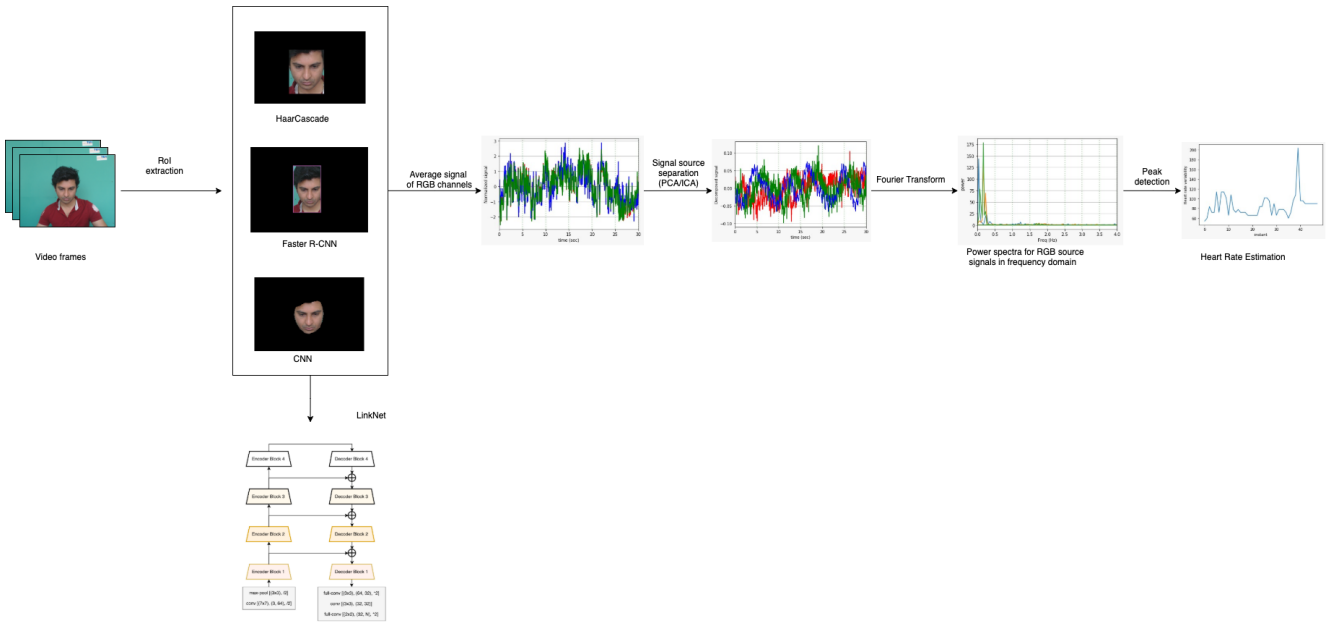

Figure 1. Heart rate extraction methodology

classifiers that deploy them. It constructs a "strong" classifier as a linear combination of weighted simple "weak" classifiers.

3.1.2 Face segmentation using CNN

We use a pre-trained PyTorch model of LinkNet to detect facial region of the subject. LinkNet architecture is represented in the Fig 2 which consists of encoder-decoder layers. The decoder layer uses full convolution and the encoder layer uses convolution with a filter of 7x7 and a stride of 2. This is followed by spatial max pooling of 3x3 with stride 2.

3.1.3 Faster R-CNN

Faster R-CNN is a popular algorithm used for Object detection in images. It consists of two main networks- Region Proposal Network(RPN) and a detection network. RPN proposes regions on the basis of feature maps it receives from a neural network(Here we have used VGG16). The second networks evaluates each of the proposed region and uses softmax function for the final classification. We trained the model to detect faces on the Caltech Faces(1999) Dataset `http://www.vision.caltech.edu/html-files/archive.html`.

3.2. Source signal separation and heart rate estimation

The RGB channels of the ROI are normalised and the mixed signal is then decomposed into it's source signals

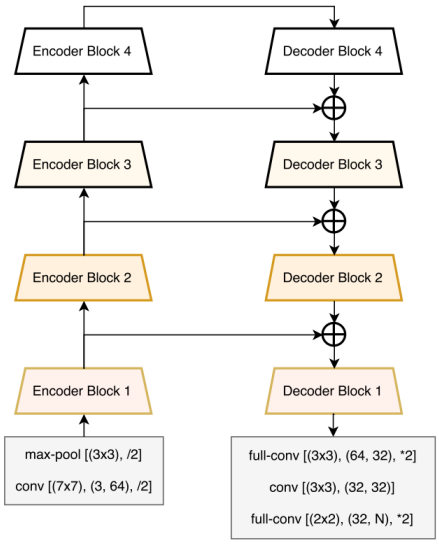

Figure 2. LinkNet architecture

using either Independent component analysis or Principal component analysis. While both the techniques lead to the discovery of basis/source signals, there are subtle differences.

After signal decomposition, the source signals are then converted into frequency domain. The measured heart rate will be the frequency of the peak with the highest peak and which is in the range of 0.5Hz to 4Hz.

$$x(t) = Bs(t) \tag{1}$$

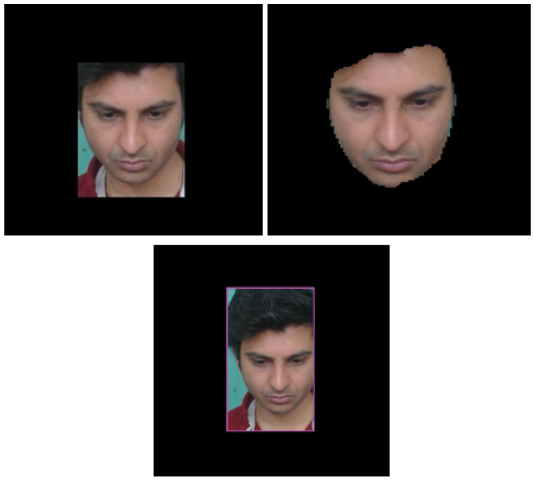

Figure 3. ROIs extracted using Haar cascade, Linknet and Faster R-CNN respectively

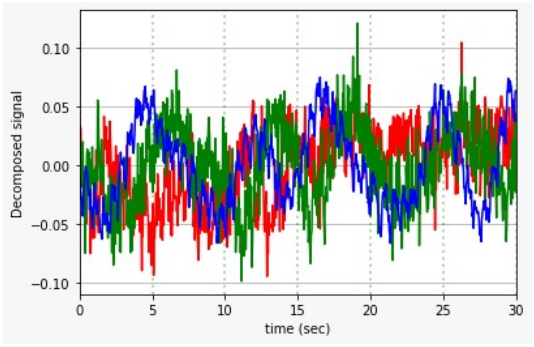

Figure 4. Source Signals from Independent Component Analysis

$$s(t) = B^{-1}x(t) \tag{2}$$

where,

$$x(t) = [x_{r(t)}\ x_{g(t)}\ x_{b(t)}]^T$$

is the vector of mixed signals,

$$s(t) = [s_{1(t)}\ s_{2(t)}\ s_{3(t)}]^T$$

is the vector of source signals, and B is a 3x3 matrix of coefficients representing the combination of source signals that form the mixed signal which we have in the beginning.

3.2.1 Independent Component Analysis

ICA always finds independent source signals from a mixed signals which is very often useful in finding independent sub-signals of a given signal.

3.2.2 Principal Component Analysis

PCA finds very specific source signals that provide the principal direction of variability. The basis vectors are orthogonal to each other but may not be independent. PCA is extremely helpful in dimensionality reduction.

Figure 5. Sample frames from the Dataset

3.3. Dataset

The UBFC-RPPG dataset [1], which is available on request from the creator, has been used for training and testing the implemented models. It contains 42 videos of around 1 min each captured using a low-cost webcam (Logitech C920 HD Pro) at 30fps with a resolution of 640x480 in uncompressed 8-bit RGB format. The subjects had to sit at a distance of around 1m from the camera and play a time-sensitive mathematical game aimed at augmenting their heart rate. The ground truth values were measured using a pulse oximeter.

4. Results

Subjects	Actual Heart Rate	Heart Rate using ICA	MAE for ICA	Heart Rate using PCA	MAE for PCA
Subject 5	98	99	1	84	14
Subject 25	91	94	3	82	9
Subject 42	94	94	0	99	5
Subject 44	82	90	8	95	13
Subject 45	111	107	4	92	19

Table 1. Heart Rate and MAE using Haar cascade for face detection

Subjects	Actual Heart Rate	Heart Rate using ICA	MAE for ICA	Heart Rate using PCA	MAE for PCA
Subject 5	98	100	2	100	2
Subject 25	91	90	1	89	2
Subject 42	94	98	4	74	20
Subject 44	82	79	3	69	13
Subject 45	111	110	1	108	3

Table 2. Heart Rate and MAE using CNN for face segmentation

	Haar cascade	CNN Link Net
PCA	12	8
ICA	3.2	2.2

Table 3. Average error rates for CNN/Haar cascade for face detection and PCA/ICA for signal decomposition.

4.1. Evaluation Metric

We have used Mean Absolute Error to compare the accuracy in beats per minute of the models.The ground truth values were extracted from the pulse-oximeter while the video was being recorded.

4.2. Analysis

- The prediction accuracy changes with change in sliding window size during the analysis. After hit-and-trial, we decided on making window sizes based on FPS (frames per second) analysis of the video in the beginning of our prediction methodology .

- Any outliers indicate that the algorithm most likely failed to pick up the pulse as one of the source signals and selected some other signal as its dominant frequency. These outlier measurements may be any random frequency within the acceptable bracket and therefore using them in calculations does not indicate how closely the algorithm matches the reference heart rate.

- The average mean absolute error using CNN for face segmentation is lower than using Haar cascade for face detection for both PCA and ICA algorithms in source signal decomposition. (Table 3) This is because the Link Net returns a contour of the face at a pixel level in contrast to Haar cascades which return a bounding rectangular box.

- The average mean absolute error is lower when using Independent component analysis, in comparison with Principal component analysis.(Table 3)As you can see,comparing PCA predictions of two models, error is reduced to +- 8 BPM in case of CNN based segmentation as compared to +- 12 BPM of Haar cascade model.

- We trained the faster R-CNN ourselves and found it is robust to head move tilt and its sideways movement which is not achieved by ea

5. Limitations

- Faster RCNN is taking a very long time to classify (2 secs to process 1 frame). Due to this, we could not test it on multiple subjects' videos and compare with other methods.

- CNN took longer than Haar Cascade for region of interest detection.

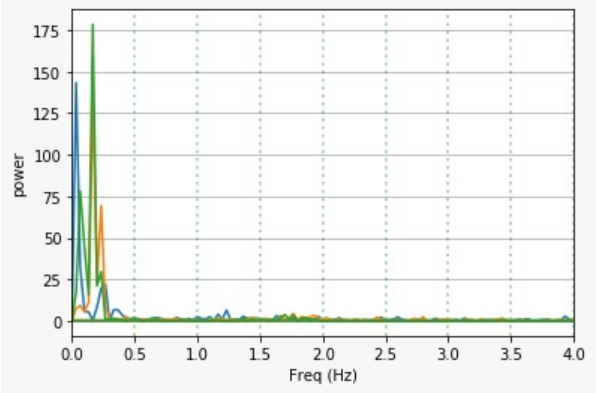

Figure 6. Power Spectrum of the Signal

6. Conclusion

As we can see from the results, Independent Component Analysis performs better than Principal Component Analysis in most cases to get the sources signals from the observed signals for heart rate estimation. Face segmentation using Convolution Neural Network gives better results than the Haar Cascade OpenCV face detection module, which is as expected. But the tradeoff is CNNs are slower to detect faces than the Open-CV module. Choosing an ROI by segmenting out facial pixels helped to keep the outliers low and therefore increased the robustness of our algorithm.

References

[1] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 2017.

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

[3] C. Tang, J. Lu, and J. Liu. Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1390–13906, June 2018.

[4] Paul Viola, Michael Jones, et al. Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, 1(511-518):3, 2001.