T.C.

MARMARA UNIVERSITY

FACULTY OF ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

CSE4197 Engineering Project I

Analysis and Design Document

rPPG BASED HEART RATE ESTIMATION USING DEEP LEARNING

Group Members

Nur Deniz ÇAYLI 150116071

Minel SAYGISEVER 150116028

Esra POLAT 150116884

Supervisor

Prof. Dr. Çiğdem EROĞLU ERDEM

INDEX

1. Introduction

1.1 Problem Description and Motivation

Heart rate estimation has great importance in determining a person's mental and physiological state. In some cases, it is not possible to use medical devices such as the finger pulse oximeter with photoplethysmography (PPG) technology due to the patient's delicate health and skin conditions. For example, such a technology is needed for continuous monitoring of premature infants. In such a case, it is necessary to measure the heart rate remotely. Remote PPG studies (rPPG) bring us a solution in such sensitive situations, allowing us to estimate the heart rate through a face video obtained with a standard webcam.

We can see an illustration of the main principles of rPPG in Figure 1. With every heartbeat, there are changes in the light and hence colour reflected from our skin caused by the cardiac cycle. We cannot see these changes with our eyes, but we can analyze the intensity of these colours with image processing techniques. If we can get the RGB values of the skin pixels in the frames of a video, we will have 3 colour signals. By processing these signals, we can estimate the heart rate.

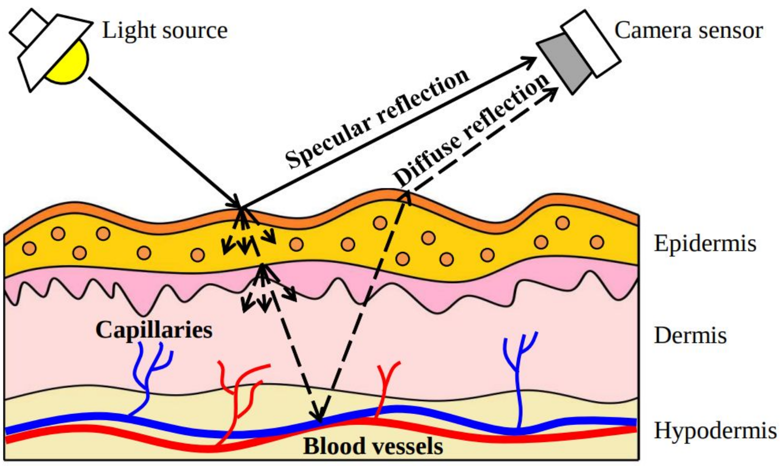


Figure1: Skin reflection model illustration [1]

1.2 Scope of the Project

There are some requirements for the project to give an accurate result. We process the signals from the skin pixels of the face captured by the camera and these skin pixels must be visible to the camera.

Our assumptions and constraints are:

- The environment must be well-lit.
- The subject's head movements should be minimal.
- The face of the subject should be visible, so the subject should not turn his/her head.
- The subject's facial expressions should be minimal.
- The subject should have minimal facial hair.

We will develop and implement a deep learning-based algorithm for heart rate estimation using rPPG and compare with signal processing based methods in the literature. We will implement the deep learning-based method based on DeepPhys[4] and PhysNet[5] frameworks. DeepPhys is an end-to-end network that can accurately recover heart rate from RGB signals. Physnet is another rPPG signal measurement framework using spatial-temporal networks. While we work on these frameworks, we will use two public datasets which are PURE [2] and UBFC-rPPG [3].
Our main goal is to measure the heart rate most accurately. If the root mean squared error (RMSE), between the estimated heart rate and the actual heart rate, is 5 beats per minute or less, we consider the result as successful.

1.3 Definitions, Acronyms and Abbreviations

Definitions:

- HR is the normal beating rate of the heart for adults which ranges from 60 to 100 beats per minute (BPM) during rest. During exercise, the heart rate may increase up to around 180 BPM.
- HRV is a measure of variability between heartbeats.
- PPG is a technique used to measure volumetric changes of the blood in the vessels.
- rPPG is a method to estimate the heart rate of a person without skin contact.
- Pulse oximetry is an instrument used to measure the oxygen saturation in a person's blood.

Acronyms and abbreviations:

- HR - Heart Rate

- HRV - Heart Rate Variability
- PPG - Photoplethysmography
- rPPG - Remote Photoplethysmography
- ROI - Region of Interest
- IBI - Interbeat Interval
- LF - Low Frequency
- HF - High Frequency
- RMSE - Root Mean Square Error
- BVP - Blood Volume Pulse

2. Related Work

The main subject of our project is to measure the heart rate of the person without touching the person. We do this with rPPG technology, so we measure heart rate with an RGB camera. We estimate the heart rate by selecting the region of interest on the skin and inferring the rPPG signal from the colour changes.
Below we give a brief overview of the methods in the literature. We group the methods as contact and remote (non-contact) methods.

2.1 Contact Methods

2.1.1 Photoplethysmography (PPG)

Figure 2: Finger pulse oximeter [12]

Photoplethysmography (PPG) is a technique used to measure the volumetric changes in the blood affected by the heartbeat. PPG is usually obtained using pulse oximetry to measure the heart rate. A normal pulse oximeter monitors the circulation of blood in the dermis layer under the skin. With each cardiac cycle, the heart pumps blood. Even though this pressure pulse is somewhat damped by the time it reaches the skin, it is enough to distend the arteries and arterioles in the subcutaneous tissue. The change in volume caused by the pressure pulse is detected by illuminating the skin with the light from a light-emitting diode (LED) and then measuring the amount of light either transmitted or reflected by a photodiode. Each cardiac cycle appears as a peak.

PPG technology is also used in other applications. For example, blood oxygen saturation, blood pressure, cardiac output, respiration, vascular assessment, arterial disease. Additionally, the shape of the PPG waveform differs from subject to subject and varies with the location and manner in which the pulse oximeter is attached as we can see in Figure 2.

2.2 Remote Methods

2.2.1 Independent Component Analysis (ICA) Method

The ICA [9] tries to separate a multivariate signal into independent non-Gaussian signals. For example, an audio signal occurs the numerical addition, at each time t, of signals from different sound sources. In this signal, the problem is whether it is possible to separate these subscripts to sources from the observed entire signal. If the statistical independence assumption is right, blind ICA separation of a mixed-signal gives very good outcomes. Also, ICA can be used for signals that are not needed to be generated by mixing for analysis purposes. We can see these processes in Figure 3.

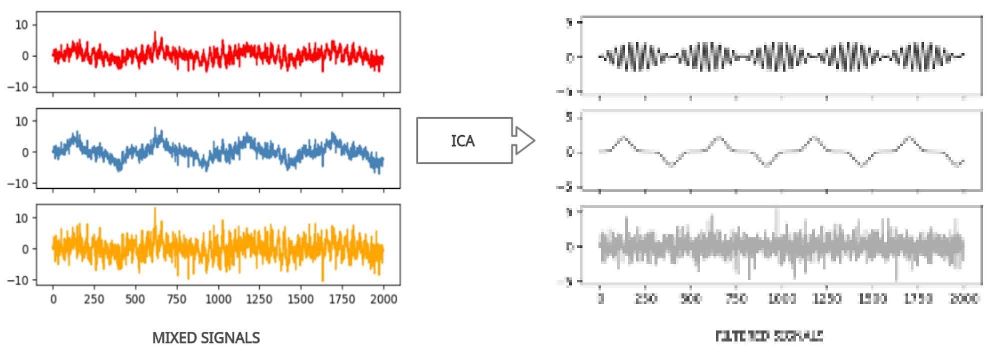


Figure 3: ICA working example

In the ICA model formulas $x_1(t)$, $x_2(t)$ and $x_3(t)$ red, green and blue signals. Source signals are represented by $s_1(t)$, $s_2(t)$ and $s_3(t)$.

$$x_i(t) = \Sigma a_{ij} s_j(t) \text{ for each } i = 1,2,3 \tag{1}$$

$$x(t) = As(t)$$

the column vectors $x(t) = [x_1(t), x_2(t), x_3(t)]T$, $s(t) = [s(t), s(t), s(t)]T$.

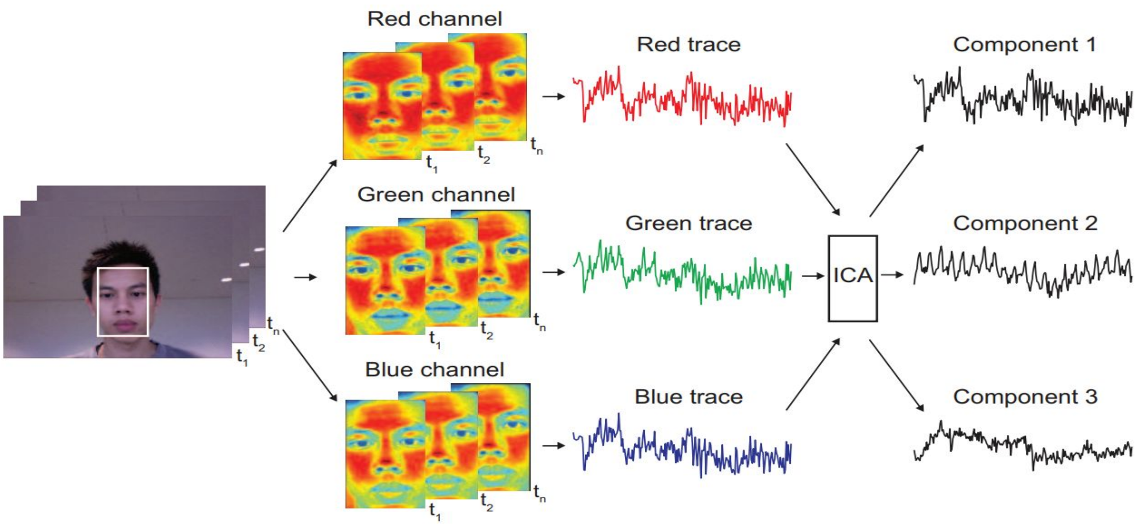


Figure 4: A diagram showing the processing of two source signals with ICA [13]

2.2.2 Chrominance-Based (CHROM) Method

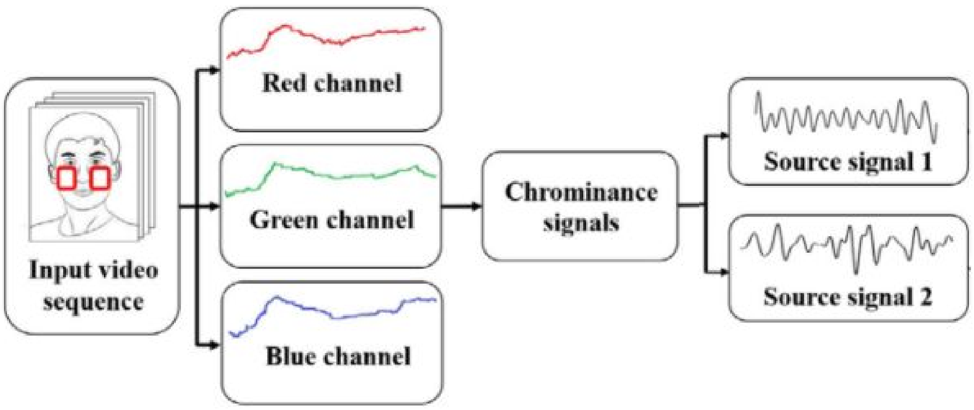


Figure 5: CHROM schema [14]

CHROM [8] signal processing method allows obtaining the pulse signal in case of specular and motion artefacts. RGB channels are reflected in a chrominance subspace. Here the movement component is largely eliminated. The CHROM method creates a vector using a standard skin tone. It obtains the pulse signal using an alpha setting. However, these settings sometimes may not match the actual situations and as a result, the method may fail. We can see the schema of CHROM in Figure 5.

2.2.3 Green - Vercruysse (GREEN) Method

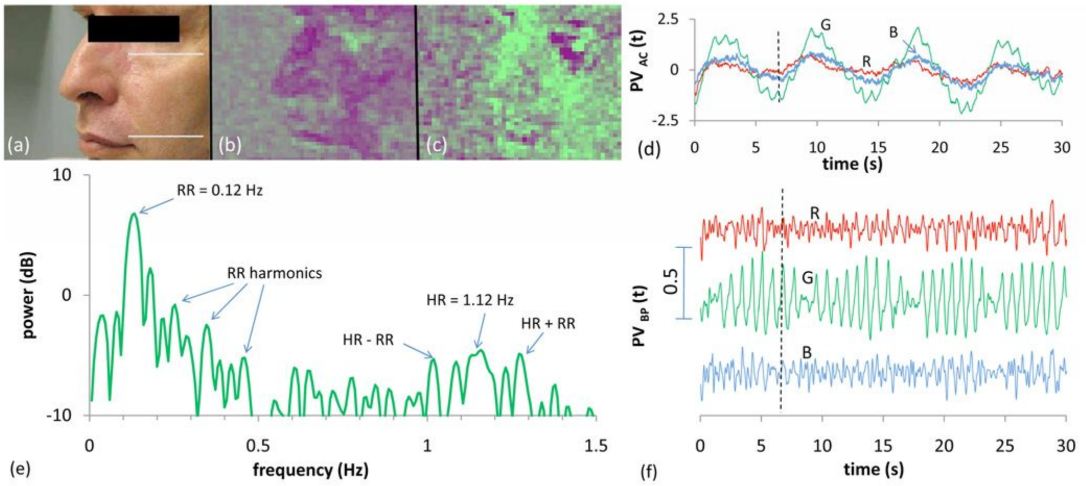


Figure 6: An example of pulse amplitude modulation [7]

According to the premise of the GREEN [7] method, the green channel contains the powerful plethysmographic signal, consistent with the fact that hemoglobin absorbs green light better than red and on the other hand passes through sufficiently deeper into the skin as compared to blue light to study the vasculature. This method used Fourier transforming for filtering. For steps, we can look at Figure 6.

2.2.4 Plane-Orthogonal-to-Skin (POS) Method

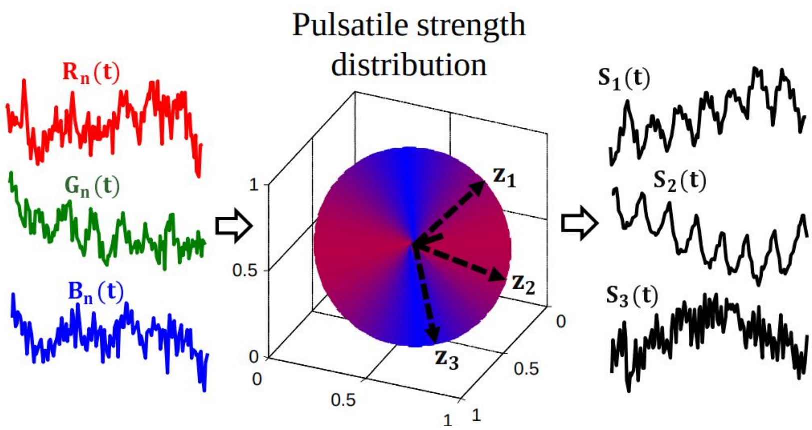


Figure 7: The distribution of the pulsatile strength on the plane orthogonal to 1 as a function of z [10]

POS [10] primarily makes skin-detection and only takes signals from the skin. The POS algorithm suggests adding the 2SR property to the model. 2SR or data-driven method is a new development. It creates a subject dependent skin-colour area and tracks the colour-change over time to measure the pulse, also the sudden colour is determined depending on the statistics of the skin pixels. From Figure 7, we can see that the projection direction is highly related to the pulsatility that determines the signal quality, different z may give very different projected-signals.

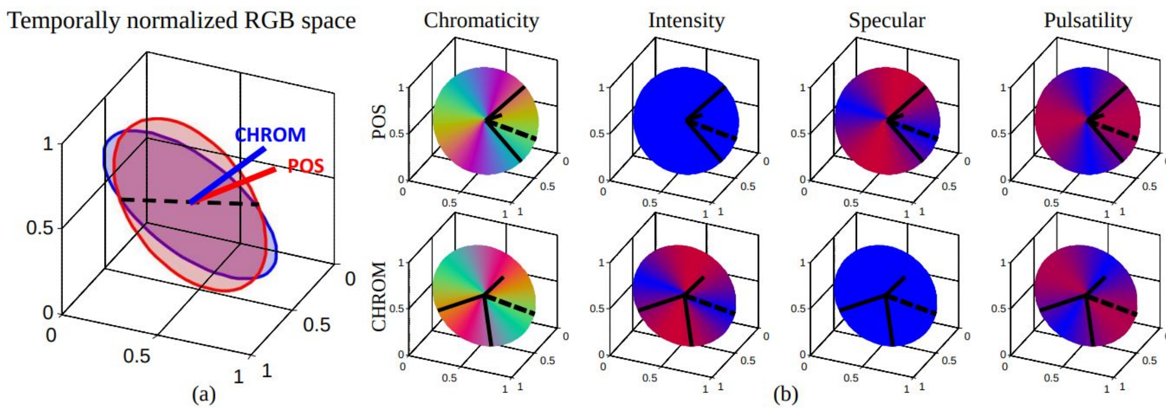


Figure 8: (a) The projection planes of POS and CHROM in the temporally normalized RGB space. (b) The projection planes of POS and CHROM have different chromaticity distributions

As we can see in Figure 8, POS and CHROM [8] have different distributions of volume and reflective variations. In this context, the solid black line shows the primary normal vector and projection axes in both. So, we can say that both have different advantages and disadvantages.

3. System Design

3.1 System Model

3.1.1 Traditional Methods

We use Viola-Jones [6] face detection technique to automatically detect the face of the subject. This step provides bounding box coordinates defining the subject face.
In the Viola-Jones algorithm, handmade simple Haar features are first created. Then the image is converted into an integral image. The integral image is the calculated

version of the source image. Each point in the integral image is the sum of pixels above and to the left of the corresponding pixel in the source image.
However, instead of making additions for each pixel value for all features - an integrated image is used to take advantage of several subtractions to achieve the same result.

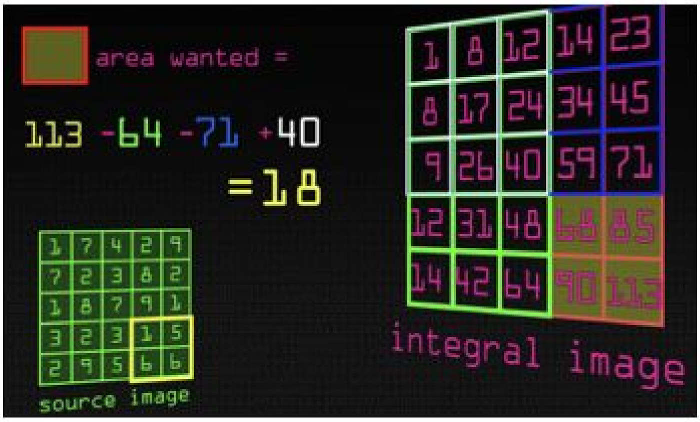


Figure 9: Using the integral image for the area wanted calculation [15]

After that, the delta values for each feature on an image region are calculated and a machine learning meta-algorithm, Adaboost, is trained for each feature. A classifier is created for each Haar feature. And these classifiers are considered "weak" classifiers. A weak classifier is trained for each feature using AdaBoost.
When the training is complete, models are sorted by error rate and select the best weak classifiers based on a threshold value and useful classifiers are added to the attentional cascade. Attentional cascade is a set of weak classifiers that are trained when used together to make a powerful classifier. After that, the cascade is loaded and the image is gradually passed through each classifier and the result is obtained.
After detecting the face, we need to make skin detection and remove non-skin pixels. The skin detection is performed on every frame to filter out non-skin pixels. The area of interest, the skin part, is our ROI [1] piece. The pixels in the ROI are spatially averaged, the process repeated for each video frame. The result of this process is then used to obtain the rPPG signal.

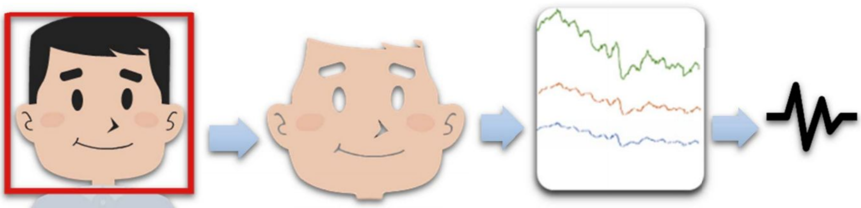


Figure 10: Obtained rPPG signal from ROI [16]

For HR extracting, FFT was applied to the selected signals and their power spectrum was obtained. The frequency corresponding to the highest power of the spectrum in an operational frequency band is determined as the pulse frequency.

3.1.2 Deep Learning Based Methods

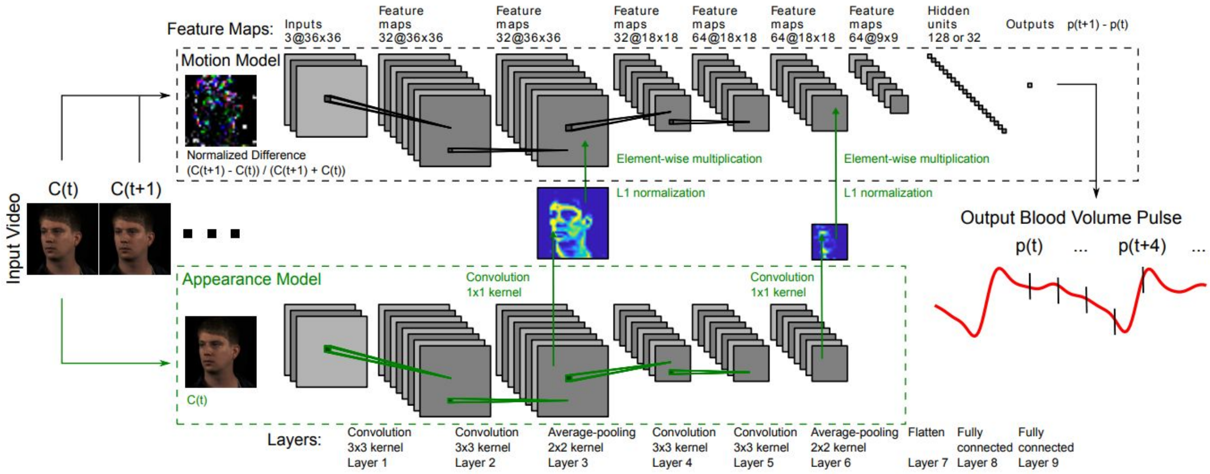


Figure 11: The illustration of DeepPhys architecture from the original paper [4]

We plan to carry out our measurements with deep learning methods, which is our main approach. We hope that deep learning will reduce error rates as a result of these measurements.

DeepPyhs [4] provides visualization of physiological information in videos using convolutional attention networks. It processes RGB or infrared videos, and can accurately obtain heart rate.

The main purpose of PhysNet [5] approach is using a Spatio-temporal network for rPPG signals from videos. Then it compares rPPG signals with Ground Truth ECG values. The method makes peak detection to find Interbeat Interval for Average Heart Rate and Heart Rate Variability.

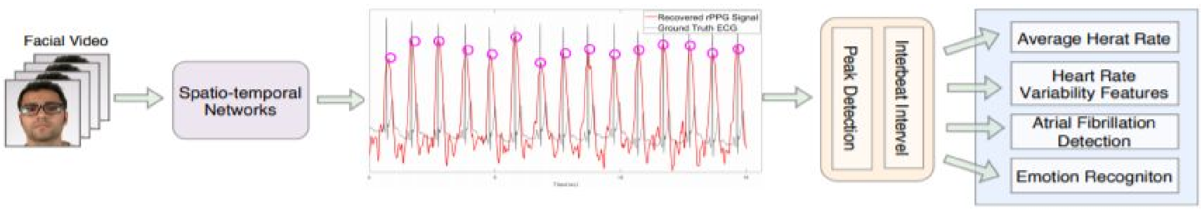


Figure 12: rPPG signal measurement using Spatio-temporal Networks [5]

3.2 Flowchart of Proposed Algorithms

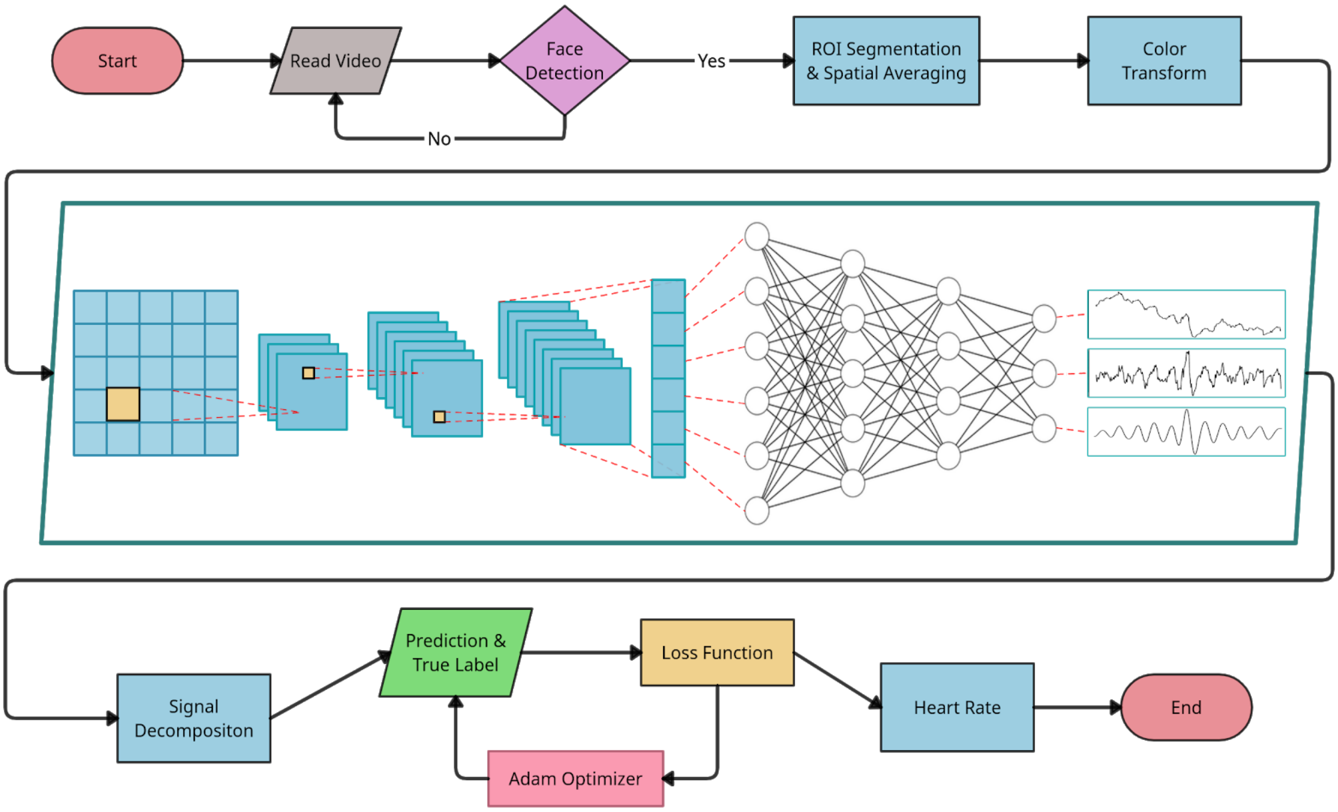


Figure 13: Flowchart of proposed algorithm

3.3 Comparison Metrics

Root Mean Square Error (RMSE) is the standard deviation of the prediction errors. Prediction errors call as residuals. Residuals can be measured by how far data points are from the regression line. The RMSE value is a measure of how far these residues have spread.

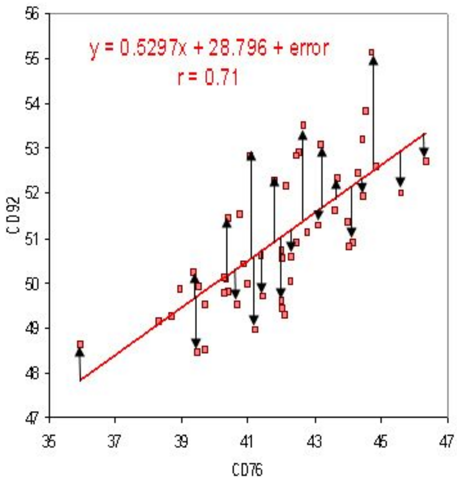


Figure 14: Residuals on a scatter plot [17]

RMSE value can be calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}} \tag{2}$$

Mean absolute error (MAE) is a measure of errors between observations expressing the same phenomenon. MAE is calculated as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{3}$$

Signal-to-Ratio (SNR) is defined as the ratio of signal power to the noise power and calculated as follows:

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{4}$$

The Mean Signal-to-Noise Ratio (MSNR) is a kind of matrix eigenvalue parsing method. Constructs the SNR function, predicts the separation matrix by eigenvalue decomposition or generalized eigenvalue decomposition. With this algorithm, the closed-form solution can be found without the iterative optimization process. MSNR can be calculated as follows:

$$MSNR = \frac{1}{N} \sum_{k=1}^{N} \left\{ 10log_{10} \left(\frac{S_k(f = f^*)}{\Sigma_{f \in F^*} S_k(f)} \right) \right\} \tag{5}$$

The correlation coefficient (r) is a measure of how close the points on a scatter plot are to the linear regression line. The correlation coefficient can be calculated as follows:

$$r = \frac{Cov(X, Y)}{\sqrt{s_x{}^2 s_y{}^2}} \tag{6}$$

where $Cov(X, Y)$ is the covariance and can be calculated as follows:

$$Cov(X, Y) = \frac{\sum (X - \underline{X})(Y - \underline{Y})}{n - 1} \tag{7}$$

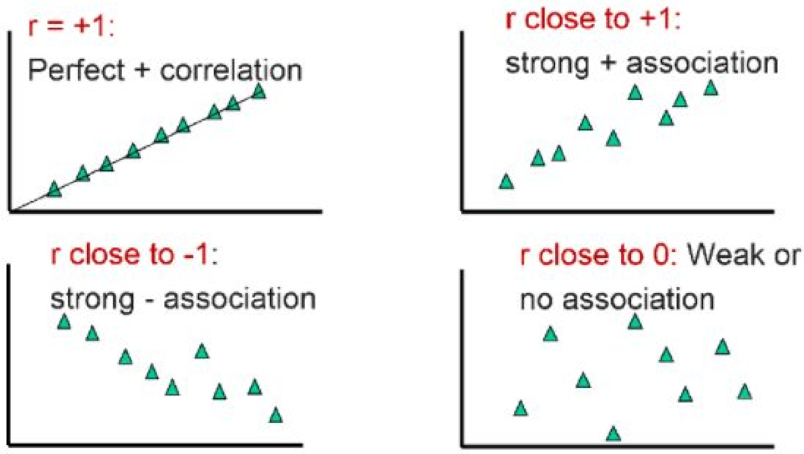


Figure 15: Example scatter plots for correlation coefficient [18]

3.4 Data Sets

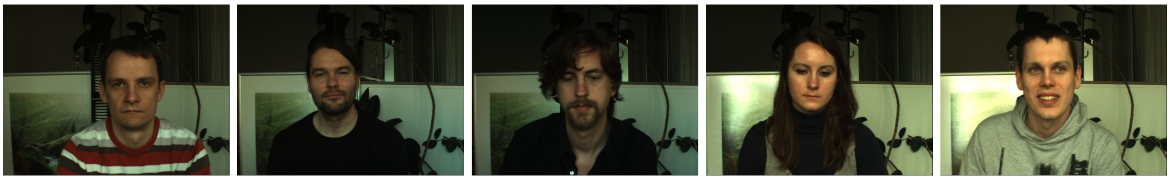


Figure 16: A few examples from PURE dataset [2]

Pulse Rate Detection Dataset - PURE [2] data set consists of 10 persons (8 male, 2 female) that were recorded in 6 different setups:

- Head is steady
- Talking without head movements
- Slow translation
- Fast translation - twice slow translation speed - average speed was 7% of the face height per second
- Small head rotation up to 20°
- Medium head rotation up to 35°

So there is a total number of 60 sequences of 1 minute each. We can see a few example frames in Figure 16. The image sequences of the head and the reference pulse measurements were recorded. The videos were captured at a frame rate of 30 Hz with a cropped resolution of 640x480 pixels and a 4.8mm lens. Reference data were captured using a finger clip pulse oximeter that provides pulse rate wave and SpO2 readings with a 60 Hz sampling rate. [2]

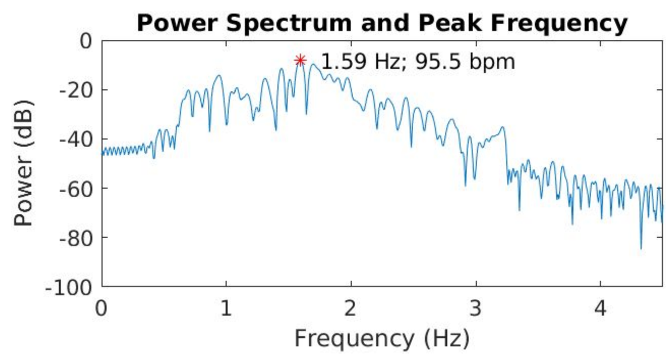
Figure 17: An example set from UBFC-RPPG dataset [3]

In UBFC-RPPG [3] database, there are 42 records created with a simple low-cost webcam at 30fps with a resolution of 640x480 in uncompressed 8-bit RGB format. A transmissive pulse oximeter was used to obtain the ground truth PPG data. The subjects sit in front of the camera at a distance of about 1m with their faces visible. The environment was well-lit. The subjects are required to play a time-sensitive mathematical game. This increases their heart rate. All experiments are conducted indoors with a varying amount of sunlight and indoor illumination. We can see some examples in Figure 17.

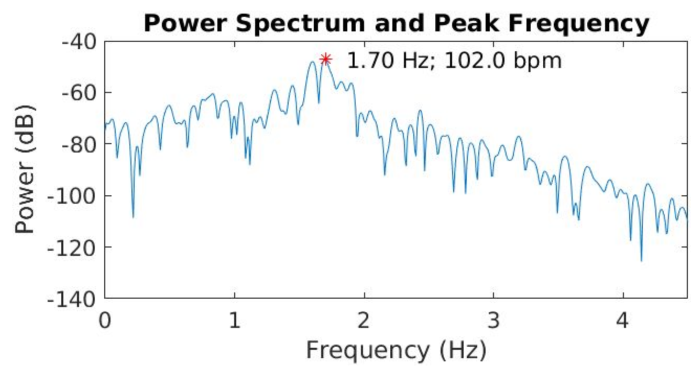
4. System Architecture

First, we will read the videos and frame them. We have to detect a face in each frame. The first thing we need to do predicting heart rate from the video should be to find the face from the video and crop it. Because foreign objects in the background can cause the algorithm to work incorrectly. At this stage, we use the Viola-Jones face detection technique [6] to automatically detect the face of the subject. This step provides bounding box coordinates defining the subject face. After the face detection, we need to make skin detection and remove non-skin pixels. The skin detection is performed on every frame to filter out non-skin pixels. The area of interest, the skin part, is our ROI piece.

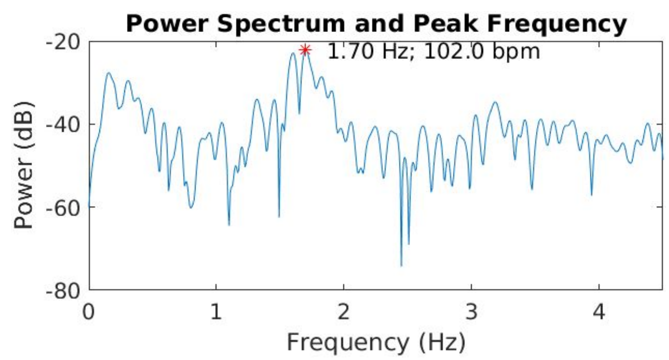
After the ROI [11] region is taken, we now have data to search for the information we want. We constitute temporal RGB signals by making colour transformation on this data.



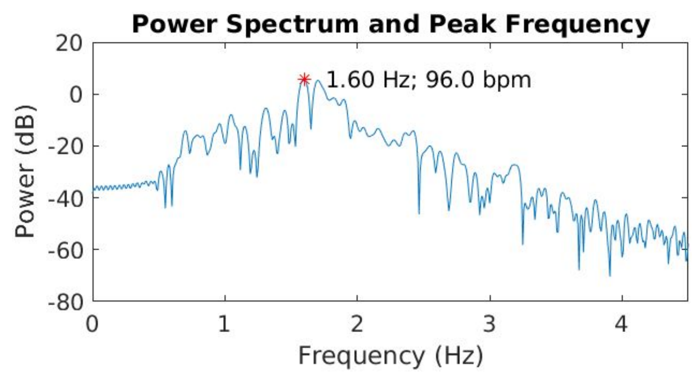
GREEN



CHROM



POS



ICA

Table 1: Results of four methods at 13.avi from UBFC-RPPG dataset

The raw RGB signals are composed by calculating the average pixel value of the skin-pixels within the ROI region over time. The RGB signals contain information about the heart rate in mixed components. Therefore, we will use deep learning approaches to recover the source signals from these mixed signals. As we can see a few results in Table 1.

And then, the spectrum of the resulting components of these methods is obtained. The peaks in the components power of these methods are determined, and the index frequency of the highest peak corresponds to the heart rate frequency.

5. Experimental Study

Experimental Setup:

- Experiments must be done indoors.
- The experiment environment must be well-lit.
- People must sit at a table in front of a laptop at a specified distance from the camera.
- The face of the subject should be visible, so the subject should not turn his/her head.

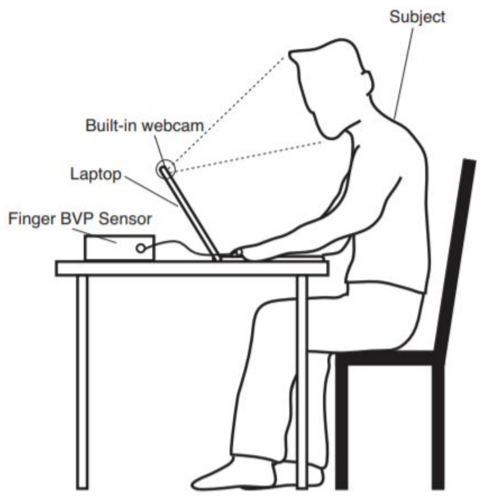


Figure 18: Experimental Setup [9]

Experimental Results:

In the below table, there are average RMSE values of videos we have obtained from UBFC [3] dataset and iPhys [19] library methods. When we look at the average RMSE values in Table 2, the best working method is CHROM DEHAN [8] because it has the smallest error rate. Also, the worst working method is GREEN VERCRUYSSE [7] because it has the highest RMSE value.

	CHROM DEHAN (RMSE)	ICA POH (RMSE)	GREEN VERCRUYSSE (RMSE)	POS WANG (RMSE)
Average RMSE	**14.5**	**16.0**	**27.2**	**20.6**

Table 2: iPhys library methods average RMSE values

But when we look at some video RMSE values in the below table, we cannot say that the CHROM gives the best result for every video. For example in video 17, CHROM is the worst and GREEN best method. Besides when we look at video 42 and 43. POS WANG method's RMSE values are far from each other.

Video	CHROM DEHAN (RMSE)	ICA POH (RMSE)	GREEN VERCRUYSSE (RMSE)	POS WANG (RMSE)
13.avi	10.4263	10.6027	12.3580	10.6355
14.avi	16.2840	22.6972	27.8859	24.4553
17.avi	18.4104	16.5873	12.8400	17.7870
31.avi	8.7691	8.6861	21.5627	1.1123
37.avi	6.0574	6.0336	27.3165	6.0389
40.avi	7.7885	7.8210	16.7349	7.8318
42.avi	17.4668	16.6722	24.4086	45.6642
43.avi	2.8240	2.8962	41.8808	3.2266

Table 3: Some videos RMSE values from iPhys library method

Discussions:

We can look at RMSE values for Meta-rPPG [20] technique in Table 4.

Method	RMSE
GREEN	20.6
ICA	18.8
CHROM	20.3
POS	10.5

Table 4: Results of average HR measurement on UBFC-rPPG from Meta-rPPG paper [20]

We compared iPhys library methods RMSE results with the Meta-rPPG technique RMSE results. As we can see in Table 5, iPhys library methods RMSE values are better than Meta-rPPG for ICA and CHROM methods. But for GREEN and POS methods, Meta-rPPG RMSE values are better than iPhys library methods. In general, we can say that we made improvements for ICA and CHROM methods.

Methods	RMSE values of iPhys library methods	RMSE values of Meta-rPPG
GREEN	27.2	20.6
ICA	16.0	18.8
POS	20.6	10.5
CHROM	14.5	20.3

Table 5: Comparing RMSE values for iPhys library and meta-rPPG methods

6. Tasks Accomplished

6.1 Current State of the Project

We dealt with traditional methods of reading videos, processing and generating signals and several calculations. We processed RGB signals and calculated heart rate using the ICA, GREEN, POS and CHROM methods and then we calculated the RMSE values and compared the success rates of our results with the equivalent results in the literature. We have remained loyal to the plan we mentioned in Project Specification Document (PSD) and we have completed four phases so far.

6.2 Task Log

Analyzing the current methods and different approaches

September 2020, on the online platform

The group members worked for a few weeks.

The results were shared in the meeting with the advisor.

Provision of equipment and establishment of development environments

November 2020, on the online platform

The group members worked for 1 week.

Studying and comparing the UBFC and PURE datasets.

December 2020, on the online platform

The group members worked for a few weeks and the results were reported.

The results obtained were evaluated in the meeting with the advisor.

Implementing the traditional methods and testing with datasets.

January 2020, on the online platform

The group members worked for a few weeks and the results were reported.

The results obtained were evaluated in the meeting with the advisor.

6.3 Task Plan with Milestones

#	**Task Description**	**Expected Output**	**2020**				**2021**				
			Sep	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**
1	Deciding the project subject.	Researching project subjects, checking existing solutions, discussing improvement ideas	✓								
2	Understanding and identifying the problem & The literature survey of rPPG	Focusing on the fundamentals of the problem, determining the solution's impact area, examining existing studies and solutions	✓								
3	Analyzing the current methods and different approaches	Understanding the solution methods of current approaches and understanding the problems they cannot solve	✓								
4	Planning general methodology	Determining the preliminary implementation to be done with traditional methods for the first term, determining the iPhys framework-based methodology		✓							
5	Preparing project specification document (PSD)	Preparation and delivery of the document			✓						
6	Finding the dataset	Finding datasets suitable for use in rPPG studies and permitted for use in academic studies, taking necessary permissions			✓						
7	Provision of equipment & Establishment of development environments	Selection of computers with suitable webcams and installation of environments required for developers			✓						
8	Studying on the UBFC Dataset	HR estimation and results storage with UBFC data using iPhys framework methods				✓					
9	Comparing methods	Comparison of results and discussion of factors affecting correct results					✓				
10	Evaluation of the results	Evaluation of results with RMSE values					✓				
11	Presentation	Presentation of the project about the current state and planned part of the project					✓				
12	Preparing analysis and design document (ADD)	Preparation and delivery of the document						✓			

13	Implementation of DeepPhys and PhysNet	Starting the implementation of the project using DeepPhys and PhysNet frameworks, estimating HR with UBFC and PURE datasets							❍		
14	Applying deep learning approaches	Understanding and applying deep learning approaches and discussing the expected improvement							❍		
15	Improvement of the errors	Focusing on the error rate which obtained with the test data set and improvement studies								❍	
16	Testing on real data	Testing the project with live subjects simultaneously using a pulse oximeter and recording results, observing the results and measuring the accuracy								❍	
17	Presentation & Preparation of poster and report	Completing the final work, preparing documents, presenting the final version of the project and its results									❍

Table 6: Task plan and status

7. References

[1] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote ppg,” IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1479–1491, 2016

[2] Stricker, R., Müller, S., Gross, H.-M. “Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot” in: Proc. 23st IEEE Int. Symposium on Robot and Human Interactive Communication (Ro-Man 2014), Edinburgh, Scotland, UK, pp. 1056 - 1062, IEEE 2014

[3] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, (2017), Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognition Letters

[4] W. Chen and D. McDuff, “Deep Phys: Video-based physiological measurement using convolutional attention networks,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 349–365

[5] Z. Yu, X. Li, and G. Zhao, “Remote photoplethysmography signal measurement from facial videos using spatio-temporal networks,” in Proc. BMVC, 2019, pp. 1–12.

[6] Paul Viola and Michael Jones (2001), Robust Real-time Object Detection, International Journal of Computer Vision

[7] Vercruysse, W., Svasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. Optics express, 16(26), 21434-21445.

[8] De Haan, G., & Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. IEEE Transactions on Biomedical Engineering, 60(10), 2878-2886

[9] Poh, M. Z., McDuff, D. J., & Picard, R. W. (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics express, 18(10), 10762-10774

[10] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote ppg,” IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1479–1491, 2016

[11] W. Wang, S. Stuijk, and G. De Haan, “Unsupervised subject detection via remote ppg,” IEEE Transactions on Biomedical Engineering, vol. 62, no. 11, pp. 2629–2637, 2015

[12] www.joom.com (access time: 06.02.2021 12:10)

[13] Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard

[14] New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method Rencheng Song, Senle Zhang, Juan Cheng, Chang Li, Xun Chen

[15] medium.com/@aaronward6210 (access time: 05.02.2021 22:26)

[16] Fouad, R. M., Osama A. Omer, Abdul-Magid M. Ali, and Moustafa H. Aly. “Refining ROI selection for real-time remote photoplethysmography using adaptive skin detection.” 2019 International Conference On Innovative Trends In Computer Engineering (ITCE’2019) 978-1-5386-5261-9, no. 19 (February): 6. 10.1109/ITCE.2019.8646412

[17] www.statisticshowto.com (access time: 23.01.2021 17:28)

[18] sphweb.bumc.bu.edu (access time: 28.01.2021 19:12)

[19] Daniel M., Ethan B., “iPhys: An Open Non-Contact Imaging-Based Physiological Measurement Toolbox” , Jan 2019, Corne University

[20] Lee, Eugene, Evan Chen, and Chen-Yi Lee. 2020. “Meta-rPPG: Remote Heart Rate Estimation Using a Transductive Meta-Learner.” (July), 26