T.C.
MARMARA UNIVERSITY
FACULTY of ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

CSE4197 – Engineering Project
Analysis and Design Document

Heart Rate Estimation from Facial Videos

Group Members

150115009 – Ferhat ÖZKAN

150114020 – Ilyas Saadeddin PORHAN

Supervised by

Prof. Çiğdem Eroğlu ERDEM

02.01.2019

Contents

1. Introduction 4
  1.1 Problem Description and Motivation 4
  1.2 Scope of the Project 8
  1.3 Definitions, Acronyms, and Abbreviations 8
  1.4 Aims of the Project 9
    Project Aim 1: 9
    Project Aim 2: 9
    Project Aim 3: 9
    Project Aim 4: 9
    Project Aim 5: 9
  1.5 Success Factors and Benefits 10
    1.5.1 Measurability / Measuring Success 10
    1.5.2 Benefits / Implications 10
2. Related Work 11
  2.1. Electrocardiographs 11
  2.2. Phonocardiograph 11
  2.3. Photoplethysmography 11
3. System Design 17
  3.1 System Model 17
  3.2 Flowchart of Proposed Algorithms 17
    3.2.1 Face Tracking 18
    3.2.2 RGB Extraction (Signal Extraction) 20
    3.2.3 Pre-Processing 20
    3.2.4 ICA 22
    3.2.5 Component Selection & Power Spectrum Analysis 24
    3.2.6 HR Prediction 26
  3.3 Comparison Metrics 27
  3.4 Data Sets or Benchmarks 27
  3.5 Professional Considerations 28
    3.5.1 Methodological Considerations / Engineering Standards: 28
    3.5.2 Societal / Ethical Considerations: 29
    3.5.3 Legal Considerations: 29
  3.6 Risk Management 29
4. System Architecture 30

Step 1: 30
Step 2: 30
Step 3: 30
Step 4: 30
Step 5: 30
Step 6: 31
5. Experimental Study 32
6. Tasks Accomplished 32
6.1 Current state of the project 32
6.2 Task Log 33
Meeting 1 - 04.06.2018 33
Meeting 2 - 19.09.2018 33
Meeting 3 - 21.09.2018 33
Meeting 4 - 02.10.2018 33
Meeting 5 - 10.10.2018 33
Meeting 6 - 17.10.2018 34
Meeting 7 - 24.10.2018 34
Meeting 8 - 31.10.2018 34
Meeting 9 - 21.11.2018 34
Meeting 10 - 28.11.2018 35
Meeting 11 - 05.12.2018 35
Meeting 12 - 12.12.2018 35
Meeting 13 - 19.12.2018 35
Meeting 14 - 26.12.2018 35
6.3 Task Plan with Milestones 35
6.3.1 Project Phases (Milestones) 35
6.3.2 Gant Chart & Work Sharing Chart 37
References 38

1. Introduction

Cardiovascular diseases are the number one cause of death globally. The key to prevention and treatment is the early detection of angiocardiopathy (diseases related to heart and blood vessels). One significant vital parameter for human cardiovascular status (CVS) assessment is Heart rate (HR).

A method to extract physiological parameters such as heart pulse rate, respiratory rate, and their variation with respect to time is photoplethysmography (PPG). The idea behind this method is to measure the blood volume change in the tissue with optical measurements. There are two Photoplethysmography approaches contact and non-contact photoplethysmography. The problem with contact photoplethysmography method is that patients can feel uncomfortable with the measurement devices and this can affect the measurements. In this project we aim to create a non-contact photoplethysmography method, which is able to measure heart rate from facial videos.

1.1 Problem Description and Motivation

Cardiovascular diseases (CVDs) are the world's biggest killers. The top causes of death related to cardiovascular diseases are ischeamic heart disease and stroke. Of the 56.9 million deaths recorded in 2016 cardiovascular diseases killed 15.2 million people as seen in Figure 1 [1]. Smoking, lack of activity, drinking alcohol are some reasons triggering these diseases. These later show up in people as raised blood pressure, high blood sugar, overweight and obesity, risks harming the good heart health.

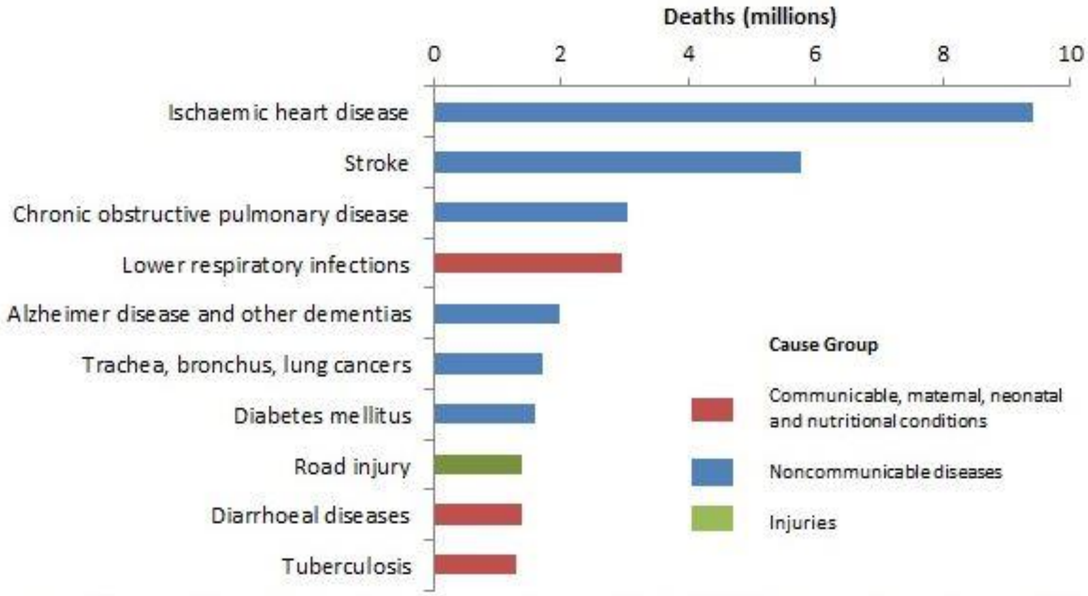

Figure 1 Top 10 Global Causes of Deaths 2016 (source: who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death)

Tachycardia, bradycardia, bradypnoea, tachpnoea, apnoea, hypoxemia are some common cardiac anomalies. The early detection of anomalies in cardiac activities are not that easy. As an example we can examine tac

hycardia. Tachycardia sometimes has no symptoms but if we don't treat it can cause serious heart function problems. It even can lead to death [2]. Some symptoms for tachycardia are rapid pulse rate, shortness of breath and chest pain. Since these symptoms are very general for cardiac anomalies it is hard to get diagnosed correctly. There are several reasons that can cause rapid heart rate and tachycardia symptoms [2]. That's why monitoring as much data as possible is important for diagnosing a patient correctly. Not only the size of data is important. As an example a rapid pulse rate recorded when someone is doing exercises is not considered as an indicator for tachycardia. The data should be collected when the human body is in a stable state. For example, the data collected when the patient is sitting on a chair reading an article can be considered as stable. Another important point in data is that it should be continuous. A healthy person would not contact a doctor before having the symptoms of a cardiac anomaly. But as we mentioned before sometimes these anomalies do not show any clear symptoms. Therefore, the cardiac anomaly could turn into a serious problem if it is not detected in early stages. A solution to this is collecting data continuously and analyzing it. These analyses can be useful in early detection and a reason to see a doctor.

Mentioned before heart rate monitoring can be useful in early detection but there are some additional signs that can be an indicator. Monitoring of human physiological signs, such as heart rate variability, respiration rate, and blood oxygen saturation plays also a role in diagnosis of health conditions and abnormal events.

Traditionally, monitoring of cardiorespiratory activity is done with adhesive sensors, electrodes, leads, wires and chest straps which in a long use case scenario may cause discomfort and restrict the activity of the patient (Figure 2). Some patients with sensitive skin can feel discomfort and the use of these sensors may cause skin damage, infection or allergic reactions on these patients. Since the equipment used for monitoring like monitoring leads and electrodes are used a single time for hygienic reasons there is a cost every time a monitoring is done. Also these monitoring methods are not suitable for every patient. A patient with physical disability will struggle with cables and all the coupling. Also the same with babies.

Monitoring with traditional equipment can cause the patient to feel uncomfortable and this can affect the measurements. As an example a baby having all the equipment attached can feel fear and the pulse rate could increase.

Listed all these problems with traditional methods people started to seek for other solutions and they came out with different alternative measurement methods for monitoring of physiological signs.

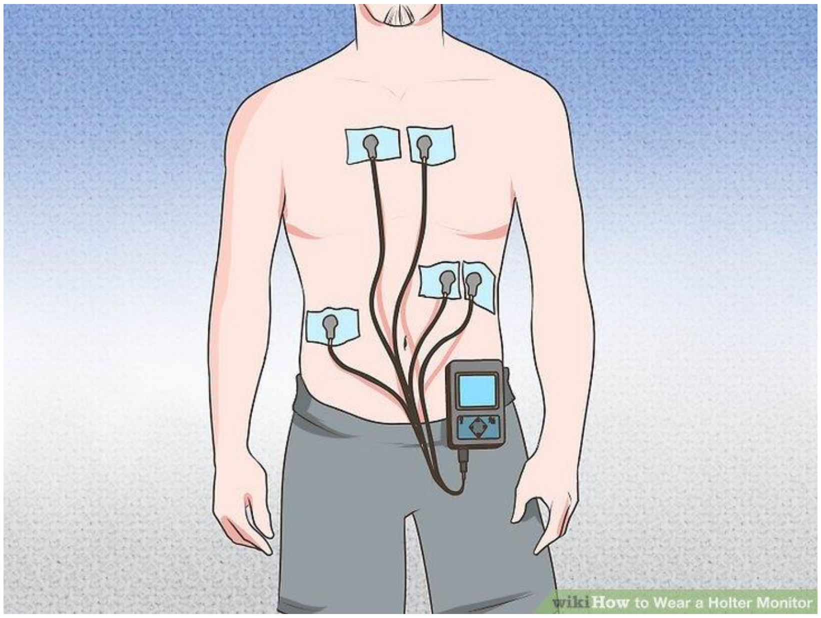

Figure 2 Holter Monitor For Heart (source: wikihow.com/Wear-a-Holter-Monitor)

Over the last 10 years the ability to measure the cardiac pulse and other vital signs via non-contact technologies has gained popularity. As an economic solution of traditional monitoring devices webcams now serve as low-cost and remote sensors. Nowadays almost every laptop comes with a webcam which also increases the use case of these kind of sensors. They are capable of recording biological signals. Pulse rate, SpO2 oxygen saturation are monitored using these signals. These indicators are continuously observed and analyzed for early detection in several health problems in clinics. Also people can use them for private purposes outside of the hospital in real life scenarios (while sitting in front of the laptop etc.). Also non-contact technologies overcome the above mentioned problems like skin damage, infection or adverse reactions. Since the whole process of monitoring is done remotely and there is nothing irritating or discomforting attached to the patient's bodies. [3]

Our motivation in this project is to develop a non-contact photo plethysmography method which is able to provide continuous information about HR measurements of users using a webcam. The system will overcome the problems with traditional measurement tools. It will deliver data which can be used in the early detection of cardiac anomalies.

1.2 Scope of the Project

Our Scope begins with the suitable seating conditions. In our project, the subjects should sit upright and still in front of the camera. In fact, we aim to make improvements in situations where subjects are moving, but our first goal will be to focus on subjects that are not moving. In addition, in the database we use, the lighting conditions in the environment will be suitable and sufficient. If we have time, we can make improvements in different light conditions.

There won't be a real-time system running. We will work on the data on a database that already exists. If we can get results from the data in the database successfully, we will test our system on a few alternative subjects.

The database we will use which contain only under 30 years of subjects which has no heart disease. The subjects have no any disability. Also, the subjects ' body temperatures will be normal for people.In additionally, the subjects will have no beard, moustache or makeup.

In this project, we aim to estimate the heart rate of person using a properly shot video. We aim to estimate heart rate with the margin of error at most +-5 beats per minute in motionless subjects. In moving experiments, we aim to see that the margin of error can be accepted up to 10 beats.

1.3 Definitions, Acronyms, and Abbreviations

HR: Heart Rate

BPM: Beats per minute

ROI: Region of Interest

RGB: Red Green Blue

ICA: Independent Component Analysis

SD: Standard deviation

PPG: Photoplethysmography

ECG: Electrocardiogram

BSS: Blind Source Seperation

HCI: Human Computer Interaction

1.4 Aims of the Project

Our aims are listed below:

Project Aim 1:

Today, the contact method used to measure heart rate (ECG) may cause serious damage to the sensitive human skin (e.g. newborns). One of the objectives of this project is to prevent the damage by doing the measurement in a non-contact way.

Project Aim 2:

People can be in stress, pressure and even excitement when they take the ECG test. This type of factors can cause the test results to deteriorate. With this project, we aim to eliminate these defects in the test results.

Project Aim 3:

Other remote camera-based PPG methods are sensitive to the motion of the subject. In this project we aim to design an algorithm, which is robust to the motion of the subject.

Project Aim 4:

We aim to estimate heart rate with the margin of error at most +-5 beats per minute in motionless subjects. In moving experiments, we aim to see that the margin of error can be accepted up to 10 beats.

Project Aim 5:

We will detect %90 of the facial key points in the ROI.

1.5 Success Factors and Benefits

1.5.1 Measurability / Measuring Success

Developing a low-cost non-contact photoplethysmography method which is able to monitor heart rate using a webcam is the main aim of the project. If the below listed requirements are satisfied, the project will be considered as successful.

- If the system is able to determine the ROI (region of interest) which is the face for our case the project is successful.
- If the system is able to detect facial key points in the ROI, the project is successful. We will use well-tested methods such as dLib [4] or CFSS [5] for this step.
- If the system is able to track this ROI throughout the video recording the project is successful.
- If the system is able to analyze the RGB signal and estimate heart rate the project is successful. The accuracy we aim to get is within 5 bpm (beats per minute) of the ground truth heart rate.

1.5.2 Benefits / Implications

The potential benefits of the project are;

- The system has the potential to be used as a heart rate monitoring device in clinics.
- The system can be used for personal usage. Data collected while people are sitting in front of the laptop can be analyzed and used for early detection of cardiac anomalies.
- The system can be used for commercial purposes. One example could be game developers. They could analyze videos of players playing their games and from the heart rate information they can get an idea where players feel happy or excited.

2. Related Work

Various methods are used to measure heart rate. Each of these methods focus on different human body activities and use different techniques to measure heart beat or cardiac cycle and determine heart rate. Beside non-contact photoplethysmography techniques we will give some traditional monitoring method examples. But the main focus will be photoplethysmography techniques.

2.1. Electrocardiographs

Muscles contract and relax this natural behavior causes the blood to travel from hearth into our vessels. During this operation of pumping blood from heart to vessels our body generates some electrical impulses. By the usage of electrical devices attached on the body, it is possible to capture these impulses and do measurements. The device that is capable of this is called ECG (Electrocardiographs). [6]

2.2. Phonocardiograph

The sounds that the hearth valves are producing during the pumping operation can be recorded using a stethoscope. These sounds are related to the HR and can also be recorded using a microphone. Sounds not related to HR like murmurs cause to noisy data but they can be filtered. The method of monitoring HR from sounds is done with phonocardiographs. [6]

2.3. Photoplethysmography

Every time the hearth pumps blood to our vessels the vessels fill with hemoglobin cells. Cells saturated with oxygen in other words when the vessel contains a lot of hemoglobin cells the skin absorbs more light. Using this some monitoring devices like the fingertip oximeter is monitoring our HR. All techniques that use this light differences in vessels are wrapped in the title Photoplethysmography.

In our project we will use a non-contact photoplethysmography method. We will extract the heart rate information of patients using a webcam. Therefore, we want to mention about some papers related to our main topic.

In the work of Poh et. al. [7], authors introduced a cheap way of monitoring several health signals with an ordinary webcam. For the experiment they have chosen 12 people with different genders. Participants with a wide range of age and skin color. The recordings were done inside a room with different illumination conditions. People in the recording are in front of a laptop and sitting on a chair 0.5m away from the built-in webcam. They have used MATLAB to build software. They have utilized the OpenCV library for face detection. After obtaining the coordinates and the box around the face they trimmed the box with full height and %60 width. After selecting the ROI, they separated the RGB values for each frame calculated an average value compared to all pixels in the frame and for each set of frame they construct the raw RGB signals. They based one the joint approximate diagonalization of eigenmatices algorithm they decomposed the normalized raw traces using ICA. After this operation they have obtained three independent source signals. ICA returns the components generated by the algorithm in random order. Thus they have selected the component whose power spectrum contained the highest peak for further analysis. They smoothed the separated source signal using a five-point moving average filter and bandpass filtered it. They have shown that it is possible to monitor several signals using just a webcam. Overall results are shown in Figure 3. In the recordings people had to breath suddenly to generate sudden increasing HR values, and to look at the camera straight and with no rigid motion. Extending their work to deal with more realistic use cases is our goal in this project.

Statistic	Heart Rate (bpm)	Respiratory Rate (breaths/min)	Heart Rate Variability		
			LF (n.u.)	HF (n.u.)	LF/HF
Mean error	0.95	0.12	7.53	7.53	0.57
SD of error	0.83	1.33	10.17	10.17	0.98
RMSE	1.24	1.28	12.3	12.3	1.1
Correlation coefficient	1.00	0.94	0.92	0.92	0.88

Figure 3 Summary of Overall Results [7]

In the work of Li et. al. [8] authors introduced a method by using recordings where subject's motion and illumination variations are involved. In the real world application people will not always look at the webcam straight the illumination of the background will change and people will have rigid motion like facial expressions and head movements. The method they have implemented have focused on these challenges. They have selected a multi-modal dataset MAHNOB-HCI because it contains recordings that can simulate real world situations as mentioned above, the dataset is big enough for their research and is publically accessible. As a first step they selected the ROI using Viola-Jones face detector. They have detected facial landmarks using Response Map Fitting (DRMF) method. Using this facial landmarks, they have generated a mask of ROI covering the chin and cheeks. To track this ROI, they have used the Kanade-Lucas-Tomasi (KLT) algorithm. Using the Distance Regularized Level Set Evolution (DRLSE) method they have reduced the interferences caused by illumination variations. They have calculated the best fitting coefficient using the Normalized Least Mean Squares (NLMS) filter. To reduce the effect of non-rigid motion noisy data computed they divided the signal into segments and discarded segments related to noise. If these noisy segments wouldn't be discarded they would end up as sudden high peaks in the signal after filtering. They limit the margin of estimated HR using several temporal filters to only calculate meaningful HR values. At the end they have employed Welch's power spectral density estimation technique to turn the HR time signal into frequency domain. In Figure 4 the black curve is the ground truth HR measured by Polar system; the green curve is HR measured from video by using their custom software. They have implemented a method for real world use cases including the challenges of motion and varying illumination conditions. In our project we have the same goal.

Figure 4 HR monitoring of one subject while playing a video. The black points denote the ground true heart rates and the green points denote the estimated values. (source: “Remote heart rate measurement from face videos under realistic situations” [8])

In the work of Demirezen and Erdem [9] authors showed that it is quite successful to use nonlinear mode decomposition (NMD) in cases where the subjects are in motion. They have done experiments using the PureDL dataset and have seen that NMD based HR estimation performs better then Independent Component Analysis (ICA). In the work they consider motion as noise to the skin color information and mention that the NMD method is extremely noise-robust. They have analyzed related work and point out the areas where improvement can be done. Then they describe how they have adopted the NMD method and in the end they analyzed the results and compared them with the results of popular signal decomposition methods in the literature. Except signal decomposition the process of estimating heart rate is quite similar with traditional methods using ICA. First they have detected the face using the Viola-Jones face detector. After that they have selected the ROI and from the signals they have estimated the heart rate. By using the NMD method they have shown that it is more effective especially in cases containing motion.

In the work of McDuff et. al [9] authors are pointing out the strengths of non-contact physiological measurements using digital cameras and image processing. They say that among all studies photoplethysmographic (PPG) imaging is the one which is the one which has the highest interest. In their work they mainly focus on researches done on PPG. They explain how different physiological parameters can be extracted via different methods. They have mentioned about challenges which may occur when these methods are used for parameter extraction. Some of the difficulties mentioned are motion, illumination, image quality and ROI selection optimization. Several studies about these topics are analyzed and reported. Finally, they have listed several approaches and analyzed them in the way how they can be used in a real world application scenario. Some of them are baby-friendly non-contact measurement, a smart phone application for recording vital signs, SpO2 saturation and monitoring applications. At the end of the survey they say that a dataset which is public with varying amount of subject and camera motion, varying illumination and diverse subject pool would be great for future research and improvement.

In the work of Poh et. al [10] authors showed that Cardiovascular functions are very important for human life. Therefore, the diagnosis of these functions and early diagnosis of chronic diseases are very important. Heart rate measurement is the first step of this diagnostic. Today, the clearest method of this test is ECG. However, ECG may cause some irritations in the body due to the conditions it requires and is also a laborious method. To resolve these issues, the use of photo plethysmography (PPG) has been investigated. In fact, PPG has always been applied as light (signal), but with the latest research show that digital cameras can use for the measurement. In other words, it is thought that the human heart rate can be estimated through videos that recorded by digital cameras. However, the probability of deterioration of the signal due to movement in the PPG has already been known. In addition, there is the effect of the signal at similar frequencies due to noise. Blind source separation (BSS) is a technique developed to prevent this. BSS has many methods here ICA (Independent Component Analysis) will be emphasized. ICA will take part in the elimination of mixed independent signals. ICA can also use to decrease the PPG’s movement related errors.

Experiments were performed under different subject conditions and ambient conditions. Different physical properties of the subjects in the experiments are important for the result of the experiment. In the experiments, the subjects were asked to stand perpendicular to the screen. Two videos were recorded for each subject during the experiments. The first video was taken while the subject was standing in the appropriate position. Secondly, the subject acted normally and made small movements. The experiment also included videos with three subjects sitting together. In this experiment, inferences made on the density changes in the blood vessels of the face. RGB sensors used while doing this. The RGB values on each frame are processed on the ICA. These signals are introduced into the matrix operations according to the ICA model. Matlab is used to analyze the results.

For each frame, the face is automatically identified and the Region of Interest is determined. While the operations doing the frame is ignored which does not contain clear face. From this study, 60% width from the center was accepted as the full length region of interest. Then the ROI is divided into three RGB channels. Finally, Fourier transform was applied to the signals to obtain the power spectrum.

As a result of the experiments, the visible effect of ICA was positively reflected in the results. In addition, the second videos taken from the subjects, that allow the subjects small movements were also evaluated. Some results have been obtained through power spectra. The second component's values were considered to be much closer to reality.

3. System Design

3.1 System Model

A System Model Diagram is shown in Figure 5.

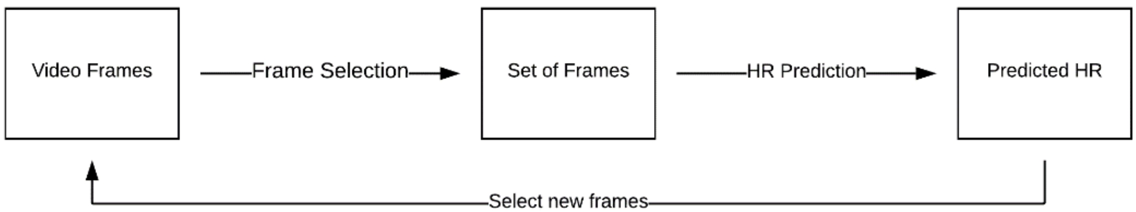

Figure 5 System Model Diagram

The system will take video recordings from PureDL and UBFC datasets. The recordings are stored as frames along with the heart rate information recorded with a pulse monitoring device. Also time-stamps are recorded together with the pulse information for each frame (more information about these datasets will be given in the “Data Sets” section in this document). Using these information’s, the videos will be analyzed offline in our custom software. We aim to estimate heart rate with a margin of error at most ± 5 beats per minute in motionless subjects. The system will continue selecting new sets of frames and predict HR till there is no frame left for corresponding video.

3.2 Flowchart of Proposed Algorithms

As mentioned before our software will take video recordings from two different datasets and predict heart rate. In this part we will give more information about the HR prediction process. A flowchart of this process is given in Figure 6.

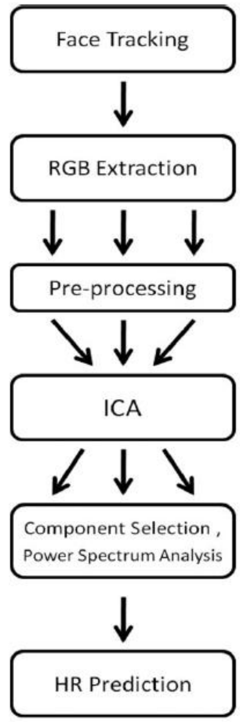

Figure 6 Flowchart of HR extraction from video recording [11]

The process starts with tracking the face in each frame. We have chosen different approaches and algorithms in the implementation so we can analyze the outcome of each and pick the one with best performance between all results.

3.2.1 Face Tracking

The HR prediction process starts with tracking the face. We will use OpenCV and dlib libraries to detect the face in each frame. OpenCV returns a rectangle covering the face on the other side dlib extracts facial landmarks. Facial landmarks can be used to determine the ROI more precise. An example of OpenCV and dlib outcomes is shown in Figure 7 and Figure 8 .

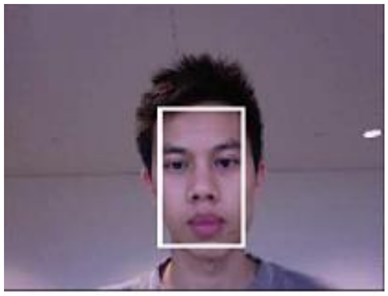

Figure 7 OpenCV ROI selection [7]

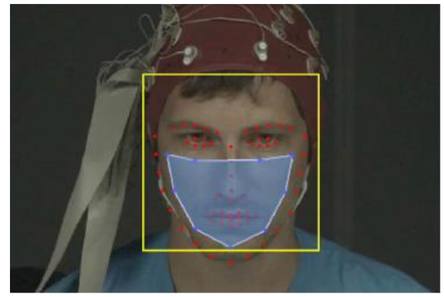

Figure 8 dlib ROI selection [8]

The region detected which contains the face is called ROI. In the ROI we are interested in frames containing skin. Eyebrows, hair and facial hair do not contain useful information about the HR. For this we will use a skin color detector. We decided to use a method called “Explicitly Defined Skin Color Detector “. In this method we look at the RGB values in each pixel and put them in an equation to decide whether it is skin or not. The equation is shown in Figure 9. The RGB values of a pixel are taken into the equation below and it is decided whether they are related to skin or not. The skin shouldn’t be grey this is obtained by the second line in the equation. R and G values also should not be close to each other this is obtained by the third line of the equation. [12]

$$\begin{aligned} R > 95 \;\&\; G > 40 \;\&\; B > 20 \;\& \\ max\{R,G,B\} - min\{R,G,B\} > 15 \;\& \\ |R-G| > 15 \;\&\; R > G \;\&\; R > B. \end{aligned}$$

Figure 9 Equation for skin color detection method [12]

3.2.2 RGB Extraction (Signal Extraction)

In the next step we continue with extracting RGB signals from our set of frames. In we can see how this step is done in detail.

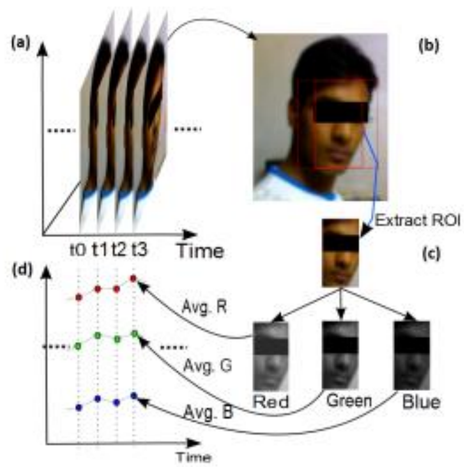

Figure 10 Above figure illustrates the process of extraction of raw RGB trace from a video sequence. [13]

In sub-image (a) set of frames of the first portion of the video are represented. In sub-image (b) The frames are represented with the corresponding ROI at each time index. In sub-image (c) Each RGB values in the ROI are separated and the average is taken and placed in the signal graph for each time index. (d) which shows the raw RGB traces. [13]

3.2.3 Pre-Processing

Before we continue with our analysis we have to do some pre-processing on the raw RGB traces. Firstly, we have to do normalize the RGB traces. Equation shown in Figure 11 is used for normalization.

$$x_i'(t) = \frac{x_i(t) - \mu_i}{\sigma_i}$$

Figure 11 Normalization Equation [13]

for each i = 1,2,3 where μ_i and σ_i are the mean and standard deviation of x_i (t) respectively. The normalization transforms x_i (t) to $x_i{}'$ (t) which has zero-mean and unit variance. [13]

Also to overcome the noise caused by non-rigid motion in the ROI we do some operations on the raw traces. This non-rigid motions can be caused by facial expressions and they lead to sharp peak values in the signal. If this noisy data recorded are not discarded, they will be represented as high peaks in the final signal after computing power spectrum density (PSD) leading to false estimations of HR. [8]

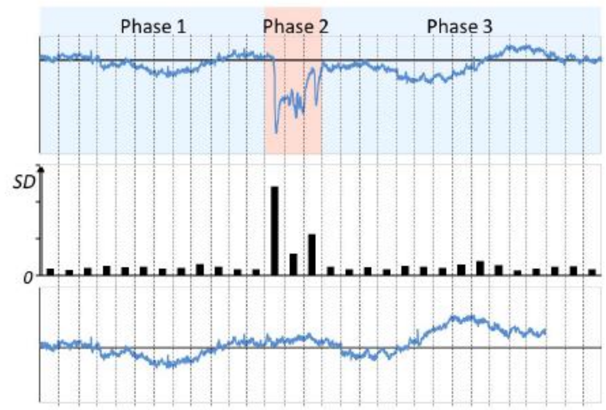

Figure 12 Discard operation of noisy signals occurring from non-rigid motions. The signal on the very top dedicates in Phase 2 a non-rigid motion. The graph in the middle shows the SD of each portion of signal. The portions with highest SD are eliminated and the resulting signal is shown below the SD graph. [8]

We are estimating HR for a set of frames, noisy segments like Phase2 shown in Figure 12 can be discarded for more accurate results. We divide the signal into m segments each with length $\frac{n}{m}$ where n is the total length of the signal. The standard derivation (SD) is calculated for each part, and %5 of all segments with the highest SD values are eliminated. [8]

Additional filters are applied after the signal is carried from time domain to frequency domain. These filters filter the computed frequencies such that the remaining frequencies are only the ones that are in a meaningful range. Our frequency range that we are interested in is [0.7, 2.5] Hz which equals to 42 bpm to 150 bpm. For this operation we will use three filters. The first one is a detrending filter based on a smoothness priors approach, which is used for reducing slow and non-stationary trend of the signal. The second one is a moving-average filter, which removes random noise using temporal average of adjacent frames. The third one is a Hamming window based finite impulse response bandpass filter with cutoff frequency of [0.7, 2.5] Hz. [8]

3.2.4 ICA

After doing pre-processing operations on raw traces they will be decomposed into three independent source signals using ICA based on the Joint Approximate Diagonalization of Eigenmatrices algorithm (JADE).

The goal of ICA is to solve BSS problems which arise from a linear mixture. The explanation of an example BSS problem is given below.

Cocktail Party Problem

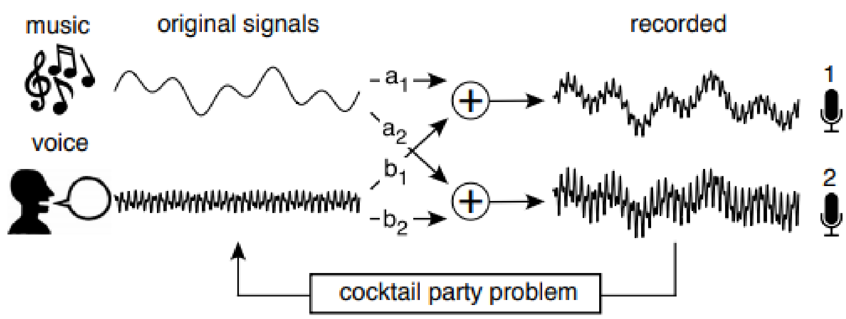

Figure 13 Example of the cocktail party problem.

In the Cocktail Party Problem, we have several sounds in the same environment and several microphones recording the sounds in the environment. Taking the example shown in Figure 13 in consideration we have sounds as music and voice denoted as s_1 and s_2. They are recorded simultaneously in two microphones denoted as 1 and 2. Sounds add up linearly with coefficients a_1 and b_1 for microphone 1 and coefficients a_2 and b_2 for microphone 2. Coefficients are related how close sounds are close to the microphones. The goal of the Cocktail Party Problem is to separate original sound signals from signals collected in the microphones.

Our problem in this project is similar to the Cocktail Party Problem. We have pre-processed RGB signals which contain information about the HR. We put the RGB signals into the ICA based Joint Approximate Diagonalization of Eigenmatrices algorithm (JADE). The algorithm takes the separated RGB signals and outputs three separated source signals. It is not clear which of the separated source signal is related to the HR. We know that they are related to HR, SpO2 (Oxygen saturation in the blood) and noise in data (caused by illumination and motion). But their order is random in the output of ICA. A visual explanation is shown in Figure 14.

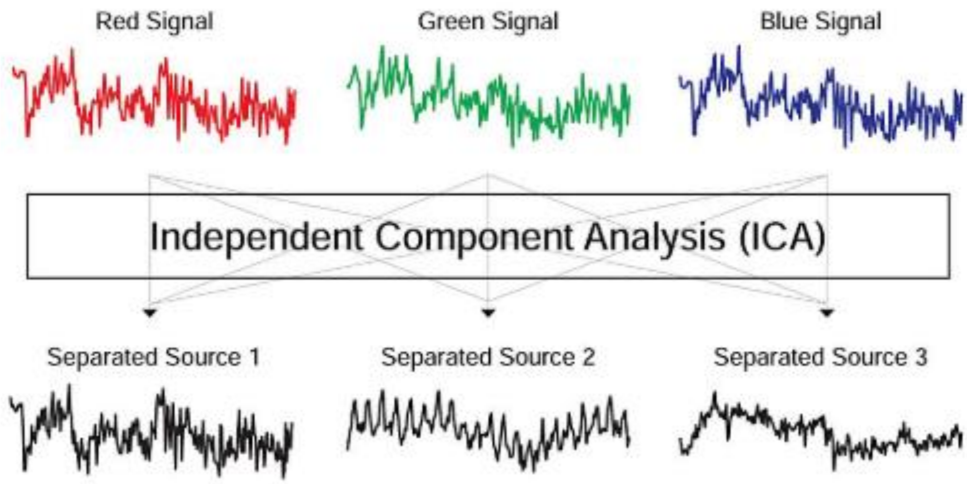

Figure 14 The separated three independent source signals obtained by applying ICA to the three detrended and normalized raw signals.[7]

<u>ICA Algorithm</u>

If we consider inputs and outputs of ICA algorithm as vectors the equation of ICA can be represented with the equation shown below.

$$x = As$$

Where x is the a vector whose element are a random mixture of $x_1,\ldots, x_n$, and similar to x, s is a random vector with elements represented by $s_1,\ldots, s_n$. And A is a coefficient matrix with elements a_{ij}. To clarify remember the Cocktail Party Problem. The microphones are the X values, the sound signals are the s values and A is the matrix that contains the coefficients that were multiplied with these source signals. We do have only a random vector called x and we are trying to calculate two unknowns A and s. ICA starts by assuming the components in s are statistically independent. It also assumes that the independent component must have Non-Gaussian distributions. For simplicity it assumes that the coefficient matrix is square [14]. When it is done estimating A, it can easily compute the inverse of A (W) and compute s using W and x by the formula:

$$s = Wx$$

In many models it would be more realistic to assume that there is some noise but we don't add any noise to our calculations since the noise-free model is difficult enough itself and is verified that it works sufficient for many applications.

There are two ambiguities of ICA:

1) <u>We cannot determine variances of the independent components</u>
   S and A are unknown therefore if we multiply A with a scalar multiplier the effect won't be recorded if we do the same operation on s using a scalar multiplier such that the result won't change. We can't even say something clear about the sign of s. We could multiply any source signal in s with -1 and the coefficient would change to its negative. Thus x would not be effected.
2) <u>We cannot determine the order of the independent components</u>
   We did mention about this problem briefly before when we explained the Cocktail Party Problem. The reason again is that A and s are unknown, we can change the indexes of rows in A or s and the resulting x matrix won't change. Therefore, we can call any of the independent components the first one.

3.2.5 Component Selection & Power Spectrum Analysis

We will be calculating the power spectral density of each signal extracted from ICA using the Welch method [15].

Power Spectrum Analysis using Welch Method

Assuming that successive sequences are offset by D points and that each sequence is L points long. Thus the overlap is L-D points and if K sequences cover the entire N data points then. Therefore, the formula of Welch method can be given as in Figure 15.

$$\hat{P}_W(e^{j\omega}) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w(n)x(n+iD)e^{-jn\omega} \right|^2$$

Figure 15 Welch Method Formula

Where W(e^{jw}) is the Fourier transform of the L-point data window w(n).

Component Selection

Each outcome of ICA is named as a component. After calculating the power spectrum for each component now we have to select one of these components and estimate the HR. For this step we will be using two different approaches the SNR component selection method and a learning based method where we use some features of the power spectrum signals and predicting which one will give the better result for HR prediction.

SNR Based Component Selection

In SNR Based Component selection, we do not select the highest peak value but we look at the overall signal and calculate a SNR value for each signal. After this we pick the signal with the highest SNR value. This method is proved to give more accurate results compared with the traditional way of picking the highest peak value. The formula for calculation of SNR value is given in Figure 16.

$$SNR(f) = \frac{\sum_{A}^{B} spec(i)}{\sum_{0.75}^{2.5} spec(i) - \sum_{A}^{B} spec(i)}$$

Where,

A= $f - 0.067$

B= $f + 0.067$

Figure 16 SNR value calculation formula

Machine Learning based Feature extraction and Component Selection

Another approach is to extract features from the Power Spectrums and build a model that learns according to these features. When given power spectrums it should pick the one that would give us the best prediction for HR according to previous training samples. The features we have determined are shown in Figure 17. Circles are the frequency peaks for each power spectrum and lines are peak frequencies. So in total we will have 6 features in our model.

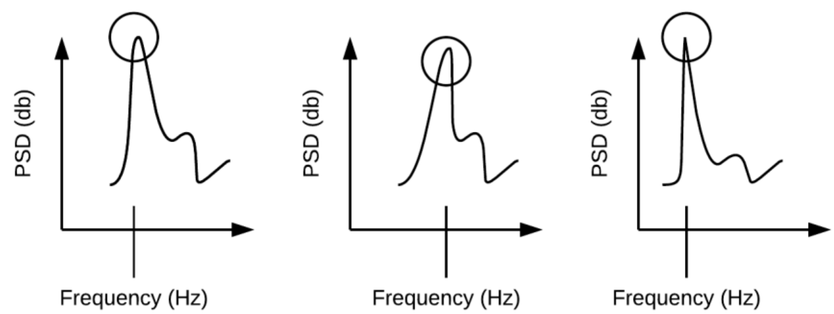

Figure 17 Features for extracted for a learning based component selection. 6 features are shown. Circles show the peak values of each power spectrum and lines show corresponding frequencies for each peak values.

3.2.6 HR Prediction

After selecting the right component in various ways described in previous step we pick the value that corresponds to the predicted HR and the algorithm stops here and returns to the first step and takes new set of frames till there are no frames left.

3.3 Comparison Metrics

There are some metrics for analyzing the results of our HR prediction software. We will use Descriptive Statistics for analyzing our data. We are interested in measurements the following measurements:

- Mean Bias(bpm)
- SD of Bias(bpm)
- RMSE (Root-mean-square deviation)
- Corr. Coefficient
- Mean absolute error
- Relative absolute error
- Root relative squared error

Additionally, we will use Bland–Altman analysis [14] to understand how well our predicted HR fits to the actual HR measured by the monitoring device.

3.4 Data Sets or Benchmarks

We will use the PureDL [16] and UBFC [17] dataset. The PureDL dataset contains face recordings of 8 different people under 6 different setups. From these 6 different setups we will mainly use the steady recordings in which the subjects look at the camera and avoid any motion. Illumination of the environment are different in the recordings. An example screenshot of one of the video recordings is shown in Figure 18.

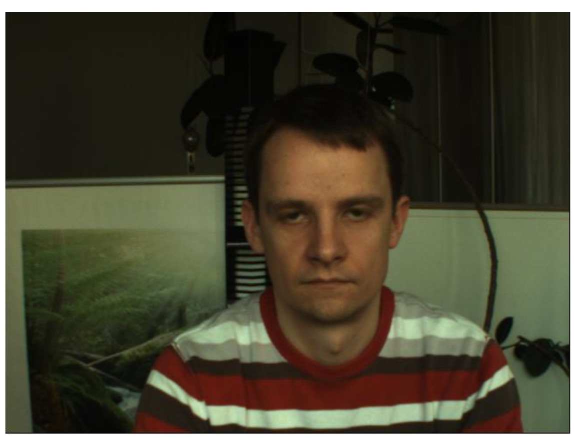

Figure 18 PureDL Subject1 [16]

Our second dataset which is UBFC dataset contains 42 videos of different subjects. Each video is recorded in different illumination conditions and subjects have varying skin color. Including subjects having scars in their face and wearing glasses. An example screenshot of one of the video recordings is shown in Figure 19.

Figure 19 UBFC Subject 46 [17]

3.5 Professional Considerations

3.5.1 Methodological Considerations / Engineering Standards:

- We will use GitKraken to manage version control [18].
- Gantt charts will be used to present our management plan.
- We decided to use MATLAB for software developing. MATLAB has many libraries especially the ones related to face detection will be useful for our project.
- As a software development process, we will use Scrum Project Management & Scrum Methodology.
- We will use Trello for management of the project [19].
- We will use the PureDL and UBFC dataset [16] , [17].

MATLAB: MATLAB® is a programming program using a matrix-based language. Since it is based on matrix operations it is very powerful in mathematical calculations.

Trello: Trello is a tool used for organization and also software developing management. You can open your own personalized Trello boards and create tasks attach files and specify due dates to each task.

3.5.2 Societal / Ethical Considerations:

3.5.2.1 Economical:

Our project aims to reduce the cost of cardiac activity monitoring. Using a webcam will decrease the cost of sensors. There will be no need for the traditional monitoring equipment.

3.5.2.2 Environmental:

Since the equipment used for monitoring like monitoring leads and electrodes are used a single time for hygienic reasons after every monitoring there will be a waste product. This leads to pollution. Since this is not the case for our system it will be a much ecosystem friendly method.

3.5.2.3 Health and Safety:

As mentioned before we aim to implement a method which is suitable for all patients. Even patients with physical disability or sensitive skin will be able to use the system. This can be considered as an improvement in health. The system may also enable early detection of heart diseases, by continuously monitoring users in their offices, home or gyms.

3.5.3 Legal Considerations:

- PureDL and UBFC are datasets which are public accessible for research purposes.
- We will use the campus labs for software developing. The computers in the lab have the licensed MATLAB software.
- The limited version of Trello allows only a single power-up per board but this is no problem since the only power-up we are going to use is GitHub integration power-up.
- We will use GitKraken licensed since its free for students.

3.6 Risk Management

One of the major risks can be changes in light conditions. We may not be able to develop a system suitable for different light conditions. Therefore, we are going to try our best to make them stable in all conditions.

Another risk could be movements of subjects. In case of too much movement of the subjects during the experiment, the results may be impaired. In addition, subjects may have an unknown disease and this may lead to incorrect results.

However, we will try to develop a stable system to reduce these risks. In the worst case, we aim to see that the margin of error can be accepted up to 10 beats.

4. System Architecture

The implementation progress of our project will be as described in each step below.

Step 1:

In first step we will be detecting the face with two different methods dlib and OpenCV. Then use this location (also known as ROI) in our next step. We plan to write the location information for each method into a text file to avoid running the face detection methods again and again.

Step 2:

In this step we average red, green and blue pixels (total number of pixels of corresponding color to every pixel in the detected area) and put them into a time domain signal where each time index corresponds to a frame processed from the video. As a result, we get three (RGB) time domain signals.

Step 3:

Then some filtering algorithm will be applied to these signals. Detailed information about filters are given in section 3.2.3 Pre-Processing.

Step 4:

In this step we use Independent Component Analysis. According to the outcomes (three components which are time domain signals) we will be predicting HR.

Step 5:

In this step we take each component and convert them to a frequency domain signal by calculating their Power Spectrum.

Step 6:

In this step we estimate HR from power spectrum signals. We will be using two different methods one is the traditional one which is directly predicting the HR from the power spectrum and the second one is Machine Learning based HR prediction where we will be using different features extracted from the power spectrum.

The main process of our project is shown in Figure 20.

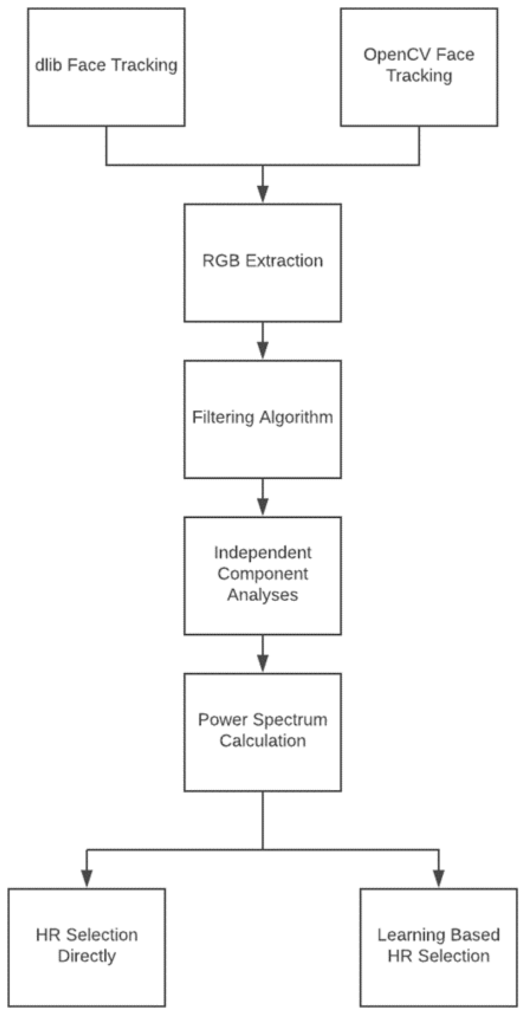

Figure 20 Detailed System Flowchart

5. Experimental Study

We plan to test our custom software on existing datasets. As we mentioned in system design part we will be using UBFC and PureDL datasets to calculate the HR of a subject. These datasets provide some different conditions which are necessary to say that the subjects were recorded in realistic situation. These conditions were also mentioned in the project constraints.

Illumination factors and rigid motion of subjects are our main challenges. We will be implementing a custom software which is robust to these changes.

We might also try to implement a demo software which takes real time data from the webcam and predict HR. But for this we have to be sure that the software is fast enough to process the data in real time and give response. Since we only used the algorithms for already existing data it is not easy to do any commend on this for now. It is a “could be good” not “must” for the project.

6. Tasks Accomplished

6.1 Current state of the project

So far we did a lot of research on Non-contact PPG and learned a lot about it. We started with very basic papers which were for starters and end up reading more advanced papers about different methods and approaches. We did a lot of research on the algorithms that are used In these methods and learned a lot about the mathematical background of them. A big part of our work done so far was doing research and trying to understand the fundamental approaches of our project. Additionally, we analyzed two datasets PureDL and UBFC. We tested two face detection libraries OpenCV and dLib. We could determine the face with both libraries. We tested these libraries in different operating systems and analyzed the performance. Also we did a Representation where we presented our project to a jury and some students. We are sticking to our time schedule that we gave in the PSD document.

6.2 Task Log

Meeting 1 - 04.06.2018

- Final Engineering Project topic selected.
- Some papers shared about non-contact PPG.
- Started to write a Project Proposal.
- Papers to be read before next year starts are determined.

Meeting 2 - 19.09.2018

- Started to write the Project Specification Document.
- Started to summarize and take notes about papers we have read.
- Started to split the project implementation into parts and assign them.
- Additional papers are given.

Meeting 3 - 21.09.2018

- Additional papers are given to be read.
- Started to do research about Machine Learning Based Methods.
- Some software like Trello and GitKraken are proposed for project management.

Meeting 4 - 02.10.2018

- Additional papers are given to be read.
- Started to prepare a PowerPoint representation about the papers we have read so far.

Meeting 5 - 10.10.2018

- A presentation about the papers we have read has been done.
- New papers are given to be read.
- CFSS will be tested for face detection.
- PureDL dataset given.
- PSD will be finalized and ready till next week for corrections.

Meeting 6 - 17.10.2018

- A paper about machine learning approach to improve contactless heart rate monitoring given to be read.
- Proposed software management tools are settled up to be used.
- PureDL dataset analyzed. Decided to not use the CFSS method.
- Most of the time busy writing the PSD.

Meeting 7 - 24.10.2018

- Another face detection library dlib has been proposed.
- This library will be tested.
- Additional papers are given to be read.

Meeting 8 - 31.10.2018

- Started to do research about extracting features from ICA components.
- An idea of predict the current heart rate from last k predictions came to mind. Limiting the candidate HR around this estimated value will be considered.
- Decided to use dlib for face detection. Each facial landmark found on the frame will be stored in a txt file.
- Ferhat will use OpenCV – Haarcascade method.
- Saadeddin will use dLib Based facial landmark detection.
- Rule based skin detector proposed to be used in the project. Link shared.

Meeting 9 - 21.11.2018

- Papers about ICA are given to be read.
- Links for UBFC dataset are given.
- Started to use OpenCV library. Tested the face detection with camera. Tested OpenCV on different operating systems for optimum performance. Tested OpenCV in Matlab and Python.
- Continuing doing research on feature extraction from ICA components.
- Dlib library tested and learned how to use it.

Meeting 10 - 28.11.2018

- Started to do research for ADD document.
- Decided to do more research for ICA.
- Decided to learn about power spectrum estimations.
- Start determining features for learning based method.
- Testing on dlib libraries continues.

Meeting 11 - 05.12.2018

- Additional papers for ICA are given to be read.
- Decided to do research on JadeR and FastICA.
- Dlib will be tested on UBFC dataset.

Meeting 12 - 12.12.2018

- Preparations for Project Presentation are done.
- Talked about how our presentation should be and what to mention in slides.

Meeting 13 - 19.12.2018

- A presentation to our supervisor was done.
- Changes for presentation slides were noted.
- Mainly focused on the presentation.

Meeting 14 - 26.12.2018

- Preparations for ADD document are done.
- Topics to be added to the document are chosen and listed.

6.3 Task Plan with Milestones

6.3.1 Project Phases (Milestones)

Each phase (aka milestone) of our project is listed below.

Phase 1:

Determining the problem: The problem is determined and different approaches for the solution of this problem are analyzed. Method of interest is selected and research on related work has started.

Phase 2:

Literature Survey: Reading projects/theses on this problem to understand different methods and their implementation.

Phase 3:

Preparation of Project Specification Document.

Phase 4:

Study and analyze the PureDL and UBFC dataset. Extracting data from the video recordings in MATLAB.

Phase 5:

Face detection in MATLAB using OpenCV and dlib.

Phase 6:

Determining the Region of Interest in the detected face.

Phase 7:

Decomposing of signals using ICA.

Phase 8:

Heart Rate estimation from extracted signals.

Phase 9:

Comparison of estimated heart rate and real heart rate values using visualization tools of MATLAB. (Plots, graphs etc.)

6.3.2 Gant Chart & Work Sharing Chart

In *Figure 21* we gave a Gant Chart for each Phase described in detail in section

Each phase is planned to be completed in the given time span.

In *Figure 22* we gave a Work Sharing Chart for each Phase determined. Blue colored boxes for Saadedin's impact and orange colored boxes for Ferhat's impact.

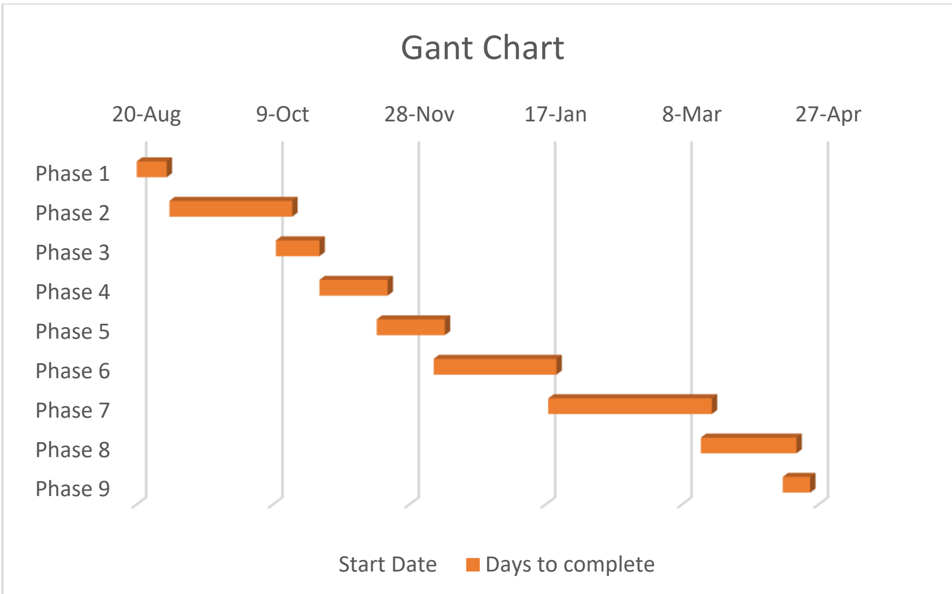

Figure 21 Gant Chart of the Project (Phases versus Due Date)

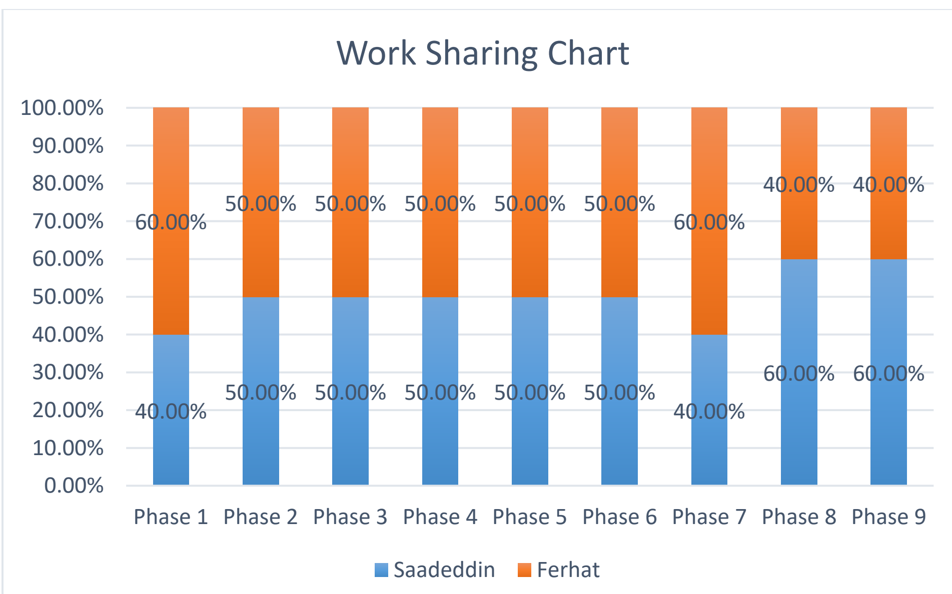

Figure 22 Work Sharing Chart Blue bars refer to the impact of Saadedin in the corresponding Phase and orange is the impact of Ferhat for the corresponding Phase.

References

[1] The Ministry of Health and Welfare (MOHW), “The top 10 causes of death, 2014,” *Animal Genetics*, 2015. [Online]. Available: http://www.mohw.gov.tw/cht/DOS/Statistic.aspx?f_list_no=312&fod_list_no=5488. [Accessed: 24-Oct-2018].

[2] “Tachycardia - Symptoms and causes.” [Online]. Available: https://www.mayoclinic.org/diseases-conditions/tachycardia/symptoms-causes/syc-20355127. [Accessed: 24-Oct-2018].

[3] F. Bousefsaf, C. Maaoui, and A. Pruski, “Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria,” *J. Med. Biol. Eng.*, vol. 37, no. 3, pp. 374–385, 2017.

[4] “dlib library Github.” [Online]. Available: https://github.com/davisking/dlib. [Accessed: 24-Oct-2018].

[5] “CVPR15-CFSS Github.” [Online]. Available: https://github.com/zhusz/CVPR15-CFSS. [Accessed: 24-Oct-2018].

[6] S. Kumar and Cypress Semiconductor Corp., “Efficient Heart Rate Monitoring,” *Cypress Semiconductor Technical Articles*, 2011. [Online]. Available: https://eu.mouser.com/applications/medical_efficient_monitoring/. [Accessed: 24-Oct-2018].

[7] M. Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, 2011.

[8] X. Li, J. Chen, G. Zhao, and M. Pietikäinen, “Remote heart rate measurement from face videos under realistic situations,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4264–4271, 2014.

[9] “REMOTE PHOTOPLETHYSMOGRAPHY USING NONLINEAR MODE DECOMPOSITION Halil Demirezen , Cigdem Eroglu Erdem Marmara University Department of Computer Engineering , Goztepe , Istanbul , Turkey,” pp. 1060–1064, 2018.

[10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” *Opt. Express*,

vol. 18, no. 10, p. 10762, 2010.
[11] H. Monkaresi, R. A. Calvo, and H. Yan, “A machine learning approach to improve contactless heart rate monitoring using a webcam,” *IEEE J Biomed Heal. Inf.*, vol. 18, no. 4, pp. 1153–1160, 2014.
[12] C. E. Erdem, S. Ulukaya, A. Karaali, and A. T. Erdem, “COMBINING HAAR FEATURE AND SKIN COLOR BASED CLASSIFIERS FOR FACE DETECTION Dept . Electrical and Electronics Engineering Dept . Electrical and Electronics Engineering,” pp. 1497–1500, 2011.
[13] P. Sahindrakar, “Improving Motion Robustness of Contact-less Monitoring of Heart Rate Using Video Analysis,” p. 52, 2011.
[14] A. Hyv and E. Oja, “Independent Component Analysis : A Tutorial 1 Motivation,” vol. 1, pp. 1–30, 1999.
[15] “Welch Method.” [Online]. Available: https://www.mathworks.com/help/signal/ref/pwelch.html. [Accessed: 01-Oct-2019].
[16] “Pulse rate detection dataset.” [Online]. Available: https://www.tu-ilmenau.de/neurob/data-sets/pulse/. [Accessed: 24-Oct-2018].
[17] “UBFC Dataset - Info.” [Online]. Available: https://sites.google.com/view/ybenezeth/ubfcrppg. [Accessed: 01-Jul-2019].
[18] “GitKraken.” [Online]. Available: https://www.gitkraken.com/. [Accessed: 24-Oct-2018].
[19] “Trello.” [Online]. Available: https://trello.com/. [Accessed: 24-Oct-2018].