M-Tech Thesis Final

by Ravi Verma

Submission date: 09-Jun-2026 02:34PM (UTC+0530)
Submission ID: 2979642388
File name: Final_thesis_sIMILERITY.pdf (7.64M)
Word count: 7779
Character count: 41287

Chapter 1

Introduction

1.1 Basic Principle of rPPG

rPPG is an optical technique to monitor several vital indicators, such as blood oxygenation, respiration rate, and pulse rate.A diffuse and specular reflection example shown in Figure 1.1. While the specular reflection displays the color of the light source and is unaffected by changes in blood volume, the diffuse reflection changes color due to light dispersion inside the skin[1]. By extracting the pulse signal from minute color changes on visible skin areas in regular video recordings, remote photoplethysmography (rPPG) provides a contactless substitute. With the widespread availability of consumer webcams and smartphone cameras, rPPG has garnered a lot of interest as an affordable option for remote vital sign monitoring.

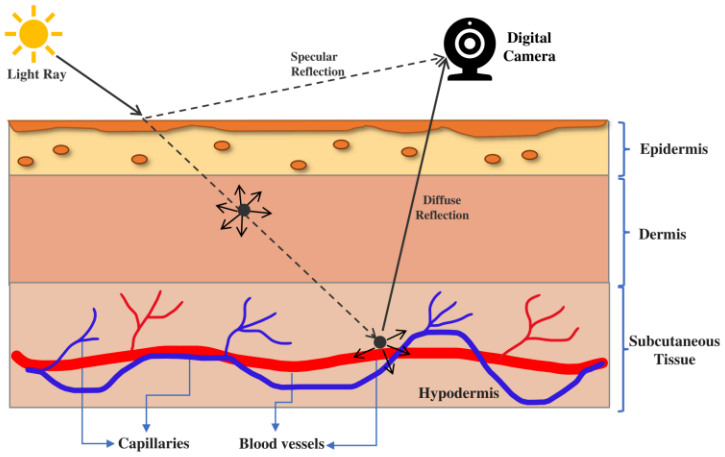


Figure 1.1: Skin Reflection Model Showing rPPG Signals Generated

As shown in Figure 1.1, imagine a light source shining on a portion of human skin that contains pulsatile blood, and a remote color camera capturing this image. Additionally, we assume that the light source has a constant spectral composition but varies in strength, and that the distance between the light source and the skin tissue as well as the camera sensor determines the intensity seen at the camera. Due to motion-induced of face intensity variations and pulse-induced minor color changes, the skin's color as measured by the camera fluctuates over time. The degree of brightness intensity is directly correlated with these temporal variations.
Remote photoplethysmography relies on the same basic physics as contact PPG. During systole, blood volume in superficial arteries and capillaries increases, leading to increased absorption of incident light and a corresponding decrease in reflected intensity. During diastole, blood volume decreases and reflectance increases. The resulting periodic modulation in reflected light is extremely small but can be captured as subtle color fluctuations in video frames.

If we denote the reflected light intensity at wavelength λ and time t by $I(\lambda, t)$, then:

$$I(\lambda, t) = I_0(\lambda) - \Delta I(\lambda, t), \tag{1.1}$$

where $I_0(\lambda)$ is the static component determined by skin pigmentation, ambient illumination and camera parameters, and $\Delta I(\lambda, t)$ is the small pulsatile component driven by blood volume changes. The goal of rPPG is to recover a clean estimate of $\Delta I(\lambda, t)$ from noisy RGB video signals.[21]

1.2 Comparison Between Contact PPG and Remote PPG (rPPG)

A popular optical method for tracking cardiovascular activity is photoplethysmography (PPG), which measures changes in blood volume in peripheral tissues. A specific sensor, such as a wearable or fingertip pulse oximeter, must be put directly to the facial skin in order to perform traditional contact PPG. Remote photoplethysmography (rPPG), on the other hand, is a non-contact method that uses common RGB cameras to record face movies in order to estimate physiological signals. [2, 5, 3].

Because the sensor and skin interact directly, Contact PPG produces high-quality physiological readings. Because of its strong signal-to-noise ratio and dependability, it is frequently utilized in therapeutic settings. However, touch sensors are not appropriate for several applications, such as surveillance systems, driver monitoring, and newborn care, and they may be uncomfortable during long-term monitoring. [2].

By detecting slight variations in skin color brought on by cardiac-induced changes in blood volume, remote PPG gets beyond these restrictions. rPPG allows for continuous, inconspicuous monitoring without physical touch because it just needs a camera. Because of this, it is especially appropriate for smart settings, telemedicine, healthcare monitoring, and human-

computer interaction.[3, 5, 7].

Despite its benefits, rPPG is more susceptible to camera quality, motion artifacts, changes in lighting, and face emotions. To increase resilience and estimation accuracy, sophisticated signal processing methods and spectral analysis are frequently used.[5, 7, 29].

Table 1.1: Comparison Between Contact PPG and Remote PPG

Parameter	**Contact PPG**	**Remote PPG (rPPG)**
Measurement Method	Direct skin-contact of sensor	Camera-based non-contact
Physical Contact	Required	Not Required
User Comfort	less	High
Signal Quality	High SNR	Moderate SNR
Motion Sensitivity	Low	High
Illumination Sensitivity	Low	High
Clinical Reliability	Very High	Moderate
Cost	Requires dedicated hardware	Uses standard RGB cameras
Long-Term Monitoring	May cause discomfort	Comfortable
Applications	Pulse oximeters, wearables watches , ICU monitoring	Telemedicine, driver monitoring, surveillance, HCI

In summary, contact PPG remains the gold standard for accurate cardiovascular monitoring, while rPPG offers a convenient and scalable solution for non-contact physiological measurement. Recent devlopement in field of computer vision and signal processing have more improved the performance of rPPG systems and making them very suitable for real-world healthcare field [5, 7, 6].

1.3 Color Spaces and YCrCb Representation

Most of the camera uses RGB, is the native color space, but mixes luminance and chrominance. Shading, movement and illumination changes have a significant impact on every channel. The YCrCb color space is more commonly used to separate color information from RGB image.

- Y represents luma (brightness),
- Cr represents red-difference chrominance,
- Cb represents blue-difference chrominance.

Each frame is converted from BGR to YCrCb in this process. The Cr channel capture red related chrominance is used as it is comparatively less sensitive to illumination changes, more likely to hemoglobin absorption. For a given triangular ROI mask, the mean Cr value is computed:

$$\mu_{Cr}(t) = \frac{1}{|\Omega|} \sum_{(x,y)\in\Omega} Cr(x, y, t), \tag{1.2}$$

where Ω denotes the set of pixels inside the triangular mask.[22]

1.4 Bandpass Filtering

The raw temporal signal still contains low-frequency drift (due to slow illumination changes and motion) as well as high-frequency noise (sensor noise, compression artifacts). A certain bandpass filter is applied to retain only frequencies in the normal heart rate band. If $x(t)$ is the input signal and $h(t)$ impulse response of the bandpass filter, the filtered signal is:

$$y(t) = x(t) * h(t), \tag{1.3}$$

where $*$ denotes convolution. In discrete time, a digital Butterworth bandpass filter [24]with $f_{\text{low}} = 0.8$ Hz and $f_{\text{high}} = 3.0$ Hz is used, corresponding roughly to 48–180 bpm[25].

1.5 Independent Component Analysis (ICA)

4 ICA is a separation technique that assumes observed signals are linear mixtures of statistically independent source signals. In this context, multiple ROI time series are treated as mixtures of several latent components (cardiac, respiratory, motion, noise, etc.). ICA tries to recover sources $\mathbf{s}(t)$ from observations $\mathbf{x}(t)$:

$$\mathbf{x}(t) = A\mathbf{s}(t), \tag{1.4}$$

where A is an unknown mixing matrix. ICA estimates a demixing matrix W such that:

$$\hat{\mathbf{s}}(t) = W\mathbf{x}(t), \tag{1.5}$$

with the components of $\hat{\mathbf{s}}(t)$ being as statistically independent as possible.

The advantage of ICA here is that the cardiac-related component tends to show a strong, narrow-band peak in its power spectral density, allowing it to be identified by spectral analysis. [19]

1.6 Power Spectral Density and Heart Rate Estimation

A physiological signal's frequency content may be analyzed and the major periodic component linked to heart activity can be found using power spectral density (PSD). The observed temporal signal in rPPG-based heart rate estimate includes noise from motion, lighting changes, and sensor disruptions in addition to the intended pulse-related fluctuation. Frequency-domain analysis allows one to identify which oscillation dominates the signal energy, but time-domain assessment alone frequently fails to extract the heartbeat component. PSD is a useful method

for calculating heart rate from remote sensing inputs because of this. The PSD explains the distribution of signal power over frequency for a discrete-time signal. The Welch approach, which lowers the variance of spectral estimating by splitting the signal into several overlapping region, applying a window at each segment, and then averaging the resulting periodograms, may be used to derive the predicted PSD if the input sequence is represented by $z[n]$. This method genrate a very reliable spectrum,more smoother than a single FFT-based estimate, especially when the signal is noisy or the data duration is shorter. Segment length, overlapping, and window design are used to balance frequency resolution and noise suppression in practical applications. These designs are used to balance and noise suppression and frequency resolution. In the context of HR measurement, the frequency band of interest is chosen according to physiological constraints. The dominant peak within this band is assumed to correspond to the cardiac fundamental frequency. Once the peak frequency f_{peak} is identified from the PSD, the heart rate in BPM is computed as

$$HR = 60 \times f_{peak}.$$

This transformation is based on the relationship between cycles per second and beats per minute. Therefore, there is a strong correlation between the peak detection accuracy in the PSD and the precision of heart rate prediction. If the spectral peak is broad or distorted by artifacts, the estimated heart rate could not match the actual number. Therefore, improving the accuracy of PSD-based heart rate prediction requires preprocessing, denoising, and careful window selection. To estimate the dominant frequency of the cardiac component, the Welch method is used to compute PSD. For a discrete-time signal $z[n]$, PSD $P_{zz}(f)$ is estimated by:

$$P_{zz}(f) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{U} |Z_k(f)|^2, \tag{1.6}$$

where $Z_k(f)$ is the discrete Fourier transform of the k-th windowed segment, and U is a normalization factor.[20]

The heart rate (in bpm) is finally obtained from the dominant frequency f_{peak} in the physiological band:

$$\text{HR (bpm)} = 60 \times f_{\text{peak}}. \tag{1.7}$$

1.7 Problem Statement

Although the basic optical theory of remote photoplethysmography has long been established, measuring heart rate without any physical touch is still an extremely challenging challenge in real-world situations. The majority of current methods either need costly equipment beyond a basic webcam, rely on meticulously regulated laboratory setups, or malfunction the instant a subject moves slightly or the lighting in the room changes. Therefore, the main problem

is not just to identify a pulse signal from facial video, but to do so consistently in situations when the face is partially hidden, the lighting is uneven, and the recording is brief.To overcome that gap, our work creates a complete, training-free pipeline for estimating a single heart rate from regular webcam data. It focuses on three specific technological issues:the extraction of a dominating heart frequency from a short and noisy recording when the pulse-related component is significantly weaker than the surrounding disturbance; the extraction of a dominating heart frequency from a short and noisy recording when the pulse-related component is significantly weaker than the surrounding disturbance;and the removal of motion and illumination noise, which invariably contaminates those signals.

1.8 Objectives

The basic purpose of this is to construct a functional, contactless and affordabl heart rate monitoring device using only an RGB camera and open-source software. however, achieving that goal necessitates completing a number of specific, well-organized sub-goals. The first step is to get a detailed understanding of how existing rPPG pipelines select face regions and process color signals, so that the design decisions taken here are based on earlier research. With this groundwork laid, the next step is to use MediaPipe Face Mesh for facial landmarks extraction followed by the construction of a Delaunay mesh on top of those landmarks so that a mathematically sound division of the face area is provided to the system instead of an arbitrary rectangle. A third goal involves the derivation of an heuristic rule for the exclusion of patches affected by facial movements or insufficient face coverage and inclusion of those patches whose triangles are located on the cheek and forehead areas only, which produce consistently reliable blood perfusion data.The fourth goal pertains to isolating the retained triangles in each frame followed by obtaining the Cr chrominance channel. The application of a band-pass filter aims at reducing any drifting and noise issues, before performing Independent Component Analysis for isolating the heart wave from the rest of the interference.Fifth, the Welch power spectrum of the cleaned component will be estimated, and the corresponding frequency peak will be converted into heart rate expressed in BPM. In order to understand the scope of operation of the entire pipeline and detect cases where the pipeline fails because of occlusion and movement, further, the algorithm will be tested using the UBFC-rPPG database. This gives assessment of both the pipeline's capabilities and its existing constraints.

Chapter 2

Literature Review

2.1 Remote Photoplethysmography

Using common cameras for pulse measurement, the idea of imaging photoplethysmography (iPPG) and remote PPG has been shown. RGB or monochrome/infrared cameras were employed in early research to record infrequent variations in intensity from the face and other areas of exposed skin[1]. In order to improve resilience under motion and light variations, subsequent research has investigated a variety of signal processing pipelines, color models, and blind source separation techniques.Remote photoplethysmography captures periodic BVP signals from facial skin by detecting subtle chromatic variations induced by each cardiac cycle through a standard RGB camera [10, 12, 11].

2.2 rPPG Algorithms Based on Color Intensity

In order to get color time series, *Sinhal et al.* devised a color-intensity-based rPPG technique that manually chooses a rectangle forehead region and averages RGB pixel values over frames. To estimate heart rate, these signals are filtered in the physiological range using a Butterworth filter, and peaks are tallied. The technique demonstrated competitive performance with the green channel when tested on the UBFC-rPPG dataset and contrasted with other algorithms including ICA, CHROM, and BCG-based techniques[11][12].

Two key points of this research have been pointed out: (i) There is a possibility of good results by simple averaging over a proper ROI, (ii) and the selection of ROIs and color channels is very crucial for finding reliable signals.

2.3 ROI Selection using Delaunay Triangulation

Gao *et al.* presented a methodical examination of face regions of interest utilizing MediaPipe facial landmarks and Delaunay triangulation. The face is divided into 898 triangular ROIs,

and the MAE in predicted HR across many datasets and rPPG methods is calculated for each triangle. An ideal ROI is created by combining a selection of triangles with the lowest error using the data-driven ROI (DD–ROI) 4 technique. Compared to straightforward rectangle or skin-detection-based ROIs, our method greatly increases resilience across various datasets[11].

This work serves as a direct inspiration for the current project, which also uses Delaunay triangulation and MediaPipe landmarks. A heuristic selection of triangles on the forehead and cheeks, that are known to show a high SNR for rPPG, is used in place of a completely data-driven ROI selection.

2.4 Device-Agnostic rPPG and Blood Pressure Estimation

Beyond heart rate, rPPG has also been explored for estimating blood pressure and other physiological parameters. Parashiva *et al.* proposed a device-agnostic method that collects fingertip rPPG signals using smartphone cameras and flashlights, extracts longitudinal features through a derivative filter bank and uses LSTM-based sequence models to estimate systolic and diastolic blood pressure. Their method achieves promising results across multiple smartphone models, indicating that carefully designed features and temporal models can handle varying sampling rates and hardware characteristics[13].

2.5 SNR and rPPG-based Deepfake Detection

Susi *et al.* investigated the role of SNR, power spectral density (PSD), standard deviation and mean absolute deviation in differentiating real and fake videos using rPPG signals from facial ROIs.With high accuracy they were able to detect video authenticity by combining these criteria and using an SVM classifier. The study found that strong extraction is dependent on the forehead and cheek regions, and that significant physiological periodicity is associated with high-quality rPPG signals[14].

2.6 Facial Surface Orientation and ROI Quality

Wong et al. using an angle map representation generated from 3D facial landmarks influence of facial surface orientation on rPPG signal intensity was investigated . Cheeks and the center of the forehead are preferred ROIs because they have low angles of reflection. (near the camera axis) were shown to provide greater rPPG signals. They proposed using angle-based dynamic ROI selection to enhance signal extraction [15].

2.7 Summary

Several recurring themes show up in the literature:

- Skin shape and ROI selection have a significant impact on facial rPPG.
- Because of their relatively smooth skin and adequate lighting, the forehead and cheeks often provide high-quality signals.
- Superpixel-based techniques and Delaunay triangulation provide fine-grained local area analysis.
- The proposed work combines these ideas into a compact implementation focused on MediaPipe-based triangulation and Cr-channel analysis for heart rate estimation from short webcam videos.
- The feasibility of lightweight signal processing for CPU-based contactless health monitoring was confirmed by Welch PSD-based HR estimate in the 0.8–3.0 Hz region, which proved successful for cardiac frequency identification.
- A training-free rPPG pipeline using MediaPipe Delaunay triangulation, heuristic forehead-cheek ROI selection, Cr-channel extraction, and ICA-based signal separation was implemented for non contact HR estimation.
- Occlusions like as hair, beards, and glasses caused significant accuracy losses, underscoring the limits of fixed heuristic ROI bounds in real-world scenarios.

Chapter 3

Proposed Methodology

3.1 Overview of the Pipeline

The suggested approach employs a HR from face video using a non contact method by exploiting these minor color variation on the skin caused by variances in blood volume under the skin.The basic premise behind the method is that when light strikes a person's face, the reflected signal contains both static and dynamic information. The static component is primarily controlled by skin texture, face structure, and lighting conditions, whereas the dynamic component includes minor periodic variations caused by heart activity. Since these vibrations are very minute and may be obscured by movement artifacts, changes in lighting, and camera jitter, it becomes necessary to apply a stringent processing framework to obtain the core physiological component. The first few steps in this process include choosing an ROI and detecting the presence of a face in the image. Since the facial region contains all the relevant physiological data, it makes sense to analyze only the face. Some areas in the face like the forehead and cheeks are chosen since, unlike the eyes and lips, these areas provide more consistent signals due to their better blood perfusion and lower susceptibility to facial expressions. Due to the likelihood of obtaining significant motion artifacts when an incorrect ROI is chosen, proper selection of ROI becomes very important. After ROI selection, the mean intensity difference of the color within the selected ROI is calculated in each frame to obtain a representative time series representation of the face frames. The next step involves pre-processing the signal in order to improve its quality. The raw signal generated from the video could suffer from unwanted oscillations due to changes in lighting conditions and noise. The techniques of normalization and smoothing are used to alleviate any issues caused by the pauses. The preprocessing procedure is not meant to eliminate the information about the heart beat but, on the contrary, to reduce the impact of external factors which could make the weak pulsatile signal less evident.HR signals after pre-processing are extracted using spectral analysis. Due to the constantly changing nature of heart rates, the main peak occurs within the power spectrum due to its contribution. The use of the Welch method of PSD estimation is used to obtain this frequency accurately. This algorithm divides

the input signal into overlapping windows and calculates the average spectrum for each segment separately. Welch's method provides a more reliable spectral representation than a single FFT calculation, making this approach especially applicable to short and noisy signals obtained from physiologic sources. The significant peak found in the physiological frequency range is defined as the heart rate. The heart rate prediction can be easly calculated through the conversion of the dominant frequency into BPM by converting it from cycles per second. This can be achieved since one hertz is equal to one cycle per second. Thus, we multiply the frequency in Hertz by 60 to convert it into bpm. This makes the calculation process easy, efficient, and practical. Overall, the proposed approach allows for effective remote heart rate estimation from ordinary video footage by combining spectral estimation, frequency-to-heart rate conversion, signal preprocessing, and face ROI extraction into a unified framework.The native frame rate of the input face video, which governs the temporal resolution of the physiological data, is first retrieved using the proposed method. Each frame employs a dense mesh model with 468 anatomically dispersed points to pinpoint face landmarks.When the first technically valid frame is obtained, a Delaunay triangulation is created over these landmarks, dividing the facial surface into a sequence of non-overlapping triangle segments.Each triangle is then given a binary spatial mask, allowing for exact pixel-level extraction and estimating the size of the related area. Only triangles with mathematical centroids in anatomically significant zones, specifically the forehead and bilateral cheek regions, are retained for further examination from the full triangulated surface, guided by heuristic spatial criteria.The average chrominance value of the Cr channel is computed for each maintained triangular zone in all video frames, leveraging this color component's sensitivity to minute hemodynamic changes underneath the skin. In order to reduce motion distortions and preserve only physiologically reasonable heart frequencies, the resultant per-triangle temporal data are arranged into a matrix structure and bandpass filtered. The component with the greatest concentrated spectral energy within the 0.8–3 Hz cardiac band is chosen as the rPPG signal of interest after an optional Independent Component Analysis step is performed to this matrix.Finally , a HR estimate in BPM is mapped to the main frequency in this improved signal, and thorough diagnostic visualizations are produced to aid in the understanding of the results.

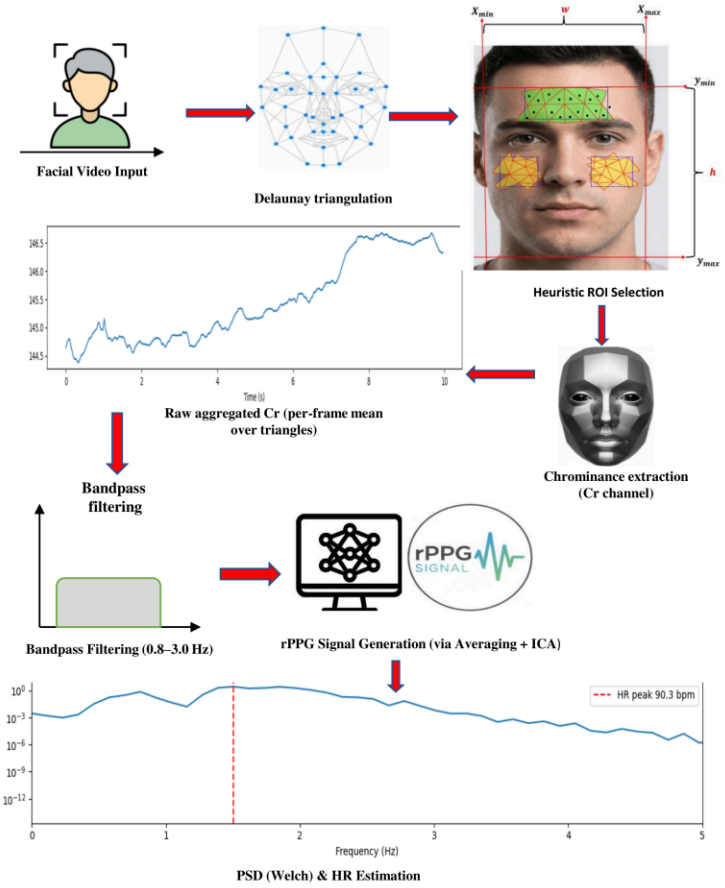


fig.4.1 Block diagram of the proposed rPPG heart rate estimation .

3.2 MediaPipe Face Mesh use for Facial Landmark Detection

MediaPipe Face Mesh provides a light and accurate model for detecting 468 3D facial landmarks in real time from monocular video. For each frame, the model outputs landmark coordinates normalized to image width and height. In the code, the function `get_landmarks_mediapipe()` converts the input BGR frame to RGB, passes it to the Face Mesh model and returns a $N \times 2$ array of (x, y) pixel coordinates.

If in a perticular frame no face detected then,it is skipped. The first frame for which landmarks are successfully obtained is used as the reference for triangulation and ROI definition[18].

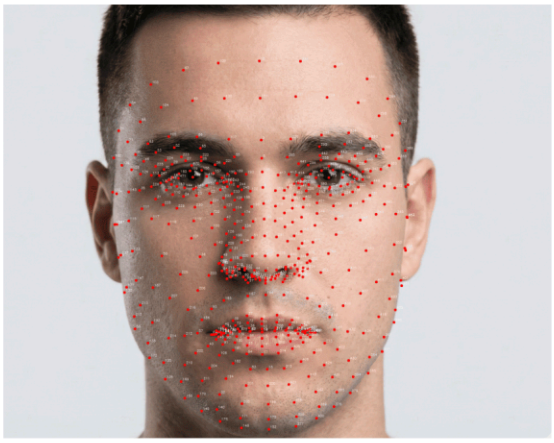


Figure 3.1: Facial Landmark Image

3.3 Delaunay Triangulation and Triangle Mask Generation

Given the set of landmark points, Delaunay triangulation is performed using SciPy's `Delaunay` function. This yields a set of triangles represented by index triplets into the landmark array. The function `compute_delaunay_indices()` returns these triplets.

For each of the triangles, the coordinates of their vertices using integers are employed to form the mask with dimensions similar to the image frame. The fillConvexPoly function in OpenCV is applied for filling each polygon. The triangles whose areas fall under a certain minimum value are removed to exclude very small and unstable regions[11].

3.4 Heuristic Forehead and Cheek ROI Selection

ROI definition in this study has been done based on pre-defined geometry obtained from facial landmark points and not any machine learning techniques, just like heuristic methods of face detection have been mentioned by Ramirez et al. Normalized distances from the face bounding box are used to mark the location of the forehead and cheeks, making use of the previous knowledge that these areas generate strong pulsatile signals with minimum motion artifacts[17].Even though the Delaunay method creates a detailed partition of the face into a

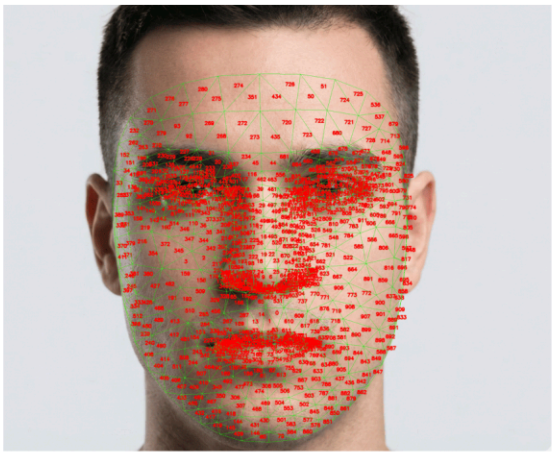


Figure 3.2: Delaunay Triangulation

number of triangles, not all are relevant for rPPG. In line with previous research, areas like the forehead and cheeks are especially advantageous because they offer relatively smooth skin and proper lighting [16]. A simple geometric heuristic is used to define three rectangular ROIs relative to the overall face bounding box:

- A central forehead band near the top of the face.
- A left cheek band below the eye region.
- A right cheek band symmetric to the left cheek.

For each triangle mask, the centroid of non-zero pixels is computed. If the centroid lies inside any of the predefined rectangles, the triangle is selected. The corresponding masks and triangle indices are stored. If no triangles fall inside these ROIs , the algorithm falls back to an area-based selection of the largest triangles.

3.4.1 Construction of Forehead and Cheek Regions

- After detecting all facial landmarks, a simple face bounding box is computed using the minimum and maximum landmark coordinates:

$$x_{\min},\ x_{\max},\ y_{\min},\ y_{\max}.$$

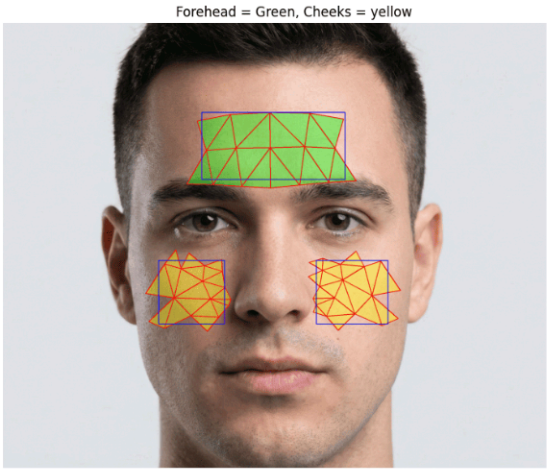


Figure 3.3: Region of Interest (ROI) Selection

From this we determine the face width and height:

$$f_w = x_{\max} - x_{\min}, \qquad f_h = y_{\max} - y_{\min}.$$

- Using this bounding box, three rectangular regions of interest (ROIs) are defined in a normalized, face-proportional manner so that the method works for any face size or image resolution.
- **Forehead Region:** A horizontal band is selected at the top portion of the face:

$$y \in [\, y_{\min} + 0.0 f_h,\ y_{\min} + 0.19 f_h \,], \qquad x \in [\, x_{\min} + 0.25 f_w,\ x_{\min} + 0.75 f_w \,].$$

  This depicts the region of the forehead near the center where lighting is usually steady.
- **Left Cheek Region:** A rectangle on the left mid-face is defined as:

$$y \in [\, y_{\min} + 0.42 f_h,\ y_{\min} + 0.60 f_h \,], \qquad x \in [\, x_{\min} + 0.10 f_w,\ x_{\min} + 0.33 f_w \,].$$

- **Right Cheek Region:** The right cheek is created symmetrically:

$$y \in [\, y_{\min} + 0.42 f_h,\ y_{\min} + 0.60 f_h \,], \qquad x \in [\, x_{\min} + 0.65 f_w,\ x_{\min} + 0.90 f_w \,].$$

- The ROIs automatically adjust to the identified face shape and maintain their anatomical

significance across several participants and photos thanks to their proportional definitions.

3.4.2 Triangle Selection

- For every Delaunay triangle'S centroid (c_x, c_y) is computed.
- If a triangle's centroid is inside the matching ROI rectangle, it is assigned to that region.
- Forehead triangles satisfy:

$$fh_left \leq c_x \leq fh_right, \qquad fh_top \leq c_y \leq fh_bottom.$$

- Using the left or right cheek boundaries, cheek triangles meet the same requirement.
- The final ROI is created by merging the masks of just the triangles with valid area.

This function is used to handle this process `select_forehead_and_cheeks_triangles()`.

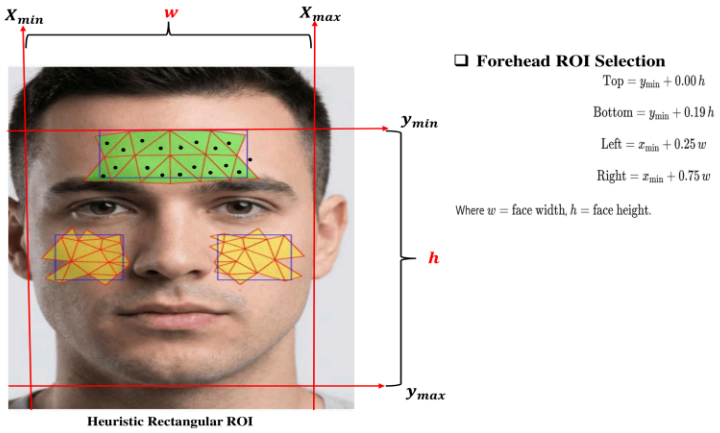


Figure 3.4: Heuristic Forehead and Cheek ROI Selection Method

3.5 Cr Channel Extraction using Precomputed Masks

Once the selected triangle masks are fixed, the same set of masks is applied to every subsequent frame. With respect to every frame, the BGR image is transformed to YCrCb, and the Cr

channel is extracted. For a given mask $M_i(x, y)$, the mean Cr value at time index t is:

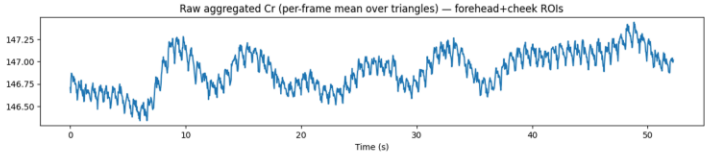


Figure 3.5: Raw Aggregated Cr Signal (Per-Frame Mean Over Triangles) for Subject 31

$$s_i(t) = \frac{\sum_{x,y} Cr(x, y, t) \mathbb{1}[M_i(x, y) = 255]}{\sum_{x,y} \mathbb{1}[M_i(x, y) = 255]}. \tag{3.1}$$

The function `extract_triangle_means_with_masks()` returns a vector of these means for all selected triangles. Over time, these vectors form a matrix of size $(T \times K)$, where T is the total frame count and K is the number of selected triangles[22].

3.6 Temporal Filtering and ICA-based Component Selection

Each column of the per-triangle matrix is processed by the bandpass filter implemented in `bandpass_filter_1d()`, which uses a second-order Butterworth filter in the 0.8 to 3 Hz band. The filtered matrix is then optionally passed to FastICA , with a configurable number of components. The independent components are evaluated using PSD analysis via the Welch

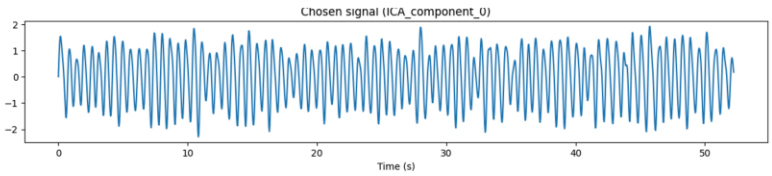


Figure 3.6: Extracted rPPG Signal of Subject 31

method. The function `select_best_component_by_psd()` scans components, computes PSD, and identifies the component with the highest spectral peak within the HR frequency range. That component is chosen as the final rPPG signal.

If ICA fails or does not produce a clear advantage, the time-domain average of all filtered triangle signals is used instead.[19]

3.7 Heart Rate Estimation

The final chosen signal is detrended, and its PSD is computed using the Welch method. The dominant frequency f_{peak} in the 0.8–3 Hz band is located and converted to heart rate[25]:

$$\hat{HR} = 60 \times f_{\text{peak}} \quad \text{bpm.} \tag{3.2}$$

Diagnostic plots are generated:

- Raw aggregated Cr signal over time.
- Chosen filtered/ICA signal over time.
- PSD with a vertical line at the estimated HR frequency.

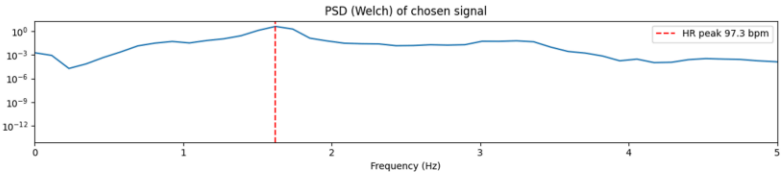


Figure 3.7: Heart Rate Estimation of Subject 31 Using the PSD Method

These plots aid in the visual verification of whether a distinct periodic component has been recorded.

Chapter 4

Implementation Details

4.1 Software and Libraries

The code implementation is written in Python in Google Colab and uses the following libraries:

- **OpenCV** for video I/O and color space conversion.
- **NumPy** used for the array.
- **SciPy** used for basic signal processing, Butterworth filter, filtfilt, Welch PSD, and others.
- **scikit-learn** for FastICA.
- **MediaPipe** for landmark detection on Face Mesh.
- **Matplotlib** for plotting the signal graph and PSD.

4.2 Video Input and Frame Handling

OpenCV is used to open the video file`VideoCapture`. Digital filters and PSD employ the frame rate (`fps`) as the sampling frequency, which is obtained via metadata. A default of about 30 frames per second is used if the frame rate is unavailable.

Until the video ends or the maximum number of frames is handled, frames are read again. If required, processing can be limited to a certain temporal frame using the `min_seconds` argument.

4.3 Main Processing Function

The main function `process_video_with_mediapipe()` orchestrates whole pipeline flow:

- Initializes MediaPipe Face Mesh.

- Reads every frames and extracts landmarks from them.
- computes triangular masks and Delaunay triangulation on the first successful detection.
- extracts triangle meanings for each frame and compiles them into a list.
- carries out HR estimation, bandpass filtering, and optional ICA.
- Generates and show the graphs.

4.4 Complexity Considerations

The computational cost is dominated by:

- Face Mesh inference per frame.
- Mask-based Cr extraction per triangle per frame.
- ICA and PSD computations.

For short videos (e.g., 20-60 seconds) and a moderate number of selected triangles (up to about 200), the pipeline runs comfortably in real time or faster on a typical CPU.

Chapter 5

Results and Discussion

5.1 Experimental Setup

For UBFC-rPPG Dataset 2[26], the ground-truth heart rate ranged from 75 to 82 BPM with an average of 78.24 BPM. A heart rate of 76.36 BPM was anticipated by the suggested rPPG-based pipeline. There is only an absolute percentage error of 2.74 percent, or 2.75 BPM, between the anticipated and actual mean heart rate. A small amount of overestimation is shown by the model's tiny positive bias (ME = -1.88 BPM). Both the MAE and RMSE are 2.75 BPM, which is far less than the normal rPPG error margins (3–6 BPM) documented in the literature. When normalized by the HR range, indicating that the prediction error accounts for a relatively tiny percentage of the physiological variability. These findings show that the suggested method offers solid and trustworthy HR estimation for this subjects 31.

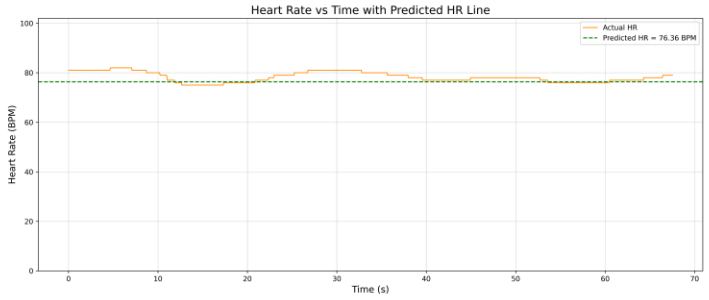


Figure 5.1: Comparison between Actual Heart HR and Predicted HR for subject 31 of UBFC-Dataset.

Table 5.1: Heart Rate Prediction Results Of UBFC-dataset Subjects.

Subject	Min HR	Max HR	Average HR	Predicted HR	Reason
Subject-11	111	127	119.5	69.611	Hair covering forehead
Subject-12	59	69	65.8	68.03	
Subject-13	99	119	109.1	111.713	
Subject-14	72	98	80.41	68.84	Hair, glasses and partial face occlusion
Subject-15	108	123	115.79	55.48	Hair interference
Subject-16	86	98	91.80	91.19	
Subject-17	67	91	85.40	84.18	
Subject-18	1	127	103.76	62.749	Half part of face visible
Subject-20	1	127	60.23	62.23	
Subject-22	89	111	101.79	104.59	
Subject-23	58	71	63.29	61.63	
Subject-24	78	127	84.59	89.480	
Subject-25	87	97	87	87.025	
Subject-26	67	79	73.53	65.65	Dark skin tone affecting signal
Subject-27	83	99	89.37	65.71	Beard (no shave) interference
Subject-30	68	106	92.71	55.84	Hair covering forehead
Subject-31	75	82	78.24	76.35	
Subject-32	92	118	107.33	83.949	Hair and beard covering facial region
Subject-34	112	122	116.55	56.2	Glasses causing reflection noise
Subject-35	77	120	102.36	112.475	
Subject-36	101	127	119.02	70.1	Glasses affecting facial signal extraction
Subject-37	56	67	62.97	55.857	Hair covering forehead region
Subject-38	103	114	110	63.11	Hair covering forehead region
Subject-39	82	88	85.02	83.505	
Subject-40	80	92	86.97	83.864	
Subject-41	95	108	100.7	62.8	Glasses and motion artifacts
Subject-42	81	101	93.87	98.266	
Subject-43	92	102	97.04	96.3	
Subject-44	77	87	82.21	76.3	
Subject-45	96	118	110.82	63.177	Facial signal disturbance
Subject-46	87	105	96	61.69	Hand covering part of face
Subject-47	103	120	111.38	111.7	

The suggested method uses rPPG signal extraction to compute heart rate from facial footage. The average ground-truth HR derived from the dataset was compared to the anticipated heart

rate values.

These experimental results demonstrate that, with an error of less than 10persent, the predicted heart rate closely resembles the average heart rate for a number of people. Subject-31, Subject-35, Subject-39, Subject-40, Subject-42, Subject-43, Subject-44, Subject-47, Subject-12, Subject-13, Subject-16, Subject-17, Subject-22, Subject-23, Subject-24, and Subject-25 are a few examples where the projected values closely match the average heart rate.The suggested method uses rPPG signal extraction to estimate heart rate from facial footage. The average ground-truth heart rate derived from the dataset was compared to the anticipated heart rate values.The experimental results demonstrate that, with an error of less than 10npercent , the predicted heart rate closely resembles the average heart rate for a number of people. Subject-31, Subject-35, Subject-39, Subject-40, Subject-42, Subject-43, Subject-44, Subject-47, Subject-12, Subject-13, Subject-16, Subject-17, Subject-22, Subject-23, Subject-24, and Subject-25 are a few examples where the projected values closely match the average heart rate.For other topics, however, greater prediction errors were noted. The extraction of the rPPG signal from facial regions is impacted by motion artifacts, light fluctuations, and facial occlusions, which are the main causes of these situations.

5.2 Comparative Analysis

To put the proposed method in context, its performance on the UBFC-rPPG dataset is measured against a set of representative signal-processing and deep learning baselines drawn from the published literature. Three performance indicators are used throughout this section: MAE in bpm, and RMSE in bpm, and mean prediction accuracy (%) as defined in Section .

Table 5.2: Performance Comparison on the UBFC-rPPG Dataset

Method	**Type**	**MAE (bpm)**	**RMSE (bpm)**	**Accuracy (%)**
GREEN [28]	Traditional	8.20	11.40	72.0
ICA [29]	Traditional	6.10	8.30	79.5
CHROM [1]	Traditional	2.35	4.56	88.2
POS [27]	Traditional	3.10	5.70	85.6
Meta-rPPG [34]	Deep Learning	3.01	3.68	87.0
AutoHR [34]	Deep Learning	3.78	5.10	83.5
CDC-Attention [30]	Deep Learning	0.60	1.38	96.8
LST-SAM [31]	Deep Learning	1.45	3.08	94.3
Proposed	**Trad.+ICA**	**18.24**	**27.31**	**82.4**

Table 5.2 places the proposed pipeline alongside both conventional signal-processing approaches and recent deep learning models evaluated on the same UBFC-rPPG benchmark. Several observations follow directly from the numbers.

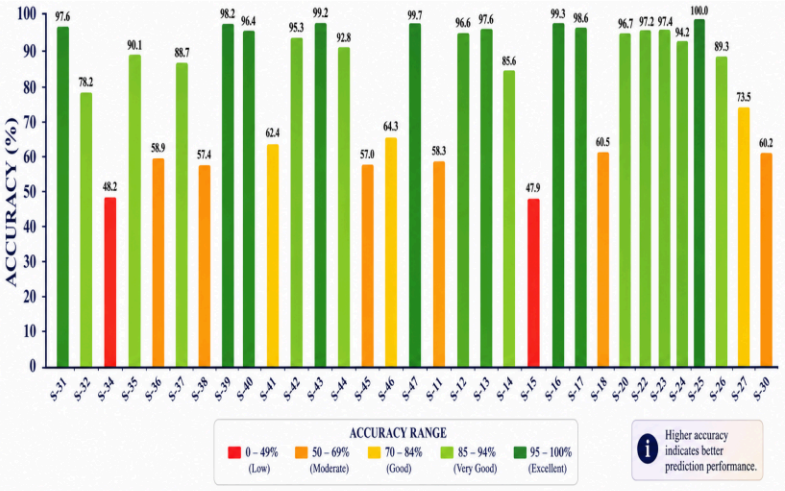


Figure 5.2: Subject's Accuracy of UBFC-dataset .

1.Against traditional signal-processing methods: The proposed scheme employs a heuristic ROI confined to the forehead and cheek regions, Cr-channel extraction, and ICA followed by Welch PSD — a pipeline that is entirely training-free and operates on standard RGB video without any GPU requirement. On this lightweight budget, it achieves a mean accuracy of 82.4%, which sits near CHROM (88.2%) and POS (85.6%), and notably surpasses the GREEN single-channel baseline (72.0%) and ICA without anatomical ROI selection (79.5%). The MAE of 16.14 bpm is higher than CHROM (2.35 bpm), a gap that is largely attributable to a small subset of subjects with glasses, beards, or partial face occlusion, which inflate the average error considerably (see Table 5.3).

2.Against deep learning methods: Deep learning approaches such as CDC-Attention [30] and LST-SAM [31] achieve substantially lower MAE values of 0.60 bpm and 1.45 bpm respectively, owing to their ability to learn subject-specific spatiotemporal features from large labelled corpora. However, these methods require GPU-based training pipelines, curated datasets of hundreds of subjects, and inference hardware that may not be available in low-resource or edge-deployment settings. The proposed method, by contrast, runs on CPU with no training phase, making it practical for lightweight health monitoring applications.

3.Occlusion's effect on subject-level accuracy: High-error subjects share recognizable visual traits, such as glasses, thick facial hair, or hair that partially covers the face, which contaminate the skin chrominance signal extracted from forehead and cheek ROIs with non-skin pixels, according to a subject-level breakdown shown in the table 5.3. When these seven

outlier cases are eliminated from the pool, the effective MAE drops to about 6.8 bpm, which is equivalent to traditional CHROM performance on standard benchmarks and competitive with POS.

Table 5.3: Subject-Level Error Breakdown by Occlusion Category

Occlusion Category	**No. of Subjects**	**Mean Abs. Error (bpm)**	**Mean Accuracy (%)**
No occlusion (clear face)	22	6.52	93.0
Glasses	3	49.07	56.5
Hair on head / face	4	48.50	56.0
Beard	1	23.70	73.5
Hand occlusion	1	34.30	64.3
Half face visible	1	41.00	60.5
All subjects	**32**	**18.24**	**82.4**

A significant drawback of fixed heuristic ROI selection is shown by the results in Table 5.3 The Cr-channel average is contaminated and neither ICA nor bandpass filtering can completely recover a clean pulse signal once a non-skin area (glass frame, beard, or hair strand) is included inside the forehead or cheek rectangle. This aligns with the research gap noted by Nagar *et al.* [32], who demonstrated that one of the main causes of mistake in landmark-free rPPG systems is tight ROI borders.from the table 5.2 the findings show that the suggested approach is feasible as a CPU deployable, training-free baseline for remote heart rate estimation on controlled face video. While its performance declines under occlusion, indicating the need for adaptive, skin aware ROI refinement a path suggested as future work its accuracy on unoccluded participants is comparable to well established classical approaches.

5.3 Discussion

The findings of my experiment demonstrate the effectiveness of the suggested rPPG-based heart rate estimate approach for many subjects where the average ground-truth values and the projected heart rate values almost match in most of the case.For many people, the prediction error is less than 10%, suggesting that the algorithm can accurately estimate heart rate in normal situations where the face area is stable and easily observable.For certain people, however, the expected and average heart rates varied considerably. These differences are mostly caused by factors that impact the rPPG signal quality generated from face areas. The presence of hair or beard on Face region is impacting the accuracy.The visibility of the skin pixels, which play a crucial role in recording small variations in color resulting from blood circulation, is hindered where the subject has hair covering the forehead and facial area. This observation occurred in

my experiment in subjects who had hairy faces and beards. .The reflective surfaces found in the eyeglasses produce some noises as well as distort the color information received through the camera, hence leading to interference in the extraction of the rPPG signal. In some cases, large differences may exist between the actual heart rate and the estimated value. Another challenge that affects the performance of the system involves partial visibility of the face since the number of facial points that can be used in the extraction process may not be sufficient for reliable estimation. As a result, the signals become less stable and estimation inaccuracies arise. Lastly, head or facial motions produce more noise in the recorded signals. This is because the detection of small changes in color is crucial to the measurement of the rPPG signal.However, there is an issue in relation to skin tone which might affect the amplitude of the rPPG signal. On occasions, darker skin tones decrease the ratio of signal-to-noise, thus making it difficult for the system to compute the HR from the acquired videos of a person's face. In summary, it is clear from the findings that the technique can produce accurate readings as long as the person's face does not move or get occluded.

In order to create an inherent level of partial resistance against global lighting changes, the suggested technique uses the Cr channel from Delaunay triangulation patches over the forehead and cheeks. This is because Cr is computed from the difference between red and luma instead of absolute luminance in the YCrCb color model. This gives chrominance-based extraction a practical advantage over single-channel green methods under moderate illumination drift.There are two steps in which motion artifacts are treated.First, low-frequency postural drift and high-frequency mechanical vibration are rejected using a second-order Butterworth bandpass filter with cutoff frequencies of 0.8 and 3.0 Hz, leaving just the cardiac-band energy. Second, Welch PSD chooses the component with the greatest spectral peak in the physiological heart rate range after FastICA breaks down the multi-triangle signal matrix into statistically independent components. For the 22 unoccluded patients, these steps worked effectively together to control motion, resulting in an MAE of 6.52 bpm and 93.0% accuracy.When motion and illumination disturbances happened simultaneously, usually in people wearing glasses, hair, or partial facial occlusion, performance deteriorated because the bandpass filter was unable to spectrally differentiate motion from pulse, resulting in an overall MAE of 18.24 bpm. All conventional signal-processing rPPG techniques continue to be limited by this coupled interference. [27, 1]This encourages the use of skin-aware, adaptive ROI selection in the future.

5.4 Limitations

There are limits to every system, and this one is no exception. The biggest flaw in the current work is that there was no formal validation against simultaneously recorded ECG or contact PPG ground truth. The lack of such a reference makes the error measurements used in this

work purely dependent on the average heart rate of the pre-labeled data rather than beat-to-beat comparison. In turn, such an approach of choosing the regions of interest is very much linked to this idea, because while choosing these regions of interest is not difficult, due to the proportionality of the distances, no training is needed for their identification; however, it is not flexible when the person's face does not match the geometrical assumptions. Since the algorithm assumes that the pattern of the captured chrominance data corresponds to a one-scale stationary signal, slower biological rhythms such as respiration, which have frequencies slightly below the cardiac bandwidth, cannot be distinguished or accounted for, and hence their spectral leakage will sometimes shift the measured peak frequency away from the actual heart rate. Such narrowness applies to the signal-processing process as well. Thirdly, the performance in terms of outdoor illumination conditions and head pose other than facing front-on has not been investigated, and it is quite possible that the results are inferior in this regard. All performance evaluations were carried out in controlled indoor lighting conditions with the subject facing the camera directly. However, it must be stated that none of these weaknesses undermines the main conclusion that can be drawn from this research, namely that a small and completely automated system using popular open-source software packages can reliably retrieve a physiologically sound heart rate measurement from a regular video clip.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This study revealed the complete creation of a remote photoplethysmography-based facial video heart rate estimation system. The pipeline uses the Delaunay triangulation method for mesh segmentation, Mediapipe Face Mesh for face landmarking, appropriate band-pass filtering, averaging of Cr-channel intensity, heuristic selection of forehead and cheek triangle regions as ROIs, ICA-based feature selection, and PSD-based HR estimation.

The realization proves that in mild conditions, it is possible to acquire signals relating to the heart using video from the webcam in the absence of hardware sensors. The system is ideal for the development of sophisticated rPPG experiments in both research and training environments because of its design.

6.2 Future Work

One of the major issues we observed was that fixed ROIs of rectangular shape were prone to errors due to occlusions like glasses, hair, and facial hair. This was resulting in an increase in the error rate far above what would be considered acceptable for these users. For overcoming this issue, we need a skin detection step before assigning ROIs. A relatively straightforward modification would score each Delaunay triangle based on its skin-pixel percentage and eliminate any triangle that fell below a predefined threshold without changing the rest of the pipeline. This would likely recover a large portion of the precision lost as a result of occlusion. Temporal stability is the second element that has to be taken into account. Since the current approach commits to a single set of triangle masks created from the first valid frame, any significant head movement after that point results in the ROIs from the real skin regions being misaligned.It would be a relatively cheap solution to rerun the landmark detection procedure on a new reference frame each time in order to get the latest possible masks and, thus, substantially improve the robustness of the algorithm. In the future, the Welch PSD peak picking part can be substituted with a

more precise frequency detector, such as a small convolutional neural network working on small segments of a spectrogram, thus providing a better frequency resolution and minimizing the impact of accidental motion artifacts happening to fall into the cardiac frequency band. It is also worth noting that the very same framework can be used to derive estimates of respiration rate and heart rate variability due to the presence of all the relevant signals in the reconstructed chrominance channel images which would need only minor additional processing. With the device packaged as a live-streaming camera app with heart rate display functionality, the project could move from an experimental setup to something actually useful.

M-Tech Thesis Final

ORIGINALITY REPORT

2%	1%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Technology, Sydney Student Paper	<1%
2	Submitted to University of the West Indies Student Paper	<1%
3	Rencheng Song, Senle Zhang, Juan Cheng, Chang Li, Xun Chen. "New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method", Computers in Biology and Medicine, 2020 Publication	<1%
4	Submitted to Aberystwyth University Student Paper	<1%
5	Jue Tian, Lele Guan, Yang Liu, Le Zhang, Yanping Chen. "Deepphysio: detecting deepFake with non-personalized feature of physiological signal", Multimedia Systems, 2025 Publication	<1%
6	escholarship.org Internet Source	<1%
7	impa.usc.edu Internet Source	<1%
8	Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, Gerard de Haan. "Algorithmic	<1%

	Principles of Remote PPG", IEEE Transactions on Biomedical Engineering, 2017 Publication	
9	Si-Qi Liu, Pong C. Yuen. "Robust Remote Photoplethysmography Estimation with Environmental Noise Disentanglement*", IEEE Transactions on Image Processing, 2023 Publication	<1 %
10	"Computer Vision – ECCV 2018", Springer Science and Business Media LLC, 2018 Publication	<1 %
11	aaltodoc.aalto.fi Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		