Remote Photoplethysmography (rPPG) for Real-Time Heart Rate Monitoring

Yash Pokar
u21cs086@coed.svnit.ac.in

Vishw Vekariya
u21cs104@coed.svnit.ac.in

Jay Vadaliya
u21cs078@coed.svnit.ac.in

Gunjan Savaliya
u21cs107@coed.svnit.ac.in

***Abstract*—Remote Remote Photoplethysmography (rPPG) is a non-contact technology that uses facial videos to monitor physiological parameters such as heart rate, offering an affordable and ac- cessible solution for home healthcare. The system processes facial videos captured through standard webcams, extracting subtle changes in skin color using computer vision techniques like Gaussian pyramids to derive Photoplethysmography (PPG) signals. Heart rate is estimated by analyzing signals from facial regions of interest, incorporating multiple heart rate detection algorithms and advanced noise suppression techniques to improve accuracy and reliability. The performance of the rPPG system is validated by comparing its output with pulse oximeter readings, showing a mean absolute error (MAE) and Pearson correlation coefficient (PCC) within acceptable limits. With these enhancements, the rPPG system demonstrates promising reliability and robustness for real-time physiological monitoring**

***Index Terms*—Remote Photoplethysmography, Heart Rate Detection, Real-time Monitoring, Computer Vision.**

I. Introduction

Remote Photoplethysmography(rPPG) is a promising technology for non-contact monitoring of physiological parameters. It is particularly useful in home health care, offering an efficient way to measure heart rate without the use of traditional contact sensors such as ECG Electrocardiography (ECG) electrodes[1]. rPPG utilizes facial videos to extract changes in skin color, which correspond to changes in blood volume. The technology has applications in health monitoring, fitness, and patient care systems.

A. Motivation

The development of this system is motivated by the need for a cost-effective and convenient solution for heart rate monitoring. Unlike traditional contact-based techniques that require electrodes or specialized sensors, remote photoplethysmography (rPPG) leverages standard webcams or cameras, significantly reducing cost and complexity while enhancing accessibility for remote and home-based applications. This technology is particularly beneficial for monitoring heart rates in infants, where contact-based methods can cause discomfort due to their delicate skin and sensitivity, offering a non-invasive, continuous, and comfortable alternative for caregivers and healthcare professionals. Additionally, rPPG has significant applications in real-time heart rate monitoring during physical activities, such as gym workouts, where it can help identify irregularities in individuals at high risk of cardiovascular events. By integrating rPPG into gym environments or wearable systems, it provides critical opportunities for timely intervention and prevention of potentially fatal incidents

B. Objectives

This project aims to develop a cost-effective, non-invasive rPPG system for accurate heart rate monitoring using standard webcams, eliminating the need for specialized sensors. By leveraging advanced noise suppression and multiple detection algorithms, the system ensures reliable performance even in challenging conditions, such as environmental disturbances, rapid movements, and varying lighting. It is designed to address unique use cases like infant heart rate monitoring, offering a safe and practical alternative to contact-based methods, and real-time monitoring during high-risk scenarios like gym workouts, providing timely alerts to detect abnormalities and potentially save lives.

C. Contributions

This project developed a computer vision-based pipeline to extract Photoplethysmographic (PPG) signals from facial videos using a Gaussian pyramid approach, enabling non-invasive heart rate monitoring. The system supports multiple-person detection, ensuring robust performance in group scenarios. Implemented in Python with open-source libraries, it emphasizes real-time efficiency and integrates advanced noise suppression for reliable measurements under challenging conditions like motion or varying lighting. Scalable and accurate, the solution is suited for diverse applications, including home healthcare, gym monitoring, and infant care.

II. Literature Survey

A. Existing Methods

Video rPPG. Human facial veins' blood volume varies synchronously with the hearts' systolic and diastolic cycles, and these variations are embedded in the light reflected from the exposed skin. Proximal video sensors (RGB [1], Near-infrared (NIR) [2], IR [3]) with sufficient sensitivity and resolution can pick up these variations embedded in the video on various channels of the video from the exposed facial skin's light reflection. HR can be obtained by counting the peaks per minute from the obtained rPPG signals. rPPG. Filtering-based methods like filtering the green channel [1], chrominance signal analysis [4], plane orthogonal to the skin tone [5], matrix decomposition [6] have been applied to estimate PPG from skin videos. Signal Source Separation based methods such as Independent Component Analysis [7], source separation [8] have shown their potential in estimating rPPG. Recent deep learning (DL) methods have found their application in the rPPG system. VitaMon [9] have proposed convolutional neural network (CNN) and fully connected layer-based architecture for rPPG estimation. Model Compression. Compression techniques reduce computations of over-parameterized DNN models by learning a fast and compact model that approximates the function learned by the baseline DNN model [10]. The existing techniques can be broadly categorized into post-training compression (pruning, quantization) and developing a new model. Pruning-based methods identify and remove the insignificant nodes/weights of models [11]. The quantization methods reduce the floatingpoint precision (FP) for each node/weight [12], eventually reducing the inference complexity and computations. The quantization also helps the model to comply with different hardware platforms [13]. Besides post-training optimization, [14] develop a compact model from a large trained model. Further, as shown in [15], these methods can also be carefully combined to achieve even greater compression gain. rPPG on limited resource devices. The limited resource devices rPPG system is an emerging field. Authors of [16] integrated NIR LED sensors with field-programmable gate array (FPGA) to detect motion and HR in real-time. [17] assessed the HR estimation performance of the FacereaderTM by Noldus.

B. Challenges in rPPG

Despite significant progress, rPPG faces challenges such as lighting variations, skin tone differences, facial obstructions, subject movement, and varying camera quality, which affect accuracy and reliability. Advanced techniques like adaptive filtering and deep learning show promise in addressing these issues by enabling dynamic adaptation and robust feature extraction. Real-time edge computing further enhances scalability, making rPPG systems more accessible for remote healthcare, fitness monitoring, and emergency response. Overcoming these hurdles will enable rPPG to deliver consistent, accurate heart rate monitoring across diverse real-world scenarios, solidifying its role as a transformative non-invasive physiological monitoring tool.

III. Proposed Methodology

A. Workflow

The workflow of the proposed rPPG system is divided into several stages:

1. **Capture Frame from Webcam:**
   - Configures video parameters such as frame size, frame rate, and color channels to ensure efficient capture of visual details.
   - Utilizes the MacBook M1's 720p FaceTime HD Camera, supporting a resolution of 1280x720 pixels at 30 fps.
   - Enhanced by the M1 chip's image signal processing, providing optimized color balance and low-light performance for reliable data capture and accurate rPPG.
   - Identifies and tracks faces in the video feed to focus processing on relevant regions like the forehead for accurate heart rate signal extraction.
   - Implements Python libraries such as OpenCV, cvzone, and a dedicated face detection module for robust real-time performance.
   - Ensures reliability under challenging conditions, including varying lighting, movement, and partial occlusions like glasses or hair.
2. **Downsampling Using Gaussian Pyramid:** The **Gaussian Pyramid** is an image processing technique that represents an image at multiple scales by down-sampling and smoothing it. The construction involves the following steps:
   - **Gaussian Smoothing:** A Gaussian filter is applied to the image to reduce high-frequency noise, resulting in a blurred version of the image.
   - **Down-Sampling:** The blurred image is subsampled by removing rows and columns, halving the width and height of the image.
   - **Repetition:** The smoothing and down-sampling steps are repeated iteratively, creating progressively smaller and smoother images at each pyramid level.

   The Gaussian Pyramid can be expressed mathematically as:

   $$G_{i+1}(x, y) = \text{DownSample}\left(\text{GaussianFilter}(G_i(x, y))\right)$$

   Where:
   - $G_i(x, y)$: The image at level i in the pyramid.
   - GaussianFilter: A 2D Gaussian function applied to the image.
   - DownSample: The process of reducing the image dimensions by a factor of 2.

   The Down-Sampling process can be mathematically represented as:

   $$g_l(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m, n) \cdot g_{l-1}(2i + m, 2j + n)$$

   Where:
   - $g_l(i, j)$: The pixel value at coordinates (i, j) in the image at level l (current pyramid level).
   - g_{l-1}: The image at level $l - 1$ (the previous, higher-resolution level in the pyramid).
   - $w(m, n)$: The weights of the Gaussian filter, which define how neighboring pixels contribute to the smoothing operation. These weights are determined by the Gaussian function.
3. **Upsampling and Face Detection:**
   - **Upsampling:** Increases the resolution of an image, typically after downsampling, restoring or enhancing image quality at higher resolutions. In Gaussian pyra-

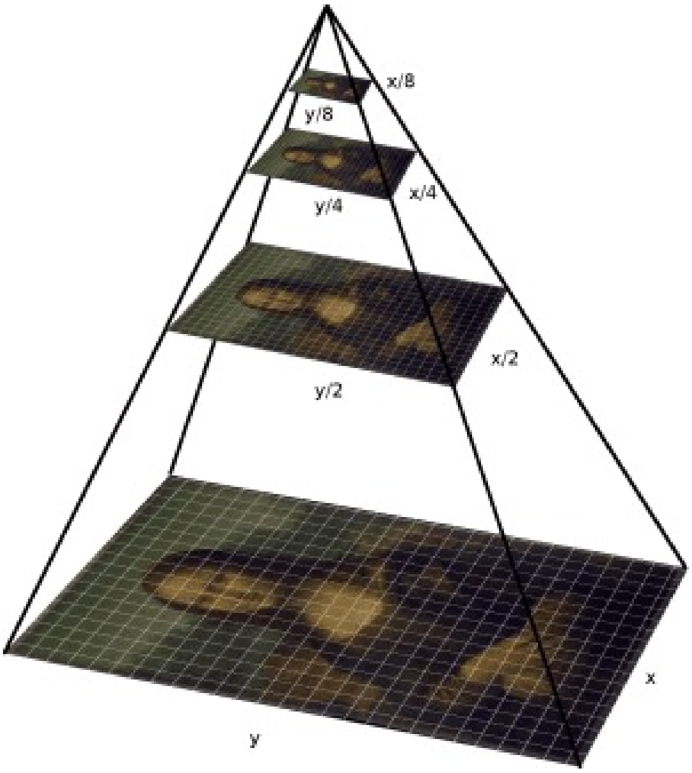

Fig. 1. Structure of Gaussian Pyramid [23]

mids, it complements downsampling by reconstructing higher-resolution images.

- **Relevance in the Code:** While primarily used for downsampling, upsampling can improve the quality of detected face regions by displaying them at higher resolutions while maintaining efficiency in earlier stages.
- **Face Detection:** Handled by the cvzone.FaceDetectionModule, it identifies faces in the webcam feed using the `findFaces()` function, which provides bounding box coordinates and allows face cropping for further processing.
- **Applications in Heart Rate Prediction:** The detected face region contains subtle color variations caused by blood flow, which is crucial for heart rate prediction. Accurate face detection is essential for reliable results.
- **Real-Time Processing:** Face detection is applied in real-time on the webcam feed, dynamically updating as the user moves or changes position.

The upsampling equation is given by:

$$g_{l,n}(i,j) = \sum_{p=-2}^{2} \sum_{q=-2}^{2} w(p,q)\, g_{l-1,n}\left(\frac{i-p}{2}, \frac{j-q}{2}\right)$$

- $g_{l,n}(i,j)$: The pixel value at coordinates (i,j) in the higher-resolution image at level l.
- $g_{l-1,n}$: The lower-resolution image at the previous level $(l-1)$ of the Gaussian Pyramid.
- $w(p,q)$: The weights of the Gaussian filter, determining how neighboring pixels contribute to the value at $g_{l,n}(i,j)$. These weights are derived from a 2D Gaussian kernel.
- $\frac{i-p}{2}, \frac{j-q}{2}$: The pixel coordinates in the lower-resolution image $g_{l-1,n}$. This maps the higher-resolution coordinates (i,j) to their corresponding positions in $g_{l-1,n}$ after upsampling.

4. **Extract Region of Interest (ROI):**

Importance of ROI Selection in rPPG:

- **Skin Tone Variations:** The ROI should capture areas with clear skin tone changes due to blood flow.
- **Noise Reduction:** Selecting an appropriate ROI minimizes interference from irrelevant regions like eyes or mouth.
- **Lighting Consistency:** Regions like the forehead and cheeks have stable lighting, aiding signal accuracy.
- **Minimized Artifacts:** Avoiding motion-prone areas ensures higher-quality signals for heart rate estimation.

Optimal ROIs:

- **Forehead:** Free from facial expression artifacts, flat, and rich in capillaries for clear blood volume changes.
- **Cheeks:** Highly vascularized with significant blood volume changes, offering a larger area for improved signal quality.

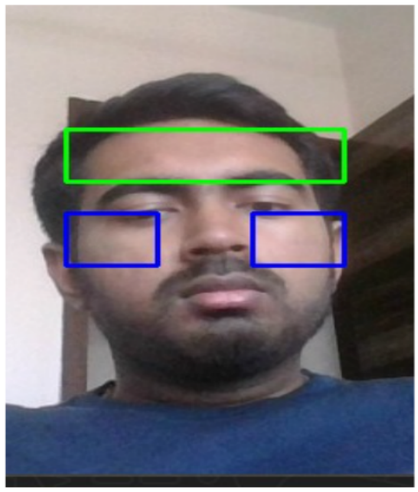

Fig. 2. The selected ROI determination for HR estimation.

Steps for ROI Extraction

a) **Face Detection**: Face detection algorithms (e.g., Haar Cascade, DNN-based, or MediaPipe) locate the face in the frame, producing a bounding box around it.
b) **Landmark Detection**: Algorithms like Dlib or MediaPipe Face Mesh detect facial landmarks to identify regions like the forehead, eyes, and cheeks.
   - **Key Landmarks:**

- **Forehead:** Above the eyes and between the eyebrows.
- **Cheeks:** Below the eyes and lateral to the nose.

c) **ROI Definition**: Based on the landmarks, ROIs are defined for the forehead and cheeks:
   - **Forehead ROI:** A rectangular or trapezoidal region above the eyebrows.
   - **Cheek ROIs:** Rectangular regions below each eye and to the sides of the nose.

d) **Signal Extraction**: The average pixel intensity across RGB channels for each ROI is computed frame by frame, forming temporal signals for further processing.For each channel (R, G, and B), the average intensity is calculated across all pixels in the ROI. This gives a single scalar value for each channel per frame. For every frame in the video, the average intensities of the Red, Green, and Blue channels are calculated and stored in a time-series array. The stored RGB values are plotted against time (frames) to visualize the intensity variation for each channel over time.

5. **Extract RGB Channels:** The RGB channels (Red, Green, Blue) are essential for remote Photoplethysmography (rPPG) in extracting heart rate signals from facial video. These channels capture subtle color changes due to blood flow beneath the skin.

   1. Physiological Basis

   - Blood absorbs light differently based on oxygenation, causing subtle skin color variations.
   - Red and Green channels are sensitive to hemoglobin absorption, making them key for detecting blood volume changes.
   - The Blue channel, while less sensitive, provides complementary data, improving robustness and noise suppression.

   2. Signal Composition

   - Temporal changes in RGB channel intensities correlate with the cardiac cycle, allowing heart rate extraction.

   3. Noise Resilience

   - Analyzing all three channels helps reduce noise from lighting, movement, and skin tone variations, improving heart rate accuracy.

6. **Apply Fourier Transformation:** The Fourier Transform converts time-domain signals into frequency-domain representations, essential for identifying periodic components like heartbeats in heart rate analysis.

   1. Purpose

   - It helps isolate heartbeat frequencies (typically 1-2 Hz) by filtering out noise and irrelevant information.
   - It reveals periodic patterns that are difficult to detect in the time domain.

   2. Steps

   - **Signal Preprocessing**: Detrend and smooth the raw signal to remove noise.
   - **Discrete Fourier Transform (DFT)**: Apply Fast Fourier Transform (FFT) to compute the frequency spectrum.
   - **Interpretation**: Identify peaks in the frequency spectrum corresponding to the heart rate.

   3. Advantages

   - Simplifies signal analysis by decomposing it into sinusoidal components.
   - Facilitates noise filtering by isolating unwanted frequency components.

   4. Limitations

   - Assumes stationarity, which may not apply to biological signals; alternative techniques like STFT or Wavelet Transform can handle non-stationary data.

7. **Apply Bandpass Filter:** Bandpass filtering isolates heart rate-related frequencies (typically 1–2 Hz) and removes irrelevant noise, enhancing the clarity of the heart rate signal.

   1. Purpose

   - Allows only heart rate frequencies (1–2 Hz) to pass, removing low-frequency noise (e.g., respiration) and high-frequency noise (e.g., electrical interference).

   2. Design

   - **Cutoff Frequencies**: 1 Hz (lower) and 2 Hz (upper).
   - **Filter Types**: Digital filters (e.g., Butterworth, Chebyshev), IIR or FIR filters for real-time systems.

   3. Steps

   - The raw signal is filtered, removing unwanted frequencies, and retaining heart rate components.

   4. Advantages

   - Enhances signal-to-noise ratio (SNR) and makes heart rate signal analysis easier.

   5. Challenges

   - Incorrect cutoff selection may distort the signal or exclude valuable data.
   - Proper calibration is necessary to avoid phase distortion or signal loss.

8. **Peak Detection and Heart Rate Calculation :**

   Peak Detection

   - **Local Maxima:** A peak is detected if:

   $$s_i > s_{i-1} \quad \text{and} \quad s_i > s_{i+1}$$

   - **Amplitude Filtering:** A threshold is applied to exclude small peaks caused by noise or motion artifacts.

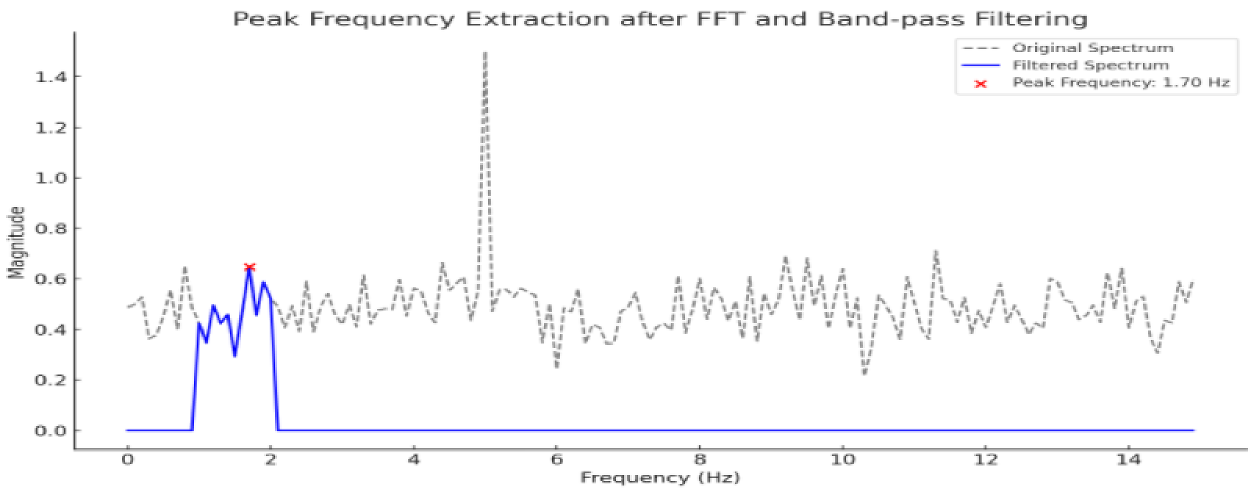

Fig. 3. Peak Frequency Extraction after FFT and Band-pass Filtering.

Heart Rate Calculation

- **Time Intervals Between Peaks:** The time intervals between peaks are:

$$\Delta t_i = t_{i+1} - t_i$$

- **Heart Rate Formula:** The heart rate in beats per minute (BPM) is calculated as:

$$\text{Heart Rate} = \frac{60}{\text{Mean}(\Delta t)}$$

Alternatively, using the dominant frequency f_{peak}:

$$\text{Heart Rate} = f_{\text{peak}} \times 60$$

IV. Simulation and Results

A. Simulation Setup

The system was implemented in Python using OpenCV and validated against pulse oximeter readings.

B. Results

The rPPG system achieved a mean absolute error (MAE) of 4.745 and a Pearson correlation coefficient (PCC) of 0.877. Compared to the deep learning models presented in the literature, the proposed method is characterized by lightweight processing as it relies on the Gaussian Pyramid that is made for image processing.This signifies the superiority of the proposed solution in terms of complexity and performance.

Metric	Value
Mean Absolute Error (MAE)	4.745
Pearson Correlation Coefficient (PCC)	0.877

TABLE I
Performance Metrics

V. Conclusion and Future Work

A. Conclusion

In conclusion, the outlined methodology provides a structured and efficient approach to real-time heart rate detection using facial video frames captured via a standard webcam. By leveraging techniques such as Gaussian pyramids for downsampling, upsampling for spatial restoration, and precise face detection, the process ensures computational efficiency and accurate facial localization. The extraction and analysis of the Region of Interest (ROI), combined with decomposition into RGB channels and signal transformation via Fourier analysis, enable the detection of subtle photoplethysmographic signals. Additionally, applying a bandpass filter ensures noise reduction and frequency isolation, resulting in a clean signal for heart rate calculation. This comprehensive pipeline demonstrates a robust framework for accurate heart rate estimation from facial video data.

B. Future Work

Future work will focus on enhancing the robustness of the system to handle varying lighting conditions and subject movements. Additionally, expanding the system's capabilities to accurately detect heart rate under various lighting conditions, as well as handling moving subjects and detecting heart rates from multiple individuals simultaneously, could further improve its practical application in real-world scenarios.

References

[1] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[2] E. Magdalena Nowara, A. Singla, A. et al., "SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1272–1281.

[3] W. Chen and D. McDuff, “DeepPhys: Video-based physiological measurement using convolutional attention networks,” in *Proc. Eur. Conf. Computer Vision (ECCV)*, 2018, pp. 349–365.

[4] G. De Haan and V. Jeanne, “Robust pulse rate from chrominance-based rPPG,” *IEEE Trans. Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[5] W. Wang et al., “Algorithmic principles of remote PPG,” *IEEE Trans. Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, Jul. 2016.

[6] B.-F. Wu, et al., “Motion resistant image-photoplethysmography based on spectral peak tracking algorithm,” *IEEE Access*, vol. 6, pp. 21621–21634, Apr. 2018.

[7] D. McDuff, S. Gontarek, and R. W. Picard, “Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera,” *IEEE Trans. Biomedical Engineering*, vol. 61, no. 12, pp. 2948–2954, Dec. 2014.

[8] R. Macwan, Y. Benezeth, and A. Mansouri, “Heart rate estimation using remote photoplethysmography with multi-objective optimization,” *Biomed. Signal Process. Control*, vol. 49, pp. 24–33, Jul. 2019.

[9] Z. Hasan, S. R. Ramamurthy, and N. Roy, “CamSense: A camera-based contact-less heart activity monitoring,” *Smart Health*, vol. 23, p. 100240, Aug. 2022.

[10] C. Bucilua, R. Caruana, and A. Niculescu-Mizil, “Model compression,” in *Proc. 12th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, Philadelphia, PA, USA, 2006, pp. 535–541.

[11] S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” in *Proc. Advances Neural Information Process. Systems (NIPS)*, 2015, vol. 28, pp. 1135–1143.

[12] J. Wu, C. Leng, Y. Wang, Q. Hu, and J. Cheng, “Quantized convolutional neural networks for mobile devices,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4820–4828.

[13] Y.-D. Kim, E. Park, S. Yoo, et al., “Compression of deep convolutional neural networks for fast and low power mobile applications,” *arXiv preprint arXiv:1511.06530*, Nov. 2015.

[14] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, Mar. 2015.

[15] Y. Gong et al., “Compressing deep convolutional networks using vector quantization,” *arXiv preprint arXiv:1412.6115*, Dec. 2014.

[16] Y.-C. Lin, N.-K. Chou, G.-Y. Lin, et al., “A real-time contactless pulse rate and motion status monitoring system based on complexion tracking,” *Sensors*, vol. 17, no. 7, p. 1490, Jul. 2017.

[17] S. Benedetto, C. Caldato, D. C. Greenwood, et al., “Remote heart rate monitoring-assessment of the FaceReader rPPG by Noldus,” *PLOS ONE*, vol. 14, no. 11, p. e0225592, Nov. 2019.

[18] Chen, W., Li, M., Song, J., Sun, C., Wang, L. (2020). Contactless heart rate measurement using remote photoplethysmography and machine learning. Journal of Medical Imaging and Health Informatics

[19] Alnaggar, M., Siam, A. I., Handosa, M., Medhat, T., Rashad, M. Z. (2023). Video-based real-time monitoring for heart rate and respiration rate. Expert Systems with Applications

[20] Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., Avolio, A. (2024). Remote photoplethysmography for heart rate measurement: A review. Biomedical Signal Processing and Control

[21] Nadrag, C., Poenaru, V., Suciu, G. (2023). Heart Rate Measurement Using Face Detection in Video.

[22] Uppal, G., Prakash, N. R., Kalra, P. (2017). Heart Rate Measurement Using Facial Videos. Advances in Computational Sciences and Technology

[23] Adelson, Edward H., Charles H. Anderson, James R. Bergen, Peter J. Burt, and Joan M. Ogden. ”Pyramid Methods in Image Processing.”