SkinMap: A Novel Weighted Full-Body Skin Segmentation for Robust Remote Photoplethysmography

Amirhossein Akbari* Zahra Maleki* Amirhossein Binesh Babak Khalaj

*Equal contribution Sharif University of Technology

CVPR Nashville JUNE 11-15, 2025

Paper:

Problem Definition and Contribution

Remote Photoplethysmography (rPPG): It is a non-contact method that measures heart rate by analyzing skin color that rely on variations in light absorption resulting from blood volume changes.

Need for Realistic Datasets: Current rPPG datasets lack the real-world complexities. They also require high-quality cameras and unsynchronized reference signals.

Contributions: We propose SkinMap, a skin segmentation model that generates a weighted skin mask. We also introduce the SYNC-rPPG dataset, which includes data in four real-world scenarios: rest, head rotation, talking, and exercise recovery, using an affordable camera and synchronized sensor.

Heart rate estimation

The signal extraction involves dataset collection, video processing with skin segmentation, RGB to rPPG conversion, and evaluation of the results.

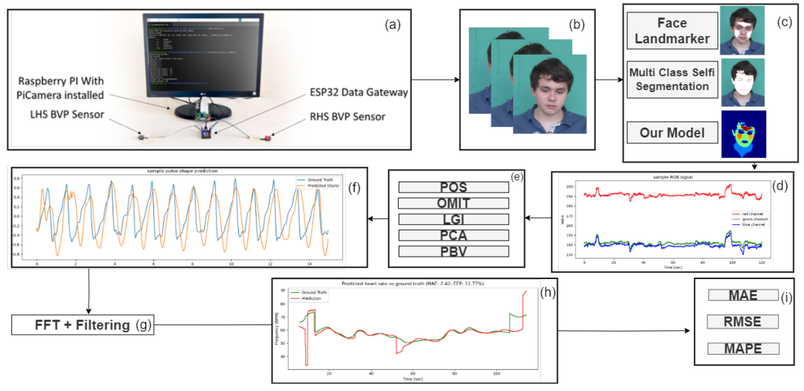

Figure 1. Unsupervised pipeline for heart rate estimation from video.

Methodology

Proposed Architecture: We use DeepLabV3-ResNet50 for skin segmentation, replacing its final layer with a single-channel convolutional layer and a sigmoid activation. Fine-tuned on a large dataset, it segments skin regions and assigns optimal weights based on subject position and lighting for pulse extraction.

Photo Dataset Creation: The model is trained and validated on a custom COCO-derived dataset using MediaPipe to prioritize skin regions: forehead/cheeks (weight=P_i), other facial skin (weight=2), and body skin (weight=1).

$$P_i = \begin{cases} 3 + \left(2\cos\left(\frac{3}{2}\theta_i\right) - 1\right), & \text{if } |\theta_i| < \frac{\pi}{3} \\ 2, & \text{otherwise} \end{cases} \quad (1)$$

Datasets & Experiments

Data acquisition: Using a Raspberry Pi 4B with a 30fps camera and dual MAX30102 pulse sensors for synchronized video and pulse capture ensures stable 30fps operation.

We evaluate our methodology using SYNC-rPPG alongside two publicly available datasets: UBFC-rPPG and UBFC-PHYS. Landmark detection fails in 0.75 frames (talking) and 118 frames (head rotations) per video on average, while SkinMap achieves flawless segmentation.

Figure 2. Model outputs on random samples.

Results

Compared to MediaPipe, SkinMap delivers superior segmentation and signal extraction, as shown in the figures. The final figures highlights SkinMap's robustness in capturing diverse skin tones.

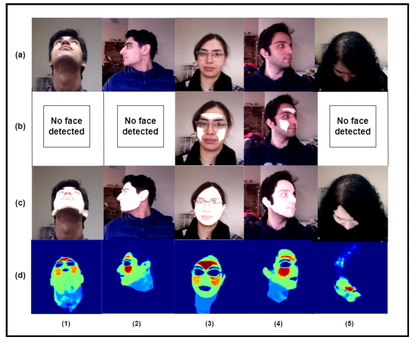

Figure 3. Segmentation results.

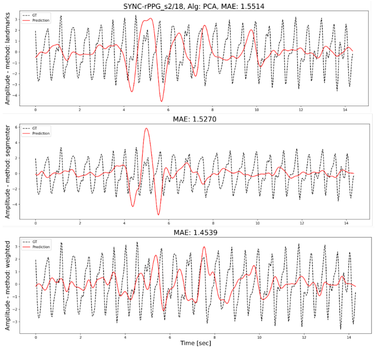

Figure 4. Extracted signal.

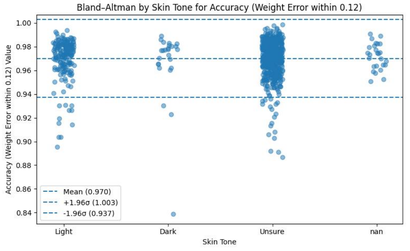

Figure 5. Accuracy

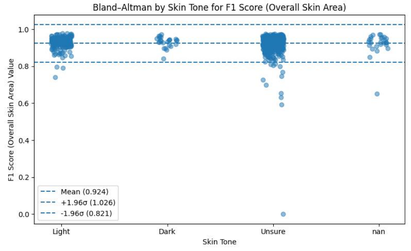

Figure 6. F1 score