# Skin Reflection Angle Useful for Region of Interest Selection in Camera-based Heart Rate Estimation?

Maurice Rohr, Philipp Witulla, Christoph Hoog Antink

KIS*MED - AI Systems in Medicine, Technische Universität Darmstadt, Darmstadt, Germany

## Abstract

*Photoplethysmography Imaging (PPGI) is considered a promising non-contact alternative for existing methods of heart rate (HR) and even blood pressure estimation. A critical processing step in PPGI involves selecting the pixels with the strongest pulse signal from a camera video, which becomes particularly challenging in real-world scenarios involving movement and changing illumination. Current pixel selection methods such as facial landmark-based region selection and skin segmentation overlook scene geometry and illumination conditions. This study evaluates whether the angle between the camera sensor and the skin surface (skin-to-camera angle) could enhance pulse estimation accuracy. The skin-to-camera angle is calculated using 3D-coordinate estimates, and thresholds ranging from 15° to 90° are applied to reject sub-regions based on their angle. The effectiveness of this angle-based selection method is tested on six publicly available datasets containing video data from 288 subjects. Results show that for the CHROM algorithm, angle thresholds of 30° to 50° marginally improve heart rate estimation accuracy by 0.1 BPM, but most algorithms do not benefit significantly from this approach. Consequently, due to the added computational complexity, the angle is unlikely to be a substantial confounder in PPGI signal quality.*

## 1. Introduction

Photoplethysmography Imaging (PPGI) shows great potential for estimating important vital markers such as heart rate, heart rate variability, blood oxygenation, respiratory rate, and even blood pressure [1] using cameras. The technique faces two major challenges: first, how to combine information from different parts of the light spectrum, such as the RGB channels of a camera; and second, how to select the best region to measure the light intensity from.

The methods for selecting this region of interest (ROI) can be grouped by their five fundamental ideas.
1. *Skin-classification* techniques work on the most basic level as they try to find skin tissue with the expectation of simultaneously showing blood perfusion that can be used to estimate the pulse, and removing all pixels such as facial hairs that pollute the PPGI signal. Current models for skin segmentation mostly employ Neural Networks (NN) [2].
2. *Facial landmark*-based approaches usually try to use domain knowledge, by manually selecting an ROI based on studies regarding the visibility, pulsation, and signal-to-noise-ratio (SNR) of a region [3] and combining that with detection and tracking algorithms.
3. *Attention maps* can be implicitly learned by training structurally crafted statistical models (e.g. NNs) on a down-stream task such as blood volume pulse estimation [4].
4. *Pulsatility based selection* is a less explored approach, where the goal is to find pulsatile regions explicitly, for example by employing temporal superpixels and signal quality-based fusion [5].
These methods do not consider scene geometry or illumination.
5. *Scene geometry* was considered a factor only recently by investigating the relevance of the camera-to-skin-angle [6].

In this study, we focus on the skin-to-camera angle-based selection and compare it extensively over a wide range of databases with the less computationally demanding facial landmark-based approach, to answer the question from the title. Therefore, we investigate if the angle between the camera sensor and the skin might be a relevant variable for estimating the pulse using cameras as suggested by [6].

## 2. Methods

We process facial videos with our evaluation pipeline (Figure 1) that extends the rPPG-Toolbox [7] by employing MediaPipe's face extraction model Blazeface [8] to extract facial landmarks. We use 3D-coordinate estimates of the landmarks to estimate the angle between the camera axis and the skin normal, short skin-to-camera angle $\theta$. To quantify improvements from only using skin parallel to the image plane, we use eleven different thresholds $\theta_{\text{th}}$ on the skin-to-camera angle between 15° and 90° to reject additional sub-regions inside the face boundaries. Addi-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.459

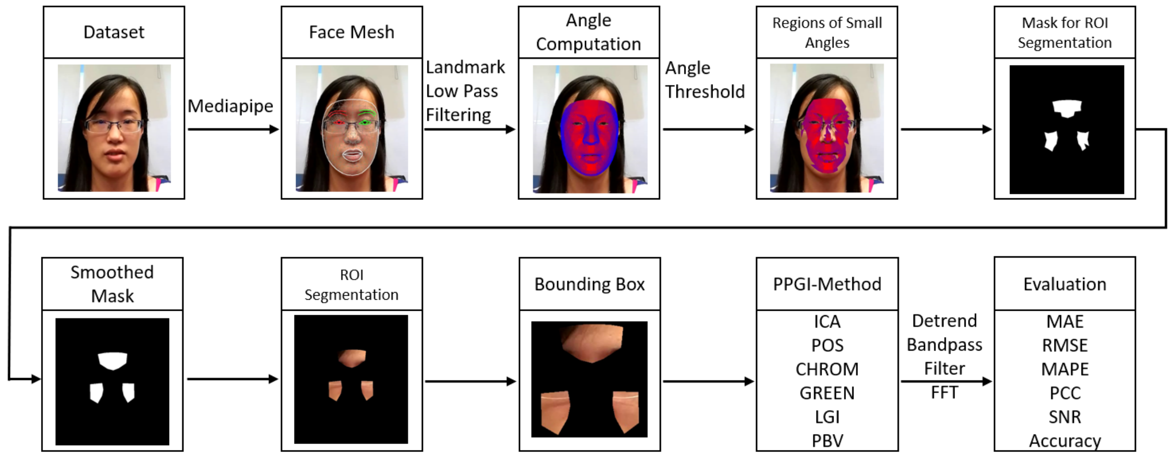


Figure 1. The pipeline uses MediaPipe to extract facial landmarks to compute the skin-to-camera angle and create an ROI segmentation of forehead and cheeks.

Table 1. Public datasets used in this study. MMPD shows a large range of skin colors, makeup and face wear

| Dataset | subjects | scenario | illumination |
|---|---|---|---|
| UBFC-rPPG [9] | 42 | sitting (no movement) | variable + ceiling |
| COHFACE [10] | 40 | sitting (no movement) | artificial / sunlight |
| VIPL-HR-V1 [11] | 107 | everyday activities | artificial |
| PURE [12] | 10 | movement | daylight |
| MMPD [13] | 33 | stationary + movements | artificial / sunlight |
| RLAP-rPPG [14] | 56 | video game | artificial |
| KISMED (private) | 10 | head rotation | ceiling |

tionally, we consider landmark-based skin regions of interest (ROI) in the face that have been shown to have the strongest signal-to-noise ratio (SNR) [3] as well as providing superior heart rate estimation under a broad spectrum of recording situations. The three most promising regions of interest are forehead, left cheek and right cheek [3]. The pipeline allows to turn off each element separately and thereby combine the ROIs with the segmentation provided by the skin-to-camera angle-based thresholding.

## 2.1. Datasets

We collected six openly available datasets listed in Table 1. We also recorded a private dataset (KISMED) with 10 subjects, to analyze the influence of camera type and twelve different scenarios. The subjects gave informed consent. We used two cameras, Logitech 930e with 640x480 pixels and Sigma fp with 1920x1080 pixels, recording uncompressed raw frames with a frame rate of 30 frames per second. The reference PPG was measured synchronously at the finger (Pulox PO-250). In the following, we only analyze the scenario where the subject performs various head rotations.

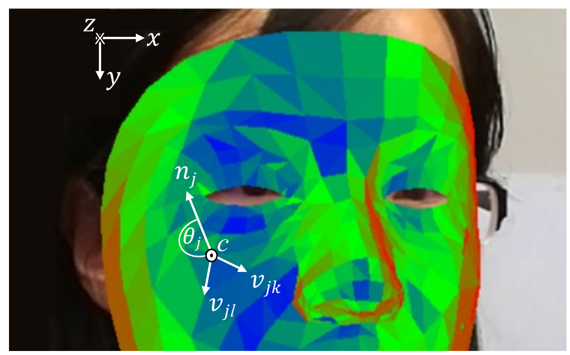


Figure 2. Estimated reflection angle $\theta_j$ between estimated skin normal $\mathbf{n}_j$ and camera axis $\mathbf{c}$. $\mathbf{v}_{j,k}$ and $\mathbf{v}_{j,l}$ are the two vectors connecting the vertices of triangle $j$.

## 2.2. Angle Estimation

The angle estimation uses MediaPipe 3D-landmarks and a tesselation that creates triangles by connecting neighbouring landmarks resulting in 851 triangles all over the face, excluding the eyes. For each triangle $j$ defined by $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathbb{R}^3$ we construct the normal vector $\mathbf{n}_j = (\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_1 - \mathbf{p}_3)$ and normalizing it with $\bar{\mathbf{n}}_j = \mathbf{n}_j / \|\mathbf{n}_j\|$. The skin-to-camera angle is then simply defined as

$$\theta_j = \arccos \frac{\bar{\mathbf{n}}_j \mathbf{c}^T}{\|\bar{\mathbf{n}}_j\| \|\mathbf{c}\|}, \tag{1}$$

where $\mathbf{c} = [0, 0, -1]^T$ is the camera vector pointing perpendicular towards the image sensor. $\theta_j$ is constructed to be between 0° and 90°. We interpolate between the angles of each triangle center by projecting it into the image space and using simple linear interpolation. We shorten it to $\theta$ to mean the angle at any pixel position.

## 2.3. Processing

A threshold $\theta_{th}$ is defined to select pixels where $\theta \leq \theta_{th}$. The resulting set of pixels is then intersected with the set of pixels defined by the optimal facial ROI. This ROI contains the forehead ROI and the left and right cheek ROI and is tracked by using MediaPipe's facial landmarks. Due to jitter over time, the landmark coordinates are filtered using a low-pass filter with 3.5 Hz cutoff frequency. Subsequently, each facial ROI is smoothed using the convex hull algorithm. In this ROI, the color channels are averaged per frame resulting in an RGB signal. We use six methods implemented in the rPPG-Toolbox to extract the PPGI signal from the RGB signals: ICA, POS, CHROM, GREEN, LGI, PBV. The PPGI signal is detrended and band-pass (BP) filtered with cutoff frequencies at 0.4 and 4 Hz before the HR is estimated using the frequency at the maximal spectral power of the FFT inside a 30 s window. The same detrending, BP filtering, and HR estimation is applied to the ground truth signal.

## 2.4. Evaluation

We evaluate the skin-to-camera angle on the six public datasets and the rotation scenario of the KISMED dataset, where we expect the largest influence of angle-based thresholding, using the following metrics: mean absolute error (MAE), root mean square error (RMSE), signal-to-noise ratio (SNR) [15], and accuracy. The accuracy is defined as the number of correctly estimated segments divided by the total number of segments. A segment is considered correct if the HR estimation is within 2 BPM of the ground truth, considering the measurement precision of a typical finger clip PPG.

## 3. Results

Figure 3 shows MAE/SNR over the angle threshold $\theta_{th}$. The average trend of MAE and SNR show a monotonous function with respect to $\theta_{th}$, where SNR is increasing with an increase in $\theta_{th}$. In case of CHROM the a convex function can be observed and an angle threshold from 25° to 60° leads to an improved performance on the UBFC-rPPG dataset. In contrast to separate analysis of single datasets, averaging over all datasets the CHROM estimation improves by 0.1 BPM for $\theta_{th}$ between 45° and 60°. The MAE shows a (small) dip at 30° for the other methods as well. POS shows no improvement from using the skin-to-camera angle to create an ROI. The comparison with the performance of the forehead-cheek ROI in Table 2 shows that POS appears to achieve better results than the best $\theta_{th}$ (in brackets), except for the UBFC-rPPG dataset where the MAE shows a minimum at 25°. POS, ICA, and GREEN (not shown) show no improvement, while PBV (not shown) and LGI improve by 0.2 BPM for 45°. The results for the accuracy metric align with the MAE. In all cases the combination of both approaches leads to worse metrics than the best approach.

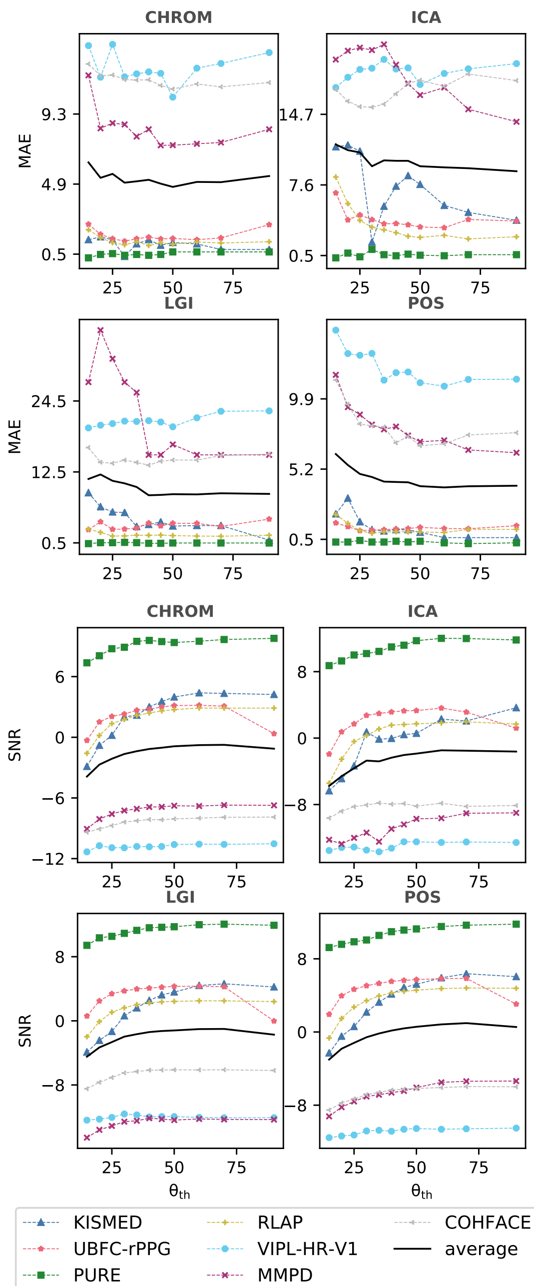


Figure 3. The dependence of HR estimation performance (MAE, SNR) on the angle threshold $\theta_{th}$ for the interpolated skin-to-camera angle. The black line is the average over all datasets.

Table 2. Comparison between using the optimal $\theta_{\text{th}}$ and the landmark-based ROI (L-ROI) using MAE and accuracy (Acc.)

| Dataset | Algorithm | $\theta_{\text{th}}$ | MAE (L-ROI) | MAE ($\theta_{\text{th}}$) | MAE (both) | Acc. (L-ROI) | Acc. ($\theta_{\text{th}}$) | Acc. (both) |
|---|---|---|---|---|---|---|---|---|
| KISMED | CHROM | 30° | 1.05 | 0.35 | 2.72 | 90 | 100 | 80 |
| KISMED | POS | 90° | 0.53 | 0.62 | 0.53 | 95 | 95 | 95 |
| UBFC-rPPG | CHROM | 30° | 1.89 | 1.33 | 1.37 | 74 | 97 | 89 |
| UBFC-rPPG | POS | 25° | 1.42 | 1.06 | 1.18 | 79 | 98 | 87 |

## 4. Discussion

The effectiveness of using the skin-to-camera angle for selecting a region of interest largely depends on the dataset and the PPGI extraction method used. For instance, the combination of UBFC-rPPG and CHROM as reported in [6] show the largest benefit of employing $\theta$. If actually the face-to-skin angle by itself was a good measure to select an ROI, data with heavy head rotation such as the rotation scenario of the KISMED dataset, should show particularly improved performance. However, we observed no relationship between $\theta$ and performance in this case. On the other hand, POS which shows equal or better performance as CHROM, is not affected by $\theta$, but benefits from the larger number of pixels when $\theta_{\text{th}}$ is chosen to be large. A decrease in the total number of pixels results also from the combination of landmark ROI and $\theta$-ROI, leading to a decrease in performance as well. Averaging a larger number of pixels is important for noise reduction, making larger $\theta$ preferable for most algorithms.

## 5. Conclusion

The skin-to-camera angle can be used to select regions for pulse estimation. However, the method is computationally expensive and, based on close inspection of the data where it performed well and the fact that small angles below 15° generally do not lead to improvements, it seems unlikely that the skin-to-camera angle is a relevant confounder to quality in PPGI estimation.

## References


[1] McDuff D. Camera Measurement of Physiological Vital Signs. ACM Computing Surveys 2023;55(9):1–40.

[2] Matthieu Scherpf, Hannes Ernst, Leo Misera, Hagen Malberg, Martin Schmidt. Skin Segmentation for Imaging Photoplethysmography Using a Specialized Deep Learning Approach. In Computing in Cardiology. 2021; .

[3] Lempe G, Zaunseder S, Withgen T, Zipse S, Malberg H. ROI Selection for Remote Photoplethymography. In Bildverarbeitung für die Medizin. 2013; .

[4] Liu X, Fromm J, Patel S, McDuff D. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. In NeurIPS. 2020; .

[5] Benezeth Y, Bobbia S, Nakamura K, Gomez R, Dubois J. Probabilistic signal quality metric for reduced complexity unsupervised remote photoplethysmography. In 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT). Piscataway, NJ: IEEE. ISBN 978-1-7281-2342-4, 2019; 1–5.

[6] Wong KL, Wei Chin J, Chan TT, Odinaev I, Suhartono K, Tianqu K, So RHY. Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2022; 2164–2170.

[7] Liu X, Zhang X, Narayanswamy G, Zhang Y, Wang Y, Patel S, McDuff D. rPPG-Toolbox: Deep Remote PPG Toolbox. arXiv 2022;.

[8] Bazarevsky V, Kartynnik Y, Vakunov A, Raveendran K, Grundmann M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. arXiv 2019;.

[9] Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J. Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography. Pattern Recognition Letters 2019; 124:82–90.

[10] Heusch G, Anjos A, Marcel S. A Reproducible Study on Remote Heart Rate Measurement. arXiv 2017; (1709.00962).

[11] Niu X, Han H, Shan S, Chen X. VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-Constrained Face Video. In Computer Vision – ACCV, volume 11365. 2019; 562–576.

[12] Stricker R, Muller S, Gross HM. Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot. In The 23rd IEEE International Symposium on Robot and Human Interactive Communication. 2014; 1056–1062.

[13] Tang J, Chen K, Wang Y, Shi Y, Patel S, McDuff D, Liu X. MMPD: Multi-Domain Mobile Video Physiology Dataset. arXiv 2023;.

[14] Wang K, Wei Y, Tong M, Gao J, Tian Y, Ma Y, Zhongjin Z. PhysBench: A Benchmark Framework for Remote Physiological Sensing with New Dataset and Baseline. arXiv 2023;.

[15] de Haan G, Jeanne V. Robust Pulse Rate from Chrominance-based rPPG. IEEE Transactions on Biomedical Engineering 2013;60(10):2878–2886.



Address for correspondence:

Maurice Rohr

Merckstraße 25, 64283 Darmstadt, Germany

rohr@kismed.tu-darmstadt.de