International Journal of Multiphysics
Volume 19, No. 1, 2025
ISSN: 1750-9548

# Unsupervised Evaluation and Selection of ROIs for Remote Photoplethysmography

**Mehdi Moghimi[1], Hadi Grailu[2]***

[1]Electrical Engineering Department, Shahrood University of Technology, Shahrood, Iran
[2]Assistant Professor, Electrical Engineering Department, Shahrood University of Technology, Shahrood, Iran

*Corresponding author

**Abstract**

In this paper, we have proposed an unsupervised method for the evaluation and selection of regions of interest (ROIs) in remote photoplethysmography (rPPG). Our approach involves several key steps: (1) face detection and tracking, (2) segmentation of the face into sub-regions designated as ROIs, (3) extraction of pulse signals from each region and computation of property scores based on sliding-window analysis and statistical assessment using signal-to-noise ratio (SNR) and accuracy metrics, and (4) integration of the selected pulse signals to estimate the final pulse signal. We compared our method against traditional techniques such as Green, CHROM, and POS, demonstrating substantial improvements in SNR and accuracy. Our method achieved a minimum SNR of 4.55, approximately 50% higher than the best-performing traditional method (POS, 1.95), with an average SNR of 7.11, outperforming POS and CHROM by 40% and 43%, respectively.In terms of accuracy, our method achieved a minimum of 95.5%, exceeding existing methods by 3.8% to 4.2%.The average accuracy of our method (96.85%) shows a clear improvement over traditional methods (Green, CHROM, POS), enhancing the reliability of heart rate estimation, especially in low SNR environments. The findings underscored our method's potential as a reliable and precise solution for heart rate estimation, especially in low SNR environments, which is critical for remote health monitoring technologies. The results not only highlighted the advancements in rPPG signal extraction compared to traditional methods, but also indicated substantial benefits for applications such as telemedicine, rPPG video compression, etc. where accurate monitoring of vital signs and physiological signal preservation are essential.

**Keywords**: Remote Photoplethysmography, rPPG, Unsupervised, Region of Interest, ROI Selection Measure, Physiological signal preservation, ROI Selection.

## 1. Introduction

Heart rate (HR) serves as a crucial physiological signal for various applications, particularly in the realm of healthcare and medical practices. Electrocardiography (ECG) and Photoplethysmograph (PPG)/Blood Volume Pulse (BVP) represent the conventional methods for monitoring heart activities. Nonetheless, these approaches necessitate physical attachment to the body, constraining their practicality and scalability. The inconvenience associated with long-term monitoring and user discomfort imposes limitations on their application, particularly in contexts like driver status assessment and patient health monitoring. Addressing these challenges, non-contact HR measurement has emerged as a thriving area of research, aiming to remotely monitor heart activity and overcome the constraints of traditional modalities. The primary approach for non-contact HR measurement revolves around remote Photoplethysmography (rPPG) techniques using facial video-based methods. Remote photoplethysmography (rPPG) offers a non-invasive means of measuring blood flow and heart rate without necessitating direct skin contact. Through the utilization of cameras to detect alterations in skin color resultant from changes in blood volume, rPPG surpasses traditional PPG by offering enhanced user comfort and convenience in the measurement process. Emanating from the rPPG framework, various methodologies are employed, with one common approach utilizing image data from video cameras for the analysis of skin color changes. These methods often integrate motion compensation techniques to mitigate the influence of motion artifacts on the signal. The analysis of the signal entails the use of diverse algorithms such as frequency analysis and adaptive filtering to ensure accurate estimation of heart rate. Furthermore, the integration of deep learning models and advanced imaging analysis techniques has facilitated more sophisticated physiological monitoring, including the measurement of blood oxygen saturation and the estimation of blood pressure. As rPPG continues to exhibit promise for non-contact HR measurement, it becomes crucial to explore unsupervised methods for evaluating and selecting regions of interest to enhance the robustness and efficacy of this innovative approach.

Region of interest selection is an essential part for rPPG algorithms. Selecting the proper ROIs is critical to ensure the accuracy, reliability, and efficiency of the algorithm and ultimately impacts the quality of physiological measurements. Effective evaluation and selection of ROIs contribute significantly to improving the performance of rPPG by enhancing signal quality and reducing noise interference. This process is essential for obtaining reliable physiological data in various conditions and environments, making it a valuable area of research for advancing the use of rPPG in real-world applications such as healthcare monitoring and affective computing.

Unsupervised methods for evaluation and selection of ROIs for Remote rPPG are valuable due to their potential applications, such as enhancing rPPG video compression and improving performance in various scenarios. By using unsupervised methods to identify and select ROIs in rPPG videos, it's possible to focus on the most relevant areas for heart rate monitoring and physiological signal extraction. This targeted approach can lead to more efficient video compression techniques, reducing data transmission and storage requirements without compromising the integrity of vital physiological information. Unsupervised methods allow for the automatic identification of ROIs without the need for manual annotation, which is particularly beneficial in scenarios where labeled training data may be limited or unavailable. This capability expands the applicability of rPPG in real-world settings, including situations with diverse skin types, lighting conditions, and facial movements, thus improving the robustness and reliability of rPPG-based systems. Unsupervised methods facilitate the automated selection of ROIs, which can streamline the deployment of rPPG technology in various settings, including healthcare, sports performance monitoring, and human-computer interaction. This can potentially lead to broader adoption and integration of rPPG-based applications in everyday devices and environments with minimal user intervention.

Given the mentioned topics, our aim is to propose a novel unsupervised approach for evaluating, ranking, and selecting ROIs in remote photoplethysmography. We seek to contribute to the field by addressing the challenge of automatically identifying the most informative ROIs without the need for manual annotation or prior knowledge about the specific application. By achieving these objectives, our paper aims to pave the way for more efficient and automated processing of rPPG signals, with broad implications for various domains, including video region-based compression, video quality assessment, healthcare, human-computer interaction, and multimedia technologies.

Remote photoplethysmography (rPPG) technology significantly impacts ROI evaluation strategies by emphasizing the importance of anatomical considerations in selecting regions of interest (ROIs) for accurate blood volume pulse (BVP) estimation. Studies have shown that the thickness of the skin varies across different facial areas, affecting the quality of diffuse reflection information obtained for rPPG [1]. By limiting the ROI to specific facial regions, the signal-to-noise ratio of rPPG signals can be improved, enhancing the accuracy of vital sign measurements [2].

Various studies have highlighted the significance of ROI selection methods based on factors like skin thickness [1], surface orientation [3], and the impact of ROI on signal extraction quality [4]. The angle map representation of the face has been proposed to study the effects of surface orientation on the extracted rPPG signal, showing that regions with small angles of reflection contain stronger signals, often found near the cheeks and forehead [3]. Additionally, the selection of optimal patch ROIs has been shown to effectively eliminate illumination noise and enhance the reliability of heart rate measurements in rPPG technology [4].

Unsupervised skin tissue segmentation for remote photoplethysmography is a crucial aspect of physiological measurements. Existing methods often rely on supervised learning or face detection, which can be limiting. The closest contribution to our work, to the best of our knowledge, is VPS, which leverages physiological features to distinguish human skin from nonhuman surfaces based on pulse signals. This approach achieves significant improvements in skin-region detection precision and pulse rate correlation [5-7]. Additionally, advancements in image preprocessing techniques, such as skin segmentation using color models like YCbCr and HSV, have been shown to enhance rPPG accuracy without compromising real-time processing speed, addressing noise issues encountered in unstable tracking trajectories [8]. Furthermore, the introduction of unsupervised contrast learning approaches, such as ST-Phys, integrates modules for low-light enhancement and spatio-temporal feature utilization, offering superior performance over existing unsupervised rPPG methods and enhancing noise robustness [9].

The rest of the paper is organized as follows. Section 2 outlines the proposed method, including face detection and tracking, ROI pulse signal extraction, and ROI quality measurement. Section 3 details the experimental setup and implementation, including the database and parameter determination. In Section 4, we present the results and discussions, focusing on how the proposed method performs in terms of accurately estimating HR and which key and impactful ROIs are significant for precise and accurate HR estimation. Finally, Section 5 concludes with insights and future research directions.

## 2. Method

The proposed method, as shown in figure 1, comprises four main steps:

**Face detection and tracking:** The algorithm begins with detecting and tracking faces within the video stream, segmenting the tracked face area into 16×16 sub-regions.

**Pulse signal extraction:** A preliminary rPPG signal is then extracted from each ROI.

**ROI quality measurement:** The temporal analysis of the facial video ROIs progresses over time using a time-framing approach, which employs overlapping sliding windows alongside ROI quality assessment.

**rPPG signal reconstruction:** Lastly, the most proper ROIs are utilized to compute the final rPPG signal and estimate HR.

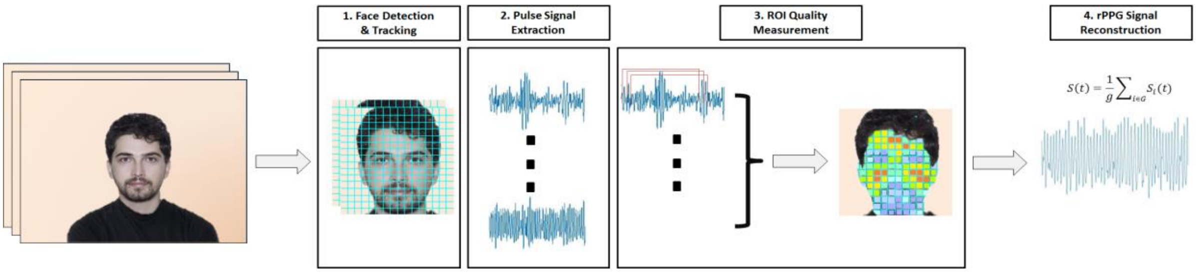



**Figure 1. Overview of the proposed method.**

### 2.1. Face Detection and Tracking

To detect and track the region of a person's face throughout a video, we utilized the Viola-Jones algorithm [10] for face detection and the Kanade-Lucas-Tomasi (KLT) algorithm [11] for face tracking. The Viola-Jones algorithm is used to efficiently detect the presence of a face in each frame of the video, while the KLT algorithm is employed to minimize motion artifacts in the extraction of remote photoplethysmography (rPPG) signals. The KLT algorithm facilitates face tracking by computing the optical flow of facial landmarks across consecutive frames. The algorithm identifies key points or components within a frame and tracks their movement to stabilize the face image. Mathematically, the KLT algorithm estimates the motion matrix, describing the displacement of these key points, which enables the stabilization of the face image. This process involves two primary steps: first, detecting the presence of a face, and second, identifying and tracking its specific facial features.

Viola-Jones object detection algorithm is known for its efficiency in detecting faces in images. The algorithm relies on a cascaded architecture of simple classifiers referred to as Haar-like features. It is based on the concept of Haar-like features and uses a technique called AdaBoost for selecting a small set of critical visual features from a large set, which can then be used to effectively detect objects in an image. The algorithm involves several steps, including: 1) constructing an integral image to efficiently calculate Haar-like features. 2) Selecting a small set of Haar-like features that can effectively classify objects from the background using AdaBoost. 3) Using a cascade of classifiers to progressively filter out negative image regions.

The integral image is defined as equation (1).

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \qquad (1)$$

where ii(x,y) the integral image at pixel location (x,y) and i(x',y') is the original image. Using the integral image to compute the sum of any rectangular area is extremely efficient. The integral image can be used to compute simple Haar-like rectangular features. These features help in capturing certain patterns in an image, such as edges or textures. Viol-Jones utilizes the AdaBoost algorithm to select a small set of important features and creates a strong classifier by using equation (2).

$$F(x) = \sum_{i=1}^{N} \alpha_i . sign(h_i(x)) \qquad (2)$$

where $\alpha_i$ are the weights assigned to weak classifiers, $h_i(x)$ are weak classifiers, and N is the total number of weak classifiers. The Viola-Jones algorithm employs a cascade of classifiers to efficiently reject non-face regions. The cascade structure helps in quickly discarding regions that are unlikely to contain faces, reducing computation time. Finally, to detect a object, the algorithm slides a window of different sizes over the image and applies the cascade classifier at each position. If the region passes all stages of the cascade, it is classified as a target object.

The KLT tracking algorithm, also known as the Kanade-Lucas-Tomasi algorithm, is a popular tracking method introduced in [11]. It is widely used across various tracking tasks. Kanade-Lucas' optical flow technique discussed in [12] was further enhanced by Shi-Tomasi [11], who improved feature selection for better tracking performance. The equation (3) depicts the motion between two consecutive frames in a video.

$$I(x,y,t+\tau)=I(x-\alpha,y-\beta,t) \qquad (3)$$

where I represents the image intensity, t and τ are time and difference of time between two frames respectively, α and β are increments of dimensions in frame of time t. Images have some noises which they are imposed in frames and could be aggregated to equation (3). So, equation (3) can be computed to reduce the noise (n).

$$n = \iint_W [I(x - \alpha, y - \beta, t) - I(x, y, t + \tau)]^2 w(x, y) dxdy \qquad (4)$$

where w(x,y) represents a weighting function, and W is the search window. As reported in papers [11, 13], we set the weight to unity. Since the displacement is small relative to the search window, we can rewrite this equation using a Taylor series approximation based on equations (5-7). Also, we do not introduce time symbols into the formulas because displacements are crucial for this purpose.

$$I(x - \alpha, y - \beta) \approx I(x, y) - \alpha \frac{\partial I}{\partial x}(x, y) - \beta \frac{\partial I}{\partial y}(x, y) \qquad (5)$$

$$g \triangleq [\frac{\partial I}{\partial x} \;\; \frac{\partial I}{\partial y}]^T \qquad (6)$$

$$I(x - \alpha, y - \beta) \approx I(x, y) - g.D, \qquad D \triangleq (\alpha, \beta) \qquad (7)$$

By substituting these approximated series, equation (4) would be rewritten as below.

$$n = \iint_W [I(x,y,t) - g.D - I(x,y,t+\tau)]^2 dxdy \quad (8)$$

$$F \triangleq I(x,y,t) - I(x,y,t+\tau) \quad (9)$$

$$n = \iint_W [F - g.D]^2 dxdy \quad (10)$$

To find the displacement, D, the equation (10) needs to be differentiated with respect to D and then equated to zero.

$$\frac{dn}{dD} = -2 \iint_W [F - g.D]\, gdxdy \quad (11)$$

$$\iint_W [F - g.D]\, gdxdy = 0 \quad (12)$$

$$\iint_W gg^T Ddxdy = \iint_W Fgdxdy \quad (13)$$

$$GD=H \quad (14)$$

$$G=\begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \quad (15)$$

Based on Shi and Tomasi's definition in [11], if λ1 and λ2 represent the eigenvalues of matrix G, the optimal features for tracking should fulfill the condition that the minimum of λ1 and λ2 is greater than the threshold λth. This threshold is derived from uniform intensity regions. This approach allows for the extraction of corners and highly textured areas in the image.

We use the Viola-Jones algorithm followed by the KLT tracking algorithm to locate the central region of the face, including the nose, lips, and eyes, in consecutive frames of the video. We track this region and use a 256*256 dimensioned facial area for decomposition and analysis by utilizing the geometric dimensions and component proportions. The facial region is approximately situated in the middle of this area. This 256 by 256 area is divided into 16*16 sub-regions, and each of these sub-regions will be used as Regions of Interest (ROI) in our further processing. An example result of this process is shown in Figure 2.

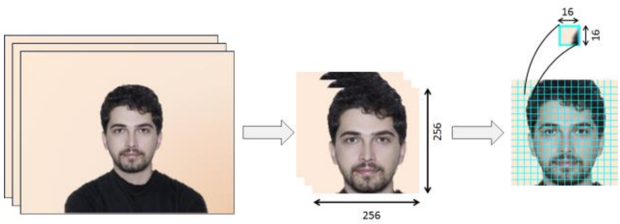



**Figure 2. Face detection and ROI gridding.**

### 2.2. ROIs Pulse Signal Extraction

To distinguish between significant clusters (such as skin areas) and insignificant ones, we generate preliminary rPPG signals for each 16×16 region. In this approach, the pixel values within each region are averaged spatially for every video frame. As a result, we obtain a collection of 256 RGB time series denoted as $x_i^c(t)$, where c represents the color channel {R, G, B}, t signifies the frame index, and i ranges from 1 to K (where K equals 256 - the total number of regions).

$$x_i^c(t) = \frac{\sum_{k=1}^{N_i(t)} I_{k,i}^c(t)}{N_i(t)} \tag{16}$$

where $N_i(t)$ is the number of pixels in the ith region at time t, which is fixed and equal to 256 due to each region's size being 16×16, and $I_{k,i}^c(t)$ represents the kth pixel value at time t in color channel c.

The RGB temporal traces are then pre-processed by normalized, detrended using a smoothness priors approach [14], and then band-pass filtered with a Butterworth filter. Subsequently, the rPPG signal can be extracted using various existing methods. In this study, we opted for the chrominance-based method (CHROM) [15] due to its simplicity and high reliability. CHROM utilizes uncomplicated linear combinations of RGB channels and demonstrates excellent performance with minimal computational complexity. Let $y_i^c(t)$ represent the pre-processed RGB time series. Using the CHROM method, these RGB values are projected onto two orthogonal chrominance vectors $X_i$ and $Y_i$ obtained from equation (17) and (18).

$$X_i(t) = 3y_i^R(t) - 2y_i^G(t) \tag{17}$$

$$Y_i(t) = 1.5y_i^R(t) + y_i^G(t) - 1.5y_i^B(t) \tag{18}$$

Since $X_i$ and $Y_i$ are two orthogonal chrominance signals, changes induced by PPG are expected to differ between $X_i$ and $Y_i$, while motion impacts both chrominance signals in the same way. Therefore, the pulse signal in the ith region can finally be determined by the equation (19).

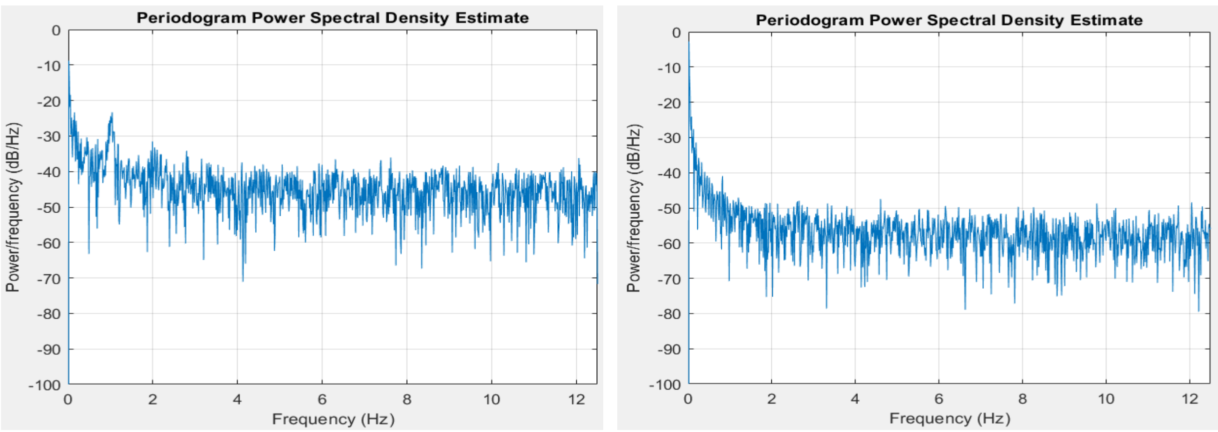
$$S_i(t) = X_i(t) - \alpha Y_i(t) \tag{19}$$

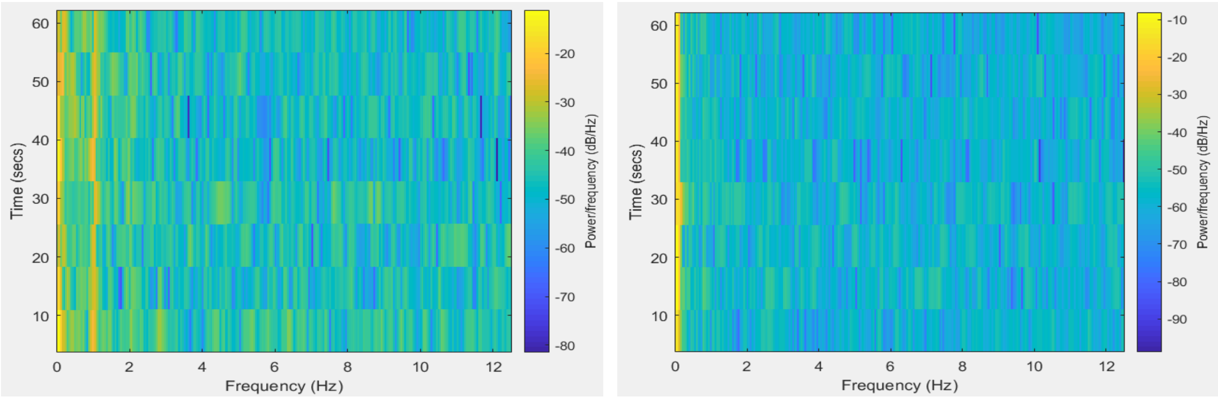
where

$$\alpha_i = \frac{\sigma(X_i)}{\sigma(Y_i)} \tag{20}$$

### 2.3. ROI Quality Measurement

Only the skin tissue of a living subject shows pulsatility, so proper pulse signals must extract from some ROIs, and it's considerable that the desired pulse is consistently present over time. Figure 3(a) presents the periodogram and spectrogram of a pulse signal estimated from the skin ROI, while Figure 3(b) shows the periodogram and spectrogram of a pulse signal estimated from the non-skin ROI. In the frequency domain, the pulsatile, cardiac-synchronous signal exhibits a significant peak centered around the fundamental frequency of the heart rate over time, along with a potential second harmonic, while providing limited information at other frequencies.






**(a)** **(b)**

**Figure 3. Periodogram and spectrogram examples of two tentative rPPG signals estimated from (a) skin area and (b) non-skin area**

Estimating HR from facial videos using rPPG faces challenges like transient movements and ambient light variations, etc. affecting accuracy [16, 17]. Complex facial regions like eyes, lips, or beard areas can introduce errors due to even minor movements [17]. The more proper ROIs are those that are less affected by challenges that lead to a decrease in heart rate estimation accuracy. This allows for the extraction of heart rate over time with appropriate precision and accuracy ensures that the rPPG signal remains of high quality. These ROIs should have characteristics such as: 1) ROIs that experience fewer optical or motion or etc. challenges, resulting in higher estimation accuracy, and 2) ROIs with good blood flow that can produce a stronger and higher quality signal. In other words, if we use a proper ROI for HR estimation at very short intervals, the estimated HR should show minimal variance. Additionally, during all these time intervals, the quality of the rPPG signal should be high.

Based on this fact, we analyze the facial video ROIs over time using a time-framing approach with overlapping sliding windows (one-frame difference) from two perspectives: the quality of the signal and the precision of HR estimation. The window used for framing the pulse signal S will be a rectangular window as shown in figure 4 and equation (21). Therefore, each ROI after framing the corresponding pulse signal extracted from it, $S_i(t)$, will have J=T-L+1 frames (windowed signal, $Sw_i^j(t)$).

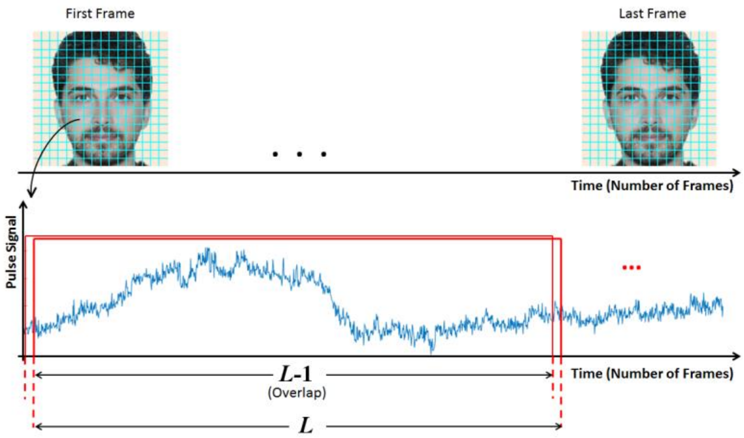



**Figure 4. Rectangular sliding window applied to ROI pulse signal.**

$$W_{framing}^{j}(n) = \begin{cases} 1 & j \le n \le L + j \\ 0 & Otherwise \end{cases} \qquad (21)$$

where L is the length of the window, and j is the parameter that slides the window over time. If we consider the duration of the signal as T, then j can take values from 0 to T-L.

To measure the quality of rPPG signals, we utilize the signal-to-noise ratio (SNR), which is calculated as the ratio of the energy around` the fundamental frequency to the remaining energy. The SNR for the jth time-frame signal of extracted pulse signal of ith ROI is determined using equation (22).

$$SNR_i^j = 10\log_{10}\left(\frac{\int_{f1}^{f2} Wsig_i^j(f)\left|F\{Sw_i^j(t)\}\right|^2 df}{\int_{f1}^{f2}(1 - Wsig_i^j(f))\left|F\{Sw_i^j(t)\}\right|^2 df}\right) \quad (22)$$

where $F\{Sw_i^j(t)\}$ represents the Fourier transform of the rPPG signal for the jth time-frame signal of ith ROI, $f_1$ and $f_2$ denote the lower and upper limits of the integral, corresponding to the physiological range of heart rates (40 to 210 bpm in our case), and $Wsig_i^j$ is a gaussian template window shown in figure 5.

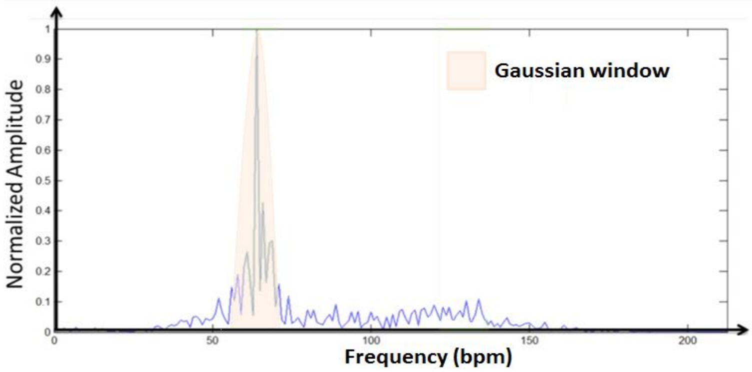



**Figure 5. In the frequency spectrum of the extracted pulse signal, the Gaussian window is defined as ± 3 bpm away from the fundamental frequency.**

As we discussed, in order to consider a ROI proper for the quality of rPPG signal presentation, it must always have a high SNR. Therefore, to evaluate the quality of the rPPG signal in the ROI, we calculate the mean and variance of the SNRs obtained from the time frames created after the time-windowing phase of that ROI (by using equation (23) and (24)). For a proper ROI, we expect a higher mean and a smaller variance in the statistical analysis of the SNRs for that ROI.

$$SNRAvg_i = \frac{1}{J}\sum_{j=1}^{i} SNR_i^j \quad (23)$$

$$SNRVar_i = \frac{1}{J}\sum_{j=1}^{J}(SNR_i^j - SNRAvg_i)^2 \quad (24)$$

To analyze the ROI from the perspective of HR estimation precision, we employ the Chrom method. As we mentioned earlier, we expect a proper ROI for rPPG-based HR estimation to be unaffected by challenging factors such as motion or local lighting changes. Therefore, HR estimates from that ROI should remain relatively consistent over closely spaced time intervals. To achieve this, we calculate the variance of the HRs obtained from the time frames created after the time-windowing phase of that ROI (by using equation (25) and (26)). For a proper ROI, we expect low variance in the statistical analysis of the HRs estimation for that ROI.

$$HRAvg_i = \frac{1}{J}\sum_{j=1}^{i} HR_i^j \quad (25)$$

$$HRVar_i = \frac{1}{J}\sum_{j=1}^{J}(HR_i^j - HRAvg_i)^2 \quad (26)$$

where $HR_i^j$ is the HR obtained from the jth time frame signal of ith ROI.

Then, the property score (PS) of the ith ROI is calculated. Given that our goal is to identify ROIs with low variance in the estimation of HRs and SNRs, and high average SNRs, we set the evaluation criteria according to equation (27).

$$PS_i = \frac{SNRAvg_i}{\gamma \times SNRVar_i + (1-\gamma) \times HRVar_i} \tag{27}$$

where, γ is a regularization parameter assumed to be between 0 and 1, and $SNRAvg$, $SNRVar$ and $HRVar$, have been normalized to a scale between 0 and 1 using unity-based normalization according to equation (28).

$$\acute{e}_i = \frac{e_i - E_{min}}{E_{max} - E_{min}} \tag{28}$$

where, $\acute{e}_i$ is the normalized value of $e_i$ for variable E in the ith row.

The final rPPG signal S(t) is then obtained by averaging the selected tentative pulse signals $S_i(t)$. This is expressed as equation (29).

$$S(t) = \frac{1}{g} \sum\nolimits_{i \in G} S_i(t) \tag{29}$$

where g is the number of elements in the set G. The elements of set G are the ROIs with a property score ($PS$) greater than a specific threshold. This threshold is determined by the average property score of all tentative skin ROIs.

## 3. Experimental Setup and Implementation Details

In this section, we outline the experimental setup for assessing the proposed method. First, we will describe the dataset utilized. Next, we will detail the parameters employed in the method. Finally, we will present the evaluation metrics used to analyze the results of the proposed approach.

### 3.1. Database

We created a self-collected benchmark database to evaluate the performance of our proposed algorithm. The study involved 16 healthy participants (11 male, 5 female, ages 23 to 55) and received approval from the DSP laboratory committee at Shahrood University of Technology. Informed consent was obtained from each participant prior to the experiments. We used a consumer-grade webcam (HD Pro Webcam C920, Logitech) for image capture, along with MATLAB's "Image Acquisition Toolbox." To establish a ground truth reference, we simultaneously recorded contact-based PPG signals using a fingertip pulse oximeter (CMS50E, Contec Medical). The experiments took place in a practical setting, specifically a general office environment without dedicated lighting or a controlled background, as typically found in lab conditions. Illumination levels were recorded during the day with mixed lighting, including sunlight and fluorescent lamps. Participants remained stationary during the recordings, with some minor involuntary movements allowed, positioned approximately 1.5 meters from the camera. The video was captured at 1280 × 720 pixels and 30 frames per second (fps) for a little over a minute, stored in WMV format. Figure 6 shows some snapshots from the benchmark dataset.

**Figure 6. Snapshots of the benchmark dataset**

### 3.2. Parameters Determination

Considering that the initial stages of the proposed method involve face detection and tracking, followed by grid partitioning of the face area, it is necessary to determine the face region size and ROI size. In the next stage of the proposed method, which includes extracting the signal variations in brightness for each ROI and then windowing that signal, it will be important to determine the type of window, the window length, and the degree of overlap between windows. Finally, in order to assess the suitability of each ROI according to Equation (29), we need to determine the regularization parameter in the mentioned equation, which balances the accuracy of heart rate estimation and the signal-to-noise ratio. These parameters will be determined experimentally as described below:

Face region size: 256×256
ROI size: 16×16
Number of ROIs: 256
Framing window ($W_{framing}$): Rectangular window
Framing window length (L): min{0.8×number of total video frames, fps×50}
Framing windows overlap: L-1
SNR calculation window ($W_{sig}$): Gaussian window is defined as ± 3 bpm away from the fundamental frequency
Regularization parameter in proper score formula (γ): 0.5

## 4. Results and Discussion

In this section, we will evaluate the proposed algorithm to address two important questions: 1) How does the proposed method perform in terms of accurately estimating HR? 2) Based on the proposed method, which key and impactful ROIs are significant for precise and accurate HR estimation? Therefore, we will analyze and evaluate the proposed method in two parts: “Evaluation of HR Estimation” and “ROIs Property”, comparing it with several conventional methods.

### 4.1. Evaluation of HR Estimation

To evaluate the proposed method, we compared it against three widely used algorithms in the literature (Green [18], CHROM [15], POS [19]), which show significant performance differences. For this comparison, we utilized the iPhys-toolbox [20] to implement and evaluate the mentioned methods. This toolbox provides MATLAB implementations for various non-contact physiological measurement algorithms, allowing researchers to replicate results on their datasets with standard public versions of the baseline methods, ensuring all parameters are well-defined. It includes implementations of many commonly used baseline methods for imaging photoplethysmography (iPPG) and image ballistocardiography (iBCG).

Given that the plethysmographic signal is unevenly distributed among the RGB channels, the green component is particularly important, overshadowing the other two channels; In the Green method, the G component is directly utilized as the rPPG signal S(t). As mentioned in section 2.2, CHROM applies simple linear combinations of RGB channels, and these combinations (based on equations (18) to (21)) are used as the rPPG signal S(t). Like the CHROM, the Plane Orthogonal to Skin (POS) method uses a linear combination of two orthogonal vectors, $X$ and $Y$ obtained from equation (30) and (31).

$$X(t) = y^G(t) - y^B(t) \tag{30}$$

$$Y(t) = -2y^R(t) + y^G(t) + 1.5y^B(t) \tag{31}$$

where $y^c(t)$ is the pre-processed RGB time series and the final result is given by: $S(t) = X(t) + \alpha Y(t)$ where $\alpha = \sigma(X)/\sigma(Y)$. The vectors $X$ and $Y$ create a plane orthogonal to the skin, effectively reducing specular information caused by variations in motion and illumination on the skin surface. Finally, it is worth mentioning again that in our proposed method, the rPPG signal, S(t), is obtained as explained in section 2.3 and according to equation (29).

Figure 7 shows the statistical results obtained from the accuracy of HR estimation for different subjects using the proposed method in comparison with other examined methods.

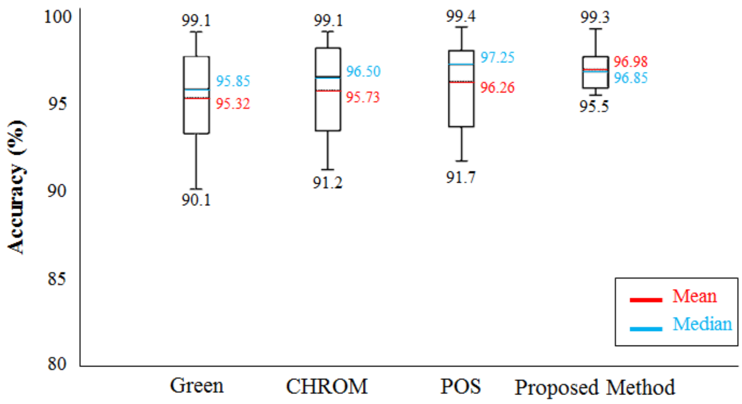



**Figure 7. Comparison of HR estimation accuracy. Box plot analysis across database subjects.**

The results reveal that our proposed method demonstrates a significant enhancement in minimum accuracy, highlighting its robust performance across various subjects, especially those who pose challenges for RPPG estimation. Additionally, the maximum accuracy attained by our method is on par with the best results from the other techniques, underscoring its effectiveness under optimal conditions. Furthermore, our method not only exceeds the average accuracies of the existing methods but does so by a considerable margin, indicating reliable performance across a wide range of subjects. These findings suggest that our proposed method provides a more dependable and precise approach to RPPG estimation compared to traditional techniques. The marked improvements in both minimum and average accuracy imply that our method excels in scenarios where others may falter, thereby broadening its applicability in real-world contexts.

Figure 8 presents the statistical results of the SNR for obtained rPPG signals across various subjects, comparing the proposed method with other evaluated techniques.

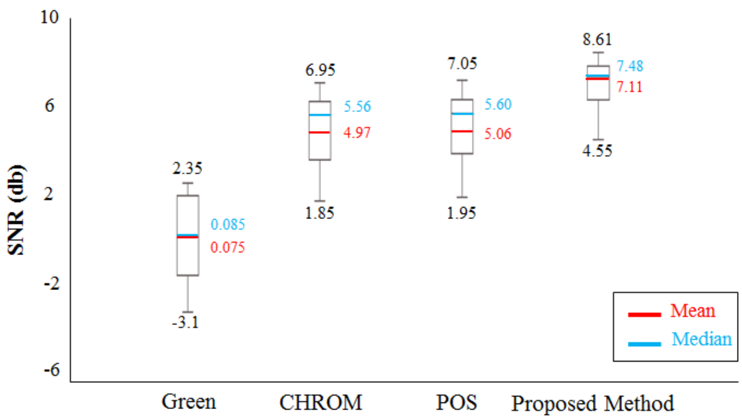



**Figure 8. Comparison of SNR for obtained rPPG signals. Box plot analysis across database subjects.**

The SNR metrics are crucial as they indicate the quality of the extracted signal relative to the background noise. Higher SNR values suggest better signal quality, which is critical for accurate heart rate estimation. The Green method shows a concerning range of SNR values, with a minimum below zero, indicating that in some cases, the noise may dominate the signal. The average SNR close to zero suggests that this method is less effective for reliable heart rate estimation. The CHROM method exhibits a significantly improved performance compared to the Green method. With all SNR values above one, it indicates a generally favorable signal quality. The average SNR suggests a moderate level of reliability for heart rate estimation. Similar to the CHROM method, the POS method demonstrates robust performance with higher minimum and maximum SNR values. The average SNR indicates that this method is also effective for extracting rPPG signals with moderate reliability. Our proposed method outperformed all existing techniques with a remarkable average SNR of 7.11, a minimum SNR of 4.55, and a maximum SNR of 8.61. These results indicate that our approach not only extracts rPPG signals more effectively but also maintains a higher level of signal integrity across subjects, thereby demonstrating its robustness and reliability.

The comparative analysis clearly shows that our proposed method significantly enhances the extraction of rPPG signals, achieving superior SNR metrics compared to traditional methods like Green, CHROM, and POS. The improvements in both minimum and maximum SNR values illustrate the efficacy of our approach in mitigating noise and enhancing signal quality. This advancement is crucial for applications requiring accurate heart rate monitoring and related physiological assessments using rPPG technology.

### 4.2. ROIs Property

Figure 9 shows the ROIs with appropriate property scores for the each subjects in the database after applying the proposed method, which are involved in extracting the final pulse signal S(t) according to Equation 29.

It is noteworthy that the areas around the forehead and cheeks have been involved in suitable participation for the majority of individuals, while regions such as the eyes, eyebrows, lips, nose, and areas covered by hair have shown the least participation in being selected as suitable areas. This observation aligns with the results obtained from our previous work [21] and the research conducted by D.Y. Kim et al. [22]. In our previous study, we concluded that one of the best regions for extracting heart rate based on rPPG is an area in the middle of the forehead, which contains the mid-nasal ridge located in the supratrochlear vein passage. This region marks the beginning of venous drainage from the forehead, originating from a dense venous network that extends widely from the forehead to the scalp. Small veins converge, gradually forming larger venous trunks. These trunks, located in the middle of the forehead, include the supratrochlear vein and the supraorbital vein, while on the sides of the forehead, there is the frontal vein. It is also worth noting that this area of skin has relatively thin skin compared to other facial areas. Furthermore, it was shown in [22] that the thinness of the skin and the proximity of blood vessels to the skin surface have a direct relationship with the suitability of an area for extracting rPPG signals. Based on this, the mid-forehead and cheek areas have been introduced as suitable regions.

**Figure 9. Recommended ROIs for each Subject based on the proposed method**.

## 5. Conclusion

In the present study, we have introduced, implemented, and evaluated an unsupervised method for assessing, ranking, and selecting ROIs for rPPG. Our approach effectively identifies suitable areas of living skin tissue based on distinct pulsatility features. By integrating multiple tentative rPPG signals from the selected ROIs, we have demonstrated that accurate rPPG signals can be estimated remotely without extensive ROI selection.

The results suggest that our method provides a more viable solution for precise heart rate estimation, especially in low signal-to-noise ratio (SNR) environments. This advancement holds significant promise for remote health monitoring technologies. Additionally, our analysis of SNR metrics shows that our proposed method significantly enhances rPPG signal extraction compared to traditional techniques.

Overall, our findings highlight the potential of this method to improve health monitoring systems, leading to more accurate and reliable heart rate estimation techniques. This could be particularly beneficial in applications such as video compression for telemedicine, where maintaining vital signals is crucial. It is important to note that due to the framing of signals from each region and the increased computational load, our proposed method is somewhat heavier in terms of computation and execution time compared to traditional methods. Therefore, optimizing the time efficiency of our approach, along with determining the optimal number and size of ROIs and fine-tuning the parameters for assessing ROI suitability, will be key focuses of our future work.

## 6. Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request. (MehdiMoghimi@chmail.ir)

## Refrences

[1] Daeyeol, Kim., Kwangkee, Lee., Chae-Bong, Sohn. (2021). Assessment of ROI Selection for Facial Video-Based rPPG. Sensors.
doi: 10.3390/S21237923

[2] Georg, Lempe., Sebastian, Zaunseder., Tom, Wirthgen., Stephan, Zipser., Hagen, Malberg. (2013). ROI Selection for Remote Photoplethysmography.
doi: 10.1007/978-3-642-36480-8_19

[3] K., Wong., Jing, Wei, Chin., T., Chan., Ismoil, Odinaev., Kristian, Suhartono., Kang, Tianqu., Richard, H., Y., So. (2022). Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation.
doi: 10.1109/CVPRW56347.2022.00235

[4] GuoPing, Wang. (2021). Influence of ROI Selection for Remote Photoplethysmography with Singular Spectrum Analysis.
doi: 10.1109/AIID51893.2021.9456548

[5] Serge, Bobbia., Yannick, Benezeth., Julien, Dubois. (2016). Remote photoplethysmography based on implicit living skin tissue segmentation.
doi: 10.1109/ICPR.2016.7899660

[6] Serge, Bobbia., Duncan, Luguern., Yannick, Benezeth., Keisuke, Nakamura., Randy, Gomez., Julien, Dubois. (2018). Real-Time Temporal Superpixels for Unsupervised Remote Photoplethysmography.
doi: 10.1109/CVPRW.2018.00182

[7] Wenjin, Wang., Sander, Stuijk., Gerard, De, Haan. (2015). Unsupervised Subject Detection via Remote PPG. IEEE Transactions on Biomedical Engineering,
doi: 10.1109/TBME.2015.2438321

[8] Kunyoung, Lee., Kyungwon, Jin., Youngwon, Kim., Jee, Hang, Lee., Eui, Chul, Lee. (2020). A Comparative Analysis on the Impact of Face Tracker and Skin Segmentation onto Improving the Performance of Real-Time Remote Photoplethysmography.
doi: 10.1007/978-3-030-68452-5_3

[9] Xu, Cheng., Xingyu, Liu., Yan, Jiang., Hao, Yu., Jingang, Shi. (2024). ST-Phys: Unsupervised Spatio-Temporal Contrastive Remote Physiological Measurement.. IEEE Journal of Biomedical and Health Informatics,
doi: 10.1109/jbhi.2024.3400869

[10] [Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I-I. Ieee, 2001.
doi: 10.1109/CVPR.2001.990517

[11] Shi, Jianbo. "Good features to track." In 1994 Proceedings of IEEE conference on computer vision and pattern recognition, pp. 593-600. IEEE, 1994.
doi: 10.1109/CVPR.1994.323794

[12] Lucas, Bruce D., and Takeo Kanade. "An iterative image registration technique with an application to stereo vision." In IJCAI'81: 7th international joint conference on Artificial intelligence, vol. 2, pp. 674-679. 1981.

[13] Mian, Ajmal S. "Realtime visual tracking of aircrafts." In 2008 Digital Image Computing: Techniques and Applications, pp. 351-356. IEEE, 2008.
doi: 10.1109/DICTA.2008.33

[14] Tarvainen, Mika P., Perttu O. Ranta-Aho, and Pasi A. Karjalainen. "An advanced detrending method with application to HRV analysis." IEEE transactions on biomedical engineering 49, no. 2 (2002): 172-175.
doi: 10.1109/10.979357

[15] De Haan, Gerard, and Vincent Jeanne. "Robust pulse rate from chrominance-based rPPG." IEEE transactions on biomedical engineering 60, no. 10 (2013): 2878-2886.
doi: 10.1109/TBME.2013.2266196

[16] V., Bharath., Ruijia, Chen., Hassan, Ali, Aden. (2024). Facial Skin-Spectra for Remote PPg Estimation.
doi: 10.31237/osf.io/qyntd

[17] J., J., Satyanarayana, Penke. (2024). An Efficient Approach to Estimating Heart Rate from Facial Videos with Accurate Region of Interest.
doi: 10.1109/inocon60754.2024.10511840

[18] Verkruysse, Wim, Lars O. Svaasand, and J. Stuart Nelson. "Remote plethysmographic imaging using ambient light." Optics express 16, no. 26 (2008): 21434-21445.
doi: 10.1364/OE.16.021434

[19] Wang, Wenjin, Albertus C. Den Brinker, Sander Stuijk, and Gerard De Haan. "Algorithmic principles of remote PPG." IEEE Transactions on Biomedical Engineering 64, no. 7 (2016): 1479-1491.
doi: 10.1109/TBME.2016.2609282

[20] McDuff, Daniel, and Ethan Blackford. "iphys: An open non-contact imaging-based physiological measurement toolbox." In 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 6521-6524. IEEE, 2019.
doi: 10.1109/EMBC.2019.8857012

[21] Moghimi Mehdi and Grailu Hadi. “General ROI Selection Approach in rPPG for Accurate Heart Rate Estimation”. Journal of Electrical systems. Vol. 20, no. 3, pp: 4445-4456 (2024)
doi:10.52783/jes.5773

[22] Kim, Dae-Yeol, Kwangkee Lee, and Chae-Bong Sohn. "Assessment of ROI selection for facial video-based rPPG." Sensors 21, no. 23 (2021): 7923.
doi: 10.3390/s21237923