

A Motion and Illumination Resistant Non-contact Method using Undercomplete Independent Component Analysis and Levenberg-Marquardt Algorithm

Ankit Gupta, Antonio G. Ravelo-García, and Fernando Morgado Dias *Member, IEEE*

Abstract—Heart Rate (HR) estimation is of utmost importance due to its applicability in diverse fields. Conventional methods for HR estimation require skin contact and are not suitable in certain scenarios such as sensitive skin or prolonged unobtrusive HR monitoring. Therefore remote photoplethysmography (rPPG) methods have become an active area of research. These methods utilize the facial videos acquired using a camera followed by extracting the Blood Volume Pulse (BVP) signal for heart rate calculation. The existing rPPG methods either utilized a single color channel or weighted color differences, which has certain limitations dealing with motion and illumination artifacts. This study considered BVP extraction as an undercomplete problem and proposed a method resistant to motion and illumination variation artifacts. This method is based on an undercomplete independent component analysis, aiming to estimate the unmixing matrix using a non-linear Cumulative Density Function (CDF) that has been optimized using the customized Levenberg-Marquardt algorithm. Therefore, the method is named U-LMA. The proposed method was tested under three scenarios: constrained, motion, and illumination variations scenarios. High Pearson correlation coefficient values and smaller lower-upper statistical limits of Bland-Altman plots justified the outstanding performance of the proposed U-LMA. Furthermore, its comparative analysis with the state-of-the-art methods demonstrated its efficacy and reliability, which was proven by the lowest error and highest correlation values (0.01 significance level). Additionally, higher accuracy satisfying the clinically accepted error differences also justified its clinical relevance.

Index Terms—Blind Source Separation, Blood volume pulse extraction, Dimensionality Reduction, Heart rate estimation, Levenberg-Marquardt Optimization algorithm, Independent Component Analysis.

I. INTRODUCTION

The cardiovascular disease growth rate has been increasing faster worldwide in recent years [1]. Therefore, Heart Rate (HR) is a vital physiological parameter. It reflects the physiological, physical, and emotional state of an individual. HR monitoring has numerous applications in diverse fields such as in criminals' false statements

This manuscript is submitted to the journal on June 23, 2021. This work is supported by the LARSys Project-UIDB/50009/2020, co-financed by Regional Development European Funds for the "Operational Programme Madeira 14–20"—PriorityAxis 1 of the Autonomous Region of Madeira, number M1420-01-0145-FEDER-000002.

Ankit Gupta is working as a researcher at Madeira Interactive Technologies Institute (ITI/LARSys/M-ITI), 9020-105 Funchal and a PhD student of Informatics Engineering, University of Madeira Caminho da Penteada, Funchal 9020-105 Portugal email(ankit.gupta@iti.larsys.pt)

Antonio G Ravelo-García is with Madeira Interactive Technologies Institute (ITI/Larsys/M-ITI), 9020-105 Funchal, Portugal and also associate professor at Institute for Technological Development and Innovation in Communications, Universidat de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain email(antonio.ravelo@ulpgc.es).

Fernando Morgado-Dias is with Madeira Interactive Technologies Institute (ITI/LARSys/M-ITI), 9020-105 Funchal, Portugal and also associate professor at Faculty of Exact Sciences and Engineering, University of Madeira, Penteada Campus, 9020-105 Funchal, Portugal (email:morgado@uma.pt).

detection, and neonatal vital signs monitoring. Photoplethysmography (PPG) and electrocardiogram (ECG) are gold standards for measuring HR. However, both methods follow a contact-based approach causing discomfort to individuals and while also proving to be unsuitable for prolonged and unobtrusive HR monitoring. Therefore, remote photoplethysmography (rPPG) has become an active area of research, which does not require any physical contact with the skin [2]–[11]. rPPG based methods are of utmost importance as they prevent contact in scenarios such as patients with sensitive skin, neonatal vital signs monitoring, or during sleep [10], [12]. The principle behind rPPG is the measurement of periodic variations due to the absorbance of hemoglobin in the blood and pulsation resulting from cardiac activity [13]–[15]. HR estimation using the rPPG method is a three-step process: Region Of Interest (ROI) selection; Blood Volume Pulse (BVP) extraction; and the average HR calculation. Most of the methods used the facial region as ROI, which has been further used for BVP extraction [16]. Capturing the facial region is predominantly done using an RGB camera because it allows less constrained conditions, unlike other methods such as NIR, radar, or ultrasound systems [12]. The rPPG based methods use reflected light acquired through a photodetector after light absorption by the skin tissues, arteries, veins, bones, and blood [17], [18]. This reflected light contains the blood volume variations along with various undesirable noise interferences [12] due to rigid and non-rigid motions [10] and illumination variations [15], which degrade the performance of HR estimation methods due to noisy BVP signal [6]. Furthermore, the noise due to these artifacts easily dominates the relatively weaker strength of the resultant BVP signal [19]. The few frequently used BVP extraction methods used in the literature are Wavelet transforms [20], Independent Component Analysis (ICA) [21], and Ensemble Empirical Mode Decomposition (EEMD) [22]. Wavelet transform requires the selection of appropriate filtering coefficients at different decomposition levels [23], whereas EEMD requires selecting amplitude and noise frequency [24]. However, ICA begins with a random initialization of unmixing matrix with just a single prerequisite of unmixing matrix dimension, depending on the number of independent components, which is comparatively trivial than the other two methods. Additionally, among these methods, ICA is a common method for BVP signal extraction [25]. It considers BVP extraction as a Blind Source Separation (BSS) problem, which deals with extracting the desired signal with no or limited information, from mixture signals. Moreover, Joint diagonalization approximation of matrices (JADE) which is a variant of ICA proposed by Poh et al. [2] has shown motion tolerance up to a certain extent. In addition, the Multichannel ICA proposed by Zhang et al. [26] conducted their experiments under low illumination as well. To the best of our knowledge, none of the ICA method based studies analyzed the impact of motion and illumination variations effect simultaneously, under constrained or natural conditions.

A general assumption regarding ICA based methods is that the number of independent signals is equal to the number of mixed signals. In other words, the signal constructed by each color channel (mixed-signal) results in an Independent Component (IC). This as-

sumption requires analyzing each IC as a potential candidate for the BVP signal while also requiring apriori knowledge. Moreover, there is no defined criterion for selecting the BVP signal from the independent components from different color channels [27]. Conventionally, BVP signal extraction includes selecting the component with the highest periodicity, which may result in selecting the incorrect IC as a BVP signal in the case of periodic motions by the subjects [11]. Most studies selected the second IC for BVP signal extraction by discarding the 1st and 3rd IC corresponding to the red and blue color channels [6], [10], [28]. This results in a loss of information present in the red and blue channels [29], which may be vital for HR estimation. Color subspace transformation methods like CHROM [11] and POS [14] were proposed to overcome this information loss, which employed orthonormal vector transformations to construct a raw signal for BVP extraction. The main drawback of these methods is the weights assigned to color channels which may degrade the BVP information [29].

Considering the limitations mentioned above, the current study proposed the BVP signal extraction as an undercomplete problem. In other words, given three mixture signals corresponding to R, G, and B color channels, the task is to extract one IC which corresponds to the motion and illumination resistant BVP signal. A novel method combining undercomplete ICA [30] with a customized Levenberg-Marquardt algorithm (LMA) [31], [32] was proposed for optimizing the unmixing matrix for BVP signal extraction without losing information from any color channel. Additionally, the proposed method eradicates the need for IC selection since the output is a single IC. The mean HR calculation was performed using power spectral density analysis by using Fast Fourier Transform (FFT), post bandpass filtering. This study contributes to the extant literature with:

- A novel non-linear optimization function constituting a cumulative density function approximated by the hyperbolic tangent to deal with the non-linearity due to rigid and non-rigid motions and illumination variation artifacts for BVP signal extraction.
- A customized LMA for optimizing the entropy of the proposed non-linear least square function, ensuring the statistical independence of the resultant BVP signal.
- A novel method constituting Undercomplete ICA with customized LMA (U-LMA), for an artifacts free BVP signal extraction, followed by its performance analysis under three scenarios (database used): Constrained (VIPL-HR [33]), motion constituting rigid and non-rigid motions (UBFC-rPPG [34]), and illumination variations (COHFACE [35]).
- Testing the performance of U-LMA with negentropy based undercomplete ICA (U-neg) and other state-of-the-art methods to analyze the impact of a non-linear function along with optimization using LMA under the scenarios mentioned above.

II. RELATED WORK

The first attempt of HR estimation under normal light conditions was performed by Verkruysse et al. [4]. The study used the PPG signal extracted using the green color channel of the ROI selected from face videos acquired by a digital camera. The PPG signal was then processed using filtering techniques and, subsequently, HR calculation. Poh et al. [28] extracted the PPG signal using JADE (ICA) from the R, G, and B signal traces acquired from the facial ROI captured using a webcam. Consequently, three ICs were extracted from each color channel, followed by selecting an appropriate IC as a PPG signal for HR estimation. This study was further extended by adding a temporal filtering component, which consisted of detrending and signal smoothing using a moving average filter for better PPG signal extraction [2]. The above methods mainly used

the green component of the ROI since it is considered to have maximum PPG information. The method given by Poh et al. [28] used kurtosis optimization, which does not have descent statistical properties to support statistical independence among components. Gill et al. [36] addressed the problem of unsorted ICs of ICA, which proves to be challenging when selecting the appropriate independent component as a BVP signal. They proposed constrained-ICA, which uses negentropy as an optimization function, avoiding local minima convergence. It is important to note that negentropy possesses better statistical properties and symmetric decorrelation than kurtosis to ensure statistical independence. Considering the periodicity of the PPG signal, Macwan et al. [37] proposed a multi-objective optimization using Autocorrelation and ICA (MAICA), which constitutes negentropy and signal autocorrelation at different time lags, for BVP signal extraction. A Kalman filter was also utilized to address motion and illumination artifacts.

A different approach was presented by De Haan et al. [10], utilizing the chrominance features of R, G, and B spectra. The method extracted the two chrominance vectors, orthogonal to each other, from the RGB color spectra. RGB to chrominance vector transformations was performed by empirically known coefficients. Finally, the ratio for the two vectors was used for PPG signal extraction. Furthermore, De Haan et al. [11] further improved this method by employing the absorption spectra changes of the RGB spectra for BVP signal extraction, where the Hulsbusch noise-free spectrum model was used to develop a normalized BVP vector. Combining chrominance-based signals and ICA's advantages, Song et al. [21] introduced a semi-blind source separation method named Kernel Density ICA, which takes chrominance signals as input and extracts the PPG signal. The kernel density ICA was used to address the problem of similar magnitude among illumination variation and PPG signal. In addition, the authors have also tested the effect of different shooting distances and image resolution for PPG signal extraction.

Realizing the need to add more channels for accurate BVP extraction, McDuff et al. [38] used a five-band lens camera to extract the orange and cyan spectra along with the three traditional color spectra. This enables monitoring the absorption of light differences between Hb and HbO_2 by creating a bigger overlap between cyan, orange, and green spectra, for accurate HR estimation, using the approach presented by Poh et al. [2]. A similar approach was presented by Gupta et al. [9] in which a magenta color filter and a thermal camera was utilized with an RGB camera, to overcome the illumination variations effect on HR estimations. Furthermore, they concluded that the red and green channel with thermal imaging can better estimate HR. FastICA was used for BVP signal separation, which uses negentropy to maximize statistical independence. Yan et al. [5] proposed an approach of using a weighted average of RGB spectra of the selected ROI for improving the Signal-to-Noise Ratio (SNR) followed by denoising the signal using Wavelet transform for PPG signal extraction. Kumar et al. [7] used a monochrome camera for extracting the green spectra followed by its weighted average using varied ROI combinations.

Purucuhu et al. [3] analyzed the effect of using different facial region parts on HR estimation and divided the facial region into three ROIs: forehead, area surrounding eyes and nose, and mouth area. The HR was computed using each ROI, applying ICA for PPG signal extraction, followed by Fourier transform. The summary of this literature review is presented in Table I.

The limitations of the existing state-of-art methods are multi-fold. First, selecting the appropriate IC containing BVP information is challenging due to the unordering of independent components. Second, existing statistical dependence metrics do not consider the non-linearity associated with the PPG extraction problem. Third,

TABLE I
DATABASE SUMMARY USED FOR THIS STUDY.

Study	Parameters	ROI Used	BVP Method	Color channel	Limitations
W Verkrusye [4]	HR, RR	face	Filtering	Green	This study was the first attempt and did not address any artifacts and bad signal to noise ratio.
M Poh [28]	HR	Face	ICA	RGB	The method would not work well under rigid movements and different illumination conditions.
M Poh [2]	HR, RR	face	ICA	RGB	The study did not address illumination variation artifacts and rigid motions.
G Tsouri [36]	PR	Face	ICA	RGB	The constrained ICA is 30 times slower than ICA.
R Macwan [37]	HR	Face	ICA	RGB	The proposed method uses periodicity as one of the criteria for BVP selection, which limits its applicability for estimation during periodic movements.
G Haan [10]	HR	Face	CHROM	RGB	The CHROM method uses skin standardization and fixed projection planes, which halts its generalizability.
G Haan [11]	HR	face	CHROM	RGB	The method did not work well with the stationary subjects due to unavoidable noise since the noise deviates the BVP vector and degrades HR estimations.
R Song [21]	HR	Cheeks	KDICA	RGB	This study examined the influence of resolution and shooting distances; therefore, the limitations were not presented.
D McDuff [38]	HR, BR, and HRV	face	ICA	RGB, cyan, and orange	Using unconventional cameras is not feasible in real-time. Furthermore, rigid movements were not addressed in this work.
O Gupta [9]	HR, HRV	Cheek, Forehead	FastICA	RGB, Magenta thermal	The study did not consider motion artifacts.
Y Yan [5]	HR	forehead	Wavelet Transform	RGB	The method could not work well under illumination variations.
M Kumar [33]	PR, PRV	Face	MRD	RGB, Mono	The method tracks ROI using the KLT algorithm, hence for larger motion; features cannot be tracked, leading to PPG information loss.
T Puruche [18]	HR, RR, and HRV	Forehead, area around eyes, and mouth	ICA	RGB	The study used the constrained conditions for estimations and did not address any artifact.

*Note: HRV- Heart Rate Variability, RR-Respiratory rate, BR- Breathing rate, PR- Pulse rate.

adding further channels for HR estimation enhances the complexity of the problem by increasing its dimensionality with an added effect due to different types of motions and illumination variations. Fourth, color difference equations proposed in color subspace transformation methods have associated coefficients with the color channels, which may affect PPG information. Finally, the semi-blind source separation may need additional information about PPG signal statistical properties for accurate signal extraction.

The problem of unordering of the independent components is resolved by assuming the BVP extraction problem as undercomplete, which deals with taking raw RGB traces as input and a single BVP signal as output. To ensure the consideration of non-linearity with statistical independence, a non-linear optimization function (CDF approximated by \tanh) is proposed. The presented method can deal with the associated non-linearity due to artifacts with three channels of the RGB color space. The proposed method does not assign the weights to color channels, ensuring that each color channel contributes to the BVP signal independently. Finally, U-LMA does not require apriori information to extract the BVP signal from raw signals. Hence the proposed method manages to overcome all the limitations pointed out by the existing literature.

III. PROPOSED METHODOLOGY

The proposed method takes a face video recorded under ambient light conditions as input and estimates the mean HR. It calculates the HR via a three-step procedure: ROI selection, BVP signal extraction, and HR Estimation. This section discusses the detailed information of the proposed U-LMA for HR estimation, explaining all the constituting steps in the following subsections. A detailed flow diagram for the proposed approach is shown in Fig. 1.

A. ROI Selection and Signal Construction

ROI selection deals with identifying the face using the Viola-Jones face detector [39], followed by skin segmentation. The skin was segmented using Cb and Cr components of the YCbCr color model using the parameters proposed by Mahmoud [40]. Subsequently, a spatial averaging operation for each channel was performed on each image frame of the video. A detrending process was also applied to remove slow non-stationary drifts in the signal using the approach by Tarvainen et al. [41]. Finally, an overlapping moving window operation was applied to each channel for constructing raw signals. Fig. 2 depicts the workflow for ROI selection and raw signal construction.

B. BVP signal extraction

The raw signals are further refined to extract the BVP signal for HR estimation. Following the standard ICA annotations, the raw signals are considered as mixture signals containing BVP and other information along with noise interferences due to motion and illumination artifacts. The goal is to extract the BVP signal as one of the ICs from them [42]. Ideally, the 2nd IC from the green channel is selected as a BVP signal, while other ICs are discarded, which may contain BVP information. Therefore, this study defines a BVP extraction as an undercomplete problem that takes three mixture signals and extracts a single IC, consisting of BVP information from all three channels [43]. This problem is solved using the proposed U-LMA, which uses a CDF of the raw signals approximated by \tanh , followed by its optimization using the customized LMA proposed in this study. The proposed approach is motivated by the work of Porcill et al. [30] for signal separation and dimensionality reduction. The difference lies in the context, optimization algorithm, and termination condition. The present work uses a customized version of LMA for optimizing the unmixing matrix W and the number of iterations as the

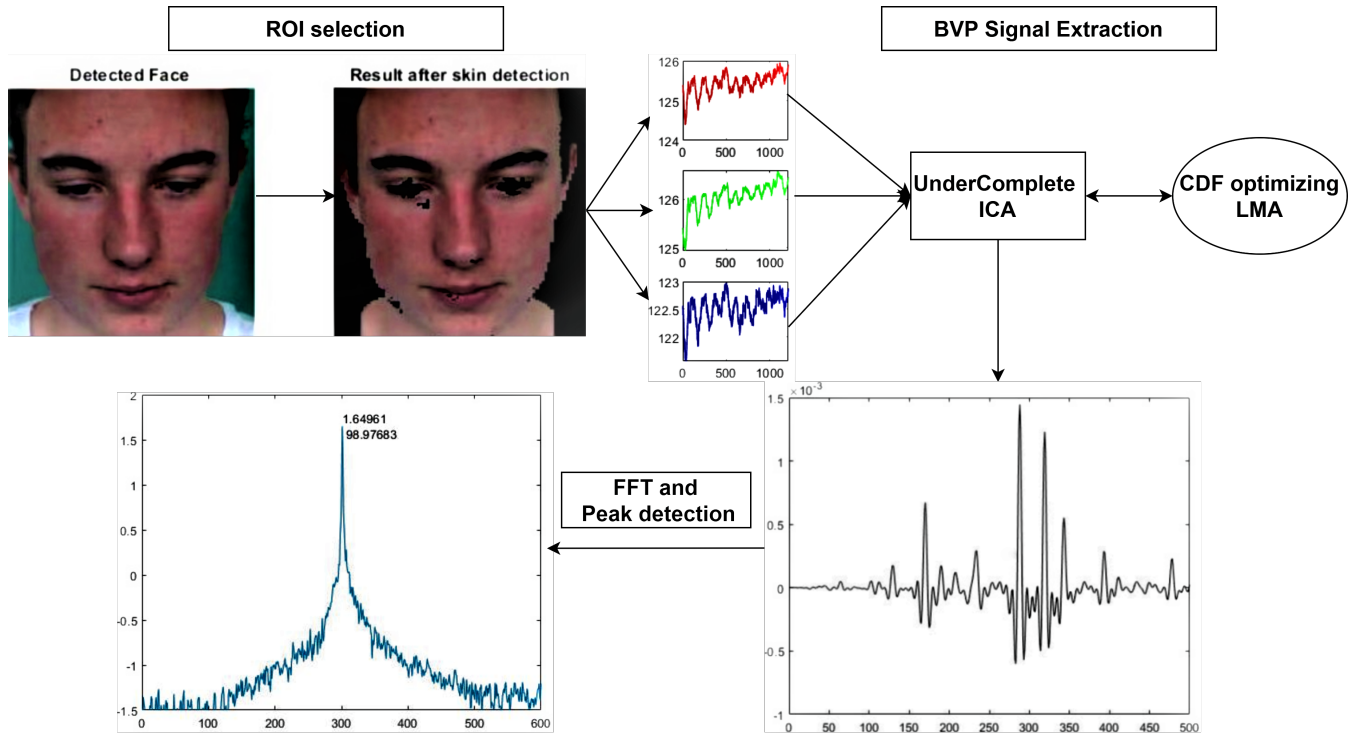


Fig. 1. The workflow of the proposed method for HR Estimation.

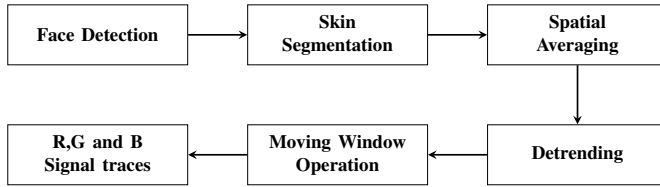


Fig. 2. ROI and raw signal construction.

only termination condition. The reason behind choosing the number of iterations as a termination condition is to consider the absence of BVP signal information. Since there is no reference BVP signal, it becomes impossible to check the correlation of the resultant signal with the original BVP signal, so the correlation criterion is not taken into consideration. LMA is chosen because initial values of W due to random initialization may or may not lie near the desired solution. Both conditions need separate ways of approaching the solution. This optimization algorithm will allow the solution to efficiently converge to the desired values of W in both conditions [44]. The details of the customized version of LMA will be discussed in the following subsection.

Mathematically, $x(t) \in R^{3 \times t}$ and $y(t) \in R^{3 \times t}$ are mixed signals and IC matrix, respectively. $x(t)$ consists of three mixed signals $x_1(t)$, $x_2(t)$, and $x_3(t)$, corresponding to the color channels, whereas $y(t)$ comprises 3 ICs $y_1(t)$, $y_2(t)$ and $y_3(t)$ corresponding to three mixed signals. A standard ICA model assumes that mixed signals are linear combinations of ICs:

$$x(t) = Ay(t) \quad (1)$$

where A is the mixing matrix which, when multiplied by independent components (signals), leads to mixed signals $x(t)$. Unfortunately, mixing matrix and independent components are unknown; therefore, the independent components can only be extracted based on their statistical properties, as mentioned before. Furthermore, the goal is

to estimate unmixing matrix W , which will be used for IC extraction as follows:

$$y(t) = Wx(t) \quad (2)$$

From (1), the matrix W can only be an approximation of A^{-1} for accurate IC extraction. Unlike the standard ICA, where W is a square matrix, for U-LMA, W will be a rectangular matrix since the number of independent components is less than the number of mixed signals. As the CDF (σ) of the statistically independent signals has maximum entropy [30], W can be determined by maximizing the entropy of the CDF, ensuring the statistical independence of ICs [45]. The entropy $H(y)$ of CDF for the BVP signal y is mathematically defined as:

$$H(y) = H(x) + E[\log \sigma'] \quad (3)$$

where $H(y)$ and $H(x)$ defines the CDF's entropy of the IC and multidimensional Gaussian mixed signals, respectively. σ' represents the CDF's derivative of the only statistically independent component (BVP signal). It is important to note that in (3), $x(t)$ and $y(t)$ are written as x and y for brevity. It is challenging to calculate $H(x)$; therefore, it can be approximated as:

$$H(x) = 0.5 * \log(c) + 0.5 * (1 + \log(2 * \pi)) \quad (4)$$

For $H(y)$ maximization, $H(x)$ can be reduced to $0.5 * \log(c)$ where c is written as $E[xx^T] = WSW^T$ and S is a diagonal matrix containing the covariance values of x . Considering the reduced form of $H(x)$ and approximating $\sigma = \tanh$, a new function can be deduced as:

$$h(W) = 0.5 * \log|WSW^T| + E[\log(\sec^2(y))] \quad (5)$$

(5) is used as a criterion for extracting the IC from the mixed signals. Differentiating (5) with respect to W_{ij} yields the ∇ vector for updating the unmixing matrix W , given by:

$$\nabla(w^h) = (SW^T/WSW^T)^T - 2E[y^T x] \quad (6)$$

where $(SW^T/WSW)^T$ is the pseudo inverse of W with respect to S , positive definite matrix.

The proposed algorithm approximates the unmixing matrix W using the customised LMA algorithm by maximizing (5) and updating the matrix W using (6).

C. Customised Levenberg Marquardt Algorithm (LMA)

LMA is a widely used optimization algorithm that is used to find the global minima for non-linear least-square functions with faster convergence property [46] and dual algorithmic adaptability depending on the current solution. In other words, the LMA can be considered as a combination of gradient descent and the gauss-newton method depending on the proximity of the current solution to the global minima. The present study customized the conventional LMA by introducing the entropy of cumulative density function (CDF) approximated by a hyperbolic tangent (\tanh) defined in (5) as an optimization function, followed by its maximization for statistical independence. The advantage of approximation using \tanh lies in the fact that it introduces processing with higher-order statistics to deal with the non-linearity associated with the optimization problem [47].

The workflow of updating W using the proposed method for BVP signal extraction is presented in Fig. 3. The process starts with the random initialization of W followed by calculating the entropy of CDF and subsequently validating the convergence condition. If the convergence condition is reached, then W_{curr} , the W at the current iteration is returned as an output; otherwise, the Jacobian, Hessian, and a diagonal matrix are calculated for updating W . The diagonal matrix consists of the highest value of the Jacobian achieved until the last iteration performed. The cost function, i.e., entropy is calculated before and after updating W and is then compared. The damping parameter λ is increased if entropy decreases, followed by calculating the cost function again after updating W using W_{prev} from the previous iteration until the entropy is increased. If there is a rise in the entropy value after updating W , λ is decreased until the convergence condition is reached. The raw signal is multiplied with W to extract the BVP signal for HR estimation as:

$$BVP(t) = W * x(t) \quad (7)$$

D. Heart Rate Estimation

The BVP signal extracted through U-LMA is processed using a bandpass filter with cut-off frequencies 0.7-4.0 Hz, respectively, corresponding to 42-240 beats per minute (bpm). Finally, the Fast Fourier Transform (FFT) is applied for analyzing the power spectral density for maximum peak estimation, which is then used for the HR calculation by taking its \log_{10} and multiplying it by 60.

IV. RESULT ANALYSIS

The proposed method was tested under constrained and natural conditions using three benchmark databases: VIPL-HR, UBFC-rPPG, and COHFACE. The VIPL-HR database was used for performance validation under constrained conditions, while the UBFC-rPPG tested the method's performance for rigid and non-rigid motions, and illumination variations effect on the proposed method was tested using the COHFACE database. It is necessary to analyze the method's performance under constrained and unconstrained conditions since testing a method under constrained conditions provides insight into its steps and their precision, whereas unconstrained conditions test its robustness. A detailed description of the databases used for this study is presented in the following subsections, with a summary in Table II.

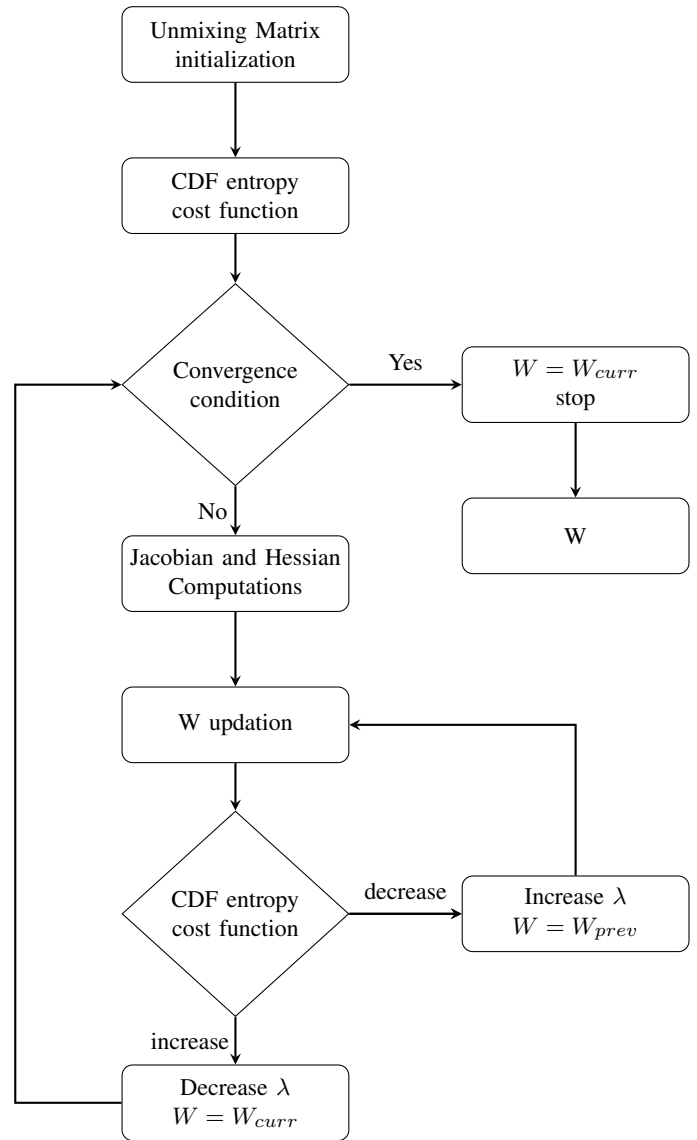


Fig. 3. Customized LMA for entropy maximization.

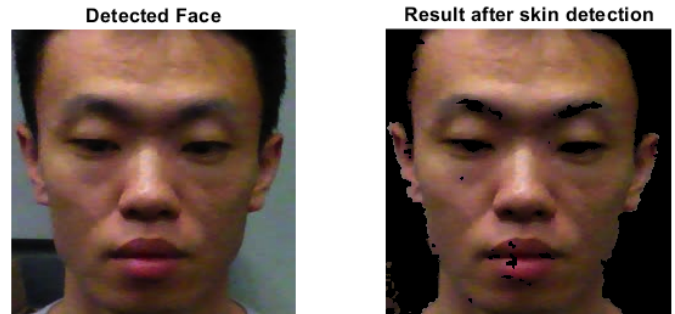


Fig. 4. Face detection and skin segmentation.

A. Databases

1) **VIPL-HR**: The database consists of 2378 videos with visible light spectra and 752 videos with Near Infrared (NIR) spectra from 107 subjects (79 males and 28 females), with age between 22 and 41 years. Nine variable scenarios were considered for sample collection. For each scenario, the samples were collected using digital cameras of different frame rates and NIR cameras. Each database sample

TABLE II
DATABASE SUMMARY USED FOR THIS STUDY.

Database Features	VIPL-HR	UBFC-rPPG	COHFACE
No. of subjects	107	50	40
Video Resolution	1920 × 1080	640 × 480	640 × 480
Frame rate	30	30	20
Video Duration	30 seconds	90 seconds	60 seconds
Ground truth sensor	CMS60C	CMS50E	SA9308, SA9311M
Shooting Distance	1 meter	1 meter	–
HR range	47-100 bpm	63-112 bpm	54-97 bpm
Considered Artifacts	Constrained*	Rigid and Non-rigid motion	Illumination

*Constrained: Conditions with minimum permissible motion and illumination.

comprises a 30 second (s) subject video, a BVP signal, HR, and SpO₂ values [33]. This study has used the subset of videos that corresponds to a frame rate of 30 frames per second (fps) with 1920 × 1080 pixels resolution, covering the HR range between 47 and 100 beats per minute (bpm). The ground truth HR was acquired using a CMS60C pulse oximeter synchronized with the subject's video. This resulted in 107 samples (one from each individual), which were analyzed for testing U-LMA. One sample (p41) was excluded due to insufficient video length (18 s) being smaller than the processing window (25s).

2) UBFC-rPPG: UBFC-rPPG is a publicly available database consisting of 50 video samples, synchronized with a CMS50E pulse oximeter (with a sampling rate of 60 Hz). The videos are available in an uncompressed form with a resolution of 640 × 480 pixels at 30 fps, covering the HR range between 63 and 112 bpm. Each video is 2 minutes (min) long in which the participants were asked to sit facing the camera and play a mathematical game that causes an abrupt rise and fall in HR value promoting rigid and non-rigid movements [21]. The database did not provide age-specific information. All videos were used to test the performance of the proposed method [34].

3) COHFACE Database: The COHFACE dataset is a collection of 160 videos with physiological recordings for the HR and the respiration rate from 40 healthy subjects with a mean age of 35.6 years. The dataset constitutes 60 seconds (s) videos from 12 females and 28 males covering the HR range between 54 to 97 bpm. The videos were recorded with a resolution of 640 × 480 pixels at 20 fps with the synchronized BVP measurements using the BVP model SA9308M, with a belt model SA9311M (with a sampling rate of 256 Hz) [35]. The dataset offers constrained and challenging natural conditions, especially in terms of illumination variations over the facial region. Therefore, this study tests the performance of the proposed method using natural conditions video samples.

B. Performance Metrics

For each video sample, the estimated HR was compared to the corresponding HR value acquired by the ground truth sensor (ECG or PPG). The degree of differences between the actual and the estimated readings were analyzed and summarized using the Bland-Altman plot [48] with upper and lower statistical limits of $\pm 1.96 * \text{standard deviation}(SD)$, mean error, standard deviation, Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), Pearson correlation at 0.01 significance level [49], regression analysis at 95% confidence intervals and accuracy defined by clinically accepted error differences between the ground truth and the estimated HR values, i.e., $\pm 5 \text{ bpm}$.

C. ROI selection and Signal construction

Before this step, the RGB image frames of the video were preprocessed to adjust the pixel intensities, using gamma correction.

The ROI selection deals with face detection followed by segmenting the skin in the YCbCr color space in which Y represents the luminance with pixel intensity ranges between 16 and 235, while for the chrominance blue (Cb) and chrominance red (Cr) components, the pixel values lie between 16 and 240. The thresholds used for Cb and Cr components are in the range of 77 to 127 and 133 to 173, respectively, with no thresholding for the luminance component [40]. Finally, the ROI is selected as 70% height and 60% width of the segmented skin region. Fig. 4 depicts the results of the face detection and skin segmentation process. For detrending the temporal RGB traces, the regularization parameter was set to an empirically defined value, i.e., 10. The raw signal was constructed using a moving window operation with a 96% overlap (1-sec increment) for each color channel.

D. BVP Signal Extraction and HR estimation

The BVP signal extraction was performed using U-LMA. The unmixing matrix W was first initialized randomly, and the values of damping parameter λ were set empirically as 5 and 2.5, respectively, as a part of standard LMA initialization. Subsequently, the customized LMA was employed to maximize the entropy of the proposed non-linear CDF optimization function using 1000 iterative steps due to the fact that none of the video samples took this many iterations for convergence to global maxima. Finally, the optimized unmixing matrix W was used to extract the BVP signal. Finally, an FFT was applied to the resultant signal, followed by the calculation of the \log_{10} value of peak maxima and then multiplying it with 60 for a mean HR estimation.

E. Performance Analysis

As mentioned before, the performance of the proposed U-LMA method is analyzed considering three scenarios: constrained, rigid and non-rigid motions, and illumination variations. Table I specifies that the VIPL-HR database has been used for performance testing under the constrained or stable scenario, UBFC-rPPG for testing its robustness in rigid and non-rigid motions and, COHFACE in illumination variations scenarios. For each scenario, Bland-Altman and regression plots will be presented and analyzed, taking into consideration the respective measured parameters for the plots.

1) Constrained Scenario: For the constrained (VIPL-HR database) scenario, the subjects were asked to sit in the still position at a distance of one meter away from the camera with the ceiling lamp on. The Bland-Altman and regression plot for the constrained scenario are shown in Fig 5 and Fig. 6, respectively. The mean bias for the proposed method is -0.35 bpm which is near to a zero error difference between the ground truth and the estimated values. In other words, on average, the HR estimated by the algorithm measures 0.35 bpm less than the conventional BVP sensor used. Furthermore, the majority of the differences lie within the upper (7.3651) and lower (-8.0632) level statistical limits, which justifies the good performance of the method. Additionally, the Pearson correlation value denoted by r for this scenario is 0.92, thus confirms a higher correlation between the ground truth and the estimated values. Therefore, the Bland-Altman's analysis and high correlation value justify the superior performance of the proposed method under the constrained scenario.

2) Rigid and Non-rigid motions Scenario: A performance analysis under this scenario was performed using all video samples of the UBFC-rPPG database. The videos were collected while subjects were playing a time-sensitive mathematical game which causes an abrupt increase or decrease in HR values along with involuntary head movements due to the subject's action. The samples have a certain amount of illumination variations, too, since the video samples were

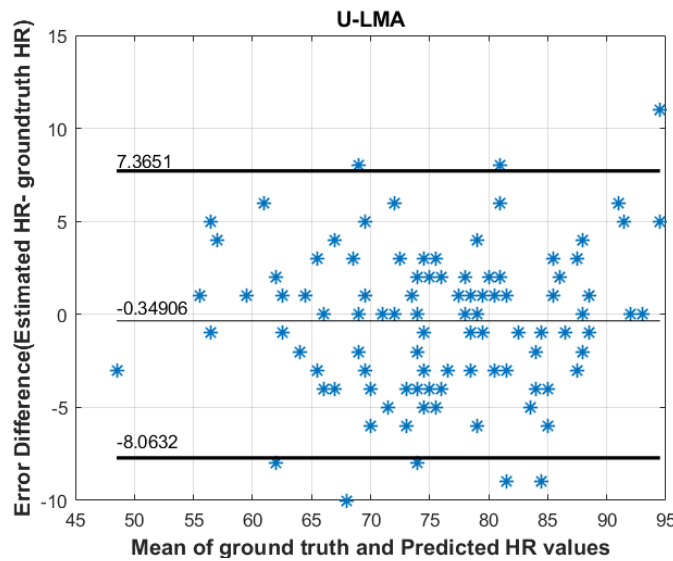


Fig. 5. Bland-Altman Plot for the constrained scenario.

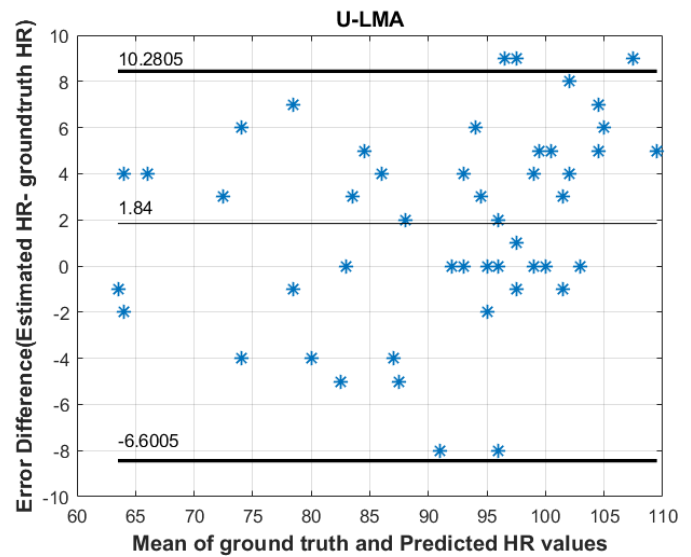


Fig. 7. Bland-Altman Plot for Rigid and Non-rigid motion scenario.

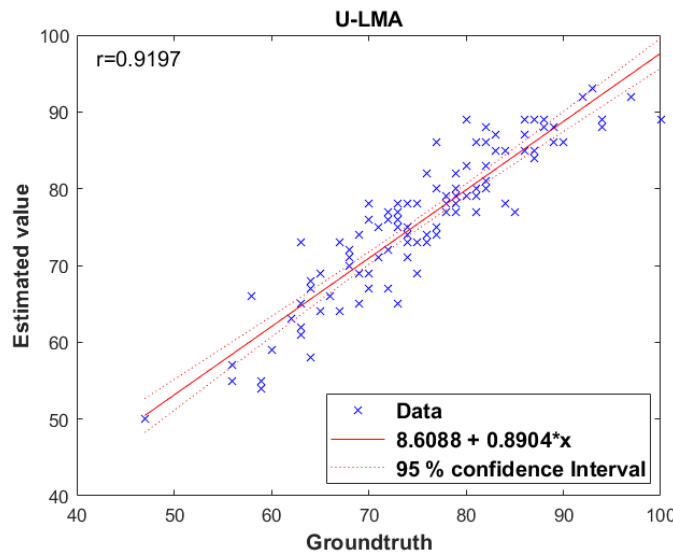


Fig. 6. Regression plot for the constrained scenario.

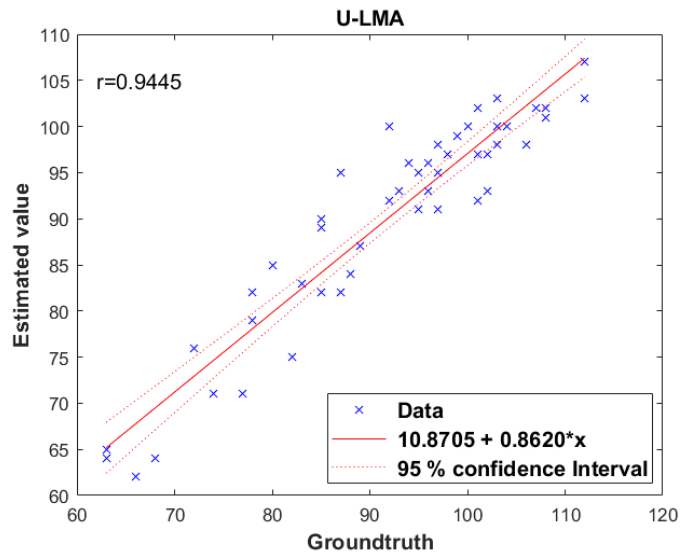


Fig. 8. Regression plot for Rigid and Non-rigid motion scenario.

collected considering natural conditions. Fig. 7 and Fig. 8 depicts the Bland-Altman, and the regression plots, respectively. As expected, the mean bias for this scenario is 1.84 bpm due to the presence of motion artifacts which means, the U-LMA predicts 1.84 bpm more than the traditional BVP. Consequently, the upper statistical limit of the Bland-Altman plot is slightly greater than 10 bpm; however, the method achieved a far lower statistical limit -6.6005. All of the data points lie between the upper and lower statistical limits. Interestingly, the ground truth and estimated HR values demonstrated a very high correlation (0.94), despite having a higher overall mean difference. Hence, the Bland-Altman analysis and the regression plot confirm the proposed method's effectiveness under challenging motion conditions, while also handling the abrupt rise and fall of HR values, when considering the mean of the HR values during the interval.

3) Illumination Variations Scenario: The COHFACE database is utilized to assess the ability of the U-LMA under different illumination scenarios. It is worth noting that the samples for the database possess motion artifacts too, but with the predominance

of uneven illumination distribution over the face due to ambient light. The performance analysis using Bland-Altman plot is presented in Fig. 9, while the regression plot is presented in Fig. 10. The mean bias achieved with the illumination scenario is -0.85 with lower and upper statistical limits of -9.7546 and 8.0546, respectively. Similar to the other scenarios, almost all the differences lie within the statistical limits. Furthermore, the Pearson correlation achieved under this scenario is 0.92. Both plots and their measured parameters proved the efficiency of the U-LMA for the HR estimation using illumination variant facial videos.

F. Comparative analysis

The available related conventional rPPG methods in the literature are based on a single color channel selection, ICA, color subspace transformations, and Wavelet based methods. U-LMA's performance was compared to all of these, except for Wavelet based methods, since these methods use the time-frequency domain and empirically set coefficients, unlike other PPG methods included in the study.

The single color channel selection method deals with utilizing a filtered signal extracted from a single color channel. Therefore, GREEN proposed by Verkryse et al. [4] was included, which extracts the BVP signal from the green color channel of the RGB color space. Most ICA based rPPG methods use ICA-Poh (JADE) [2], [28], [37], [38] and FastICA [9], [12], [50], [51]; hence they were included in this analysis. JADE uses kurtosis, whereas FastICA uses a negentropy based optimization function for an unmixing matrix estimation. Two color subspace transformations CHROM [10] and POS [14], were also included in the analysis due to their dependence on optimization procedures like ICA based methods. CHROM is a motion intolerant algorithm, while POS performs better for uneven illumination variations. The BCG method [52] was also included since it works on tracking the periodic movements of the head. As the effect of both types of motion on the proposed method is also tested in this study, it is worth including BCG as one of the state-of-art methods for performance comparison. Finally, the method presented by Song et al. [21], which is a combination of the color subspace transformation method and KDICA proposed by Ajou Chen [53], was also used to test the performance of the proposed method. The KDICA uses a Laplacian kernel for kernel density estimations which requires the pulse and artifacts spectrum to be in antiphase.

Furthermore, to assess the performance of the proposed non-linear cumulative density function which is approximated by tanh and a customized LMA, another variant of the proposed undercomplete analysis (U-neg) is also introduced, which utilizes the differential entropy or negentropy as an objective function. This objective function is optimized using the standard ICA procedure given by Hyvärinen et al. [42].

GREEN, ICA-Poh, CHROM, POS, and BCG were implemented using the standard implementation included in the iPhys toolbox by Mc Duff et al. [54]. Furthermore, the MATLAB implemented versions of FastICA by Hyvärinen et al. [42] and KDICA by Ajou Chen [53] were used to simulate the respective HR estimation methods as mentioned above, keeping other steps (ROI selection, bandpass filtering, and FFT) identical to the proposed U-LMA. All methods were tested under three scenarios, as explained in previous subsections. In other words, the video samples from all three databases were tested for all the methods used for comparative analysis by calculating RMSE, MAPE, mean error, standard deviation, accuracy, and Pearson correlation values under 0.01 significance level (α).

1) **Constrained scenario:** Similar to section IV-E.1, the videos from the VIPL-HR database were used for comparative analysis. The performance metrics for the comparative analysis are presented in Table III. Among all methods, BCG was the worst performing method for this scenario, despite minimal motion and illumination variation artifacts. BCG is susceptible to perform poorly in the presence of involuntary head movements, which may lead to false identification of face tracking points for estimation [52]. The CHROM, GREEN, ICA-Poh, and CHROM methods exhibited almost the same performance. The poor performance of GREEN becomes apparent due to inappropriate method formulations and the performance validations [28]. Furthermore, the original study suggested that along with green, the red and blue channel also contains complimentary PPG information [4], which is also confirmed in this study. The poor performance of ICA-Poh is due to lower frame rate as compared to the ground truth sensor value since this leads to an inappropriate mapping of BVP peaks, which in turn leads to inaccurate interbeat intervals for HR calculation [2]. CHROM's performance depends greatly on its alpha tuning procedure which works better for different magnitudes of specular distortions and pulse signal. The noise due to involuntary movements is inevitable, which might have degraded the performance [21]. Like CHROM, the POS method's accuracy depends significantly

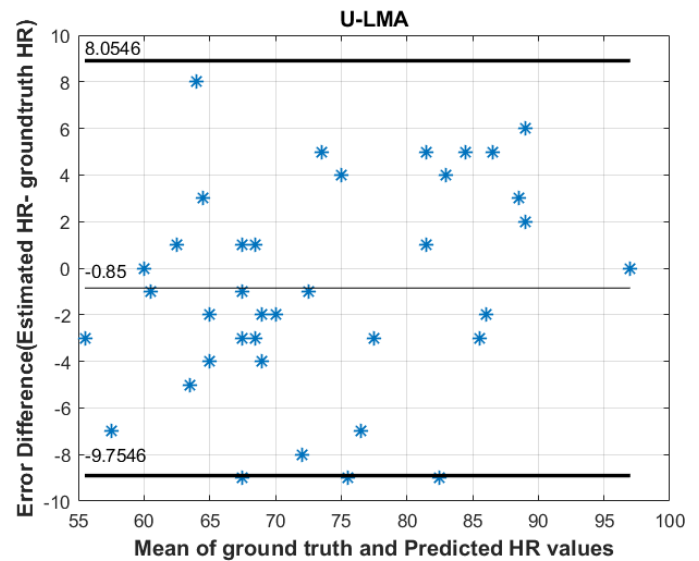


Fig. 9. Bland-Altman plot for Illumination variations scenario.

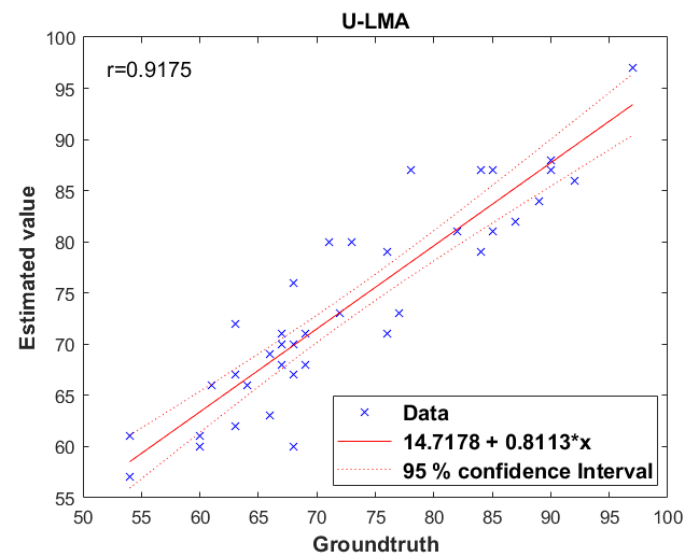


Fig. 10. Regression plot for Illumination variations scenario.

on its alpha tuning procedure which is suboptimal in the case of similar specular and pulse components magnitude. In this case, the specular variation components projected on the two axis may not be in absolute antiphase due to the presence of noise, which in turn leads to false estimations of alpha, and consequently into a poor BVP signal extraction. U-neg has performed relatively better than the above-mentioned state-of-the-art methods due to an effective information gathering from red, green and blue color channels. However, it failed to suppress the effect of inevitable noise.

On the other hand, Kernel ICA and FastICA have performed considerably better than other methods and U-neg. However, the performance of KernelICA suffered due to the same reason as CHROM. However, FastICA showed better performance than other methods except U-LMA which once more proved the effectiveness of negentropy based optimization function, by ensuring statistical independence among independent components. The proposed U-LMA achieved the best results, justifying its performance due to its ability to use higher order statistics for processing non-linear signals and effective optimization procedures using LMA. Moreover, the

TABLE III

PERFORMANCE METRICS FOR THE METHODS UNDER CONSTRAINED SCENARIO.

Methods	RMSE (bpm)	MAPE (%)	SD* (bpm)	Mean (bpm)	Accuracy (%)	r*
Green	21.4	24.7	12.3	17.6	1.89	0.23
ICA - Poh	19.15	22.37	11.70	15.21	2.83	0.24
CHROM	16.16	19.19	15.67	4.20	9.43	0.22
POS	18.80	22.30	11.94	14.56	0.94	0.31
BCG	24.87	27.79	12.82	21.35	7.55	-0.04
KernelICA	13.15	14.93	11.20	6.97	18.69	0.51
FastICA	13.10	15.37	10.38	8.06	19.62	0.60
U-neg	17.03	17.21	13.84	10.01	23.58	0.47
U-LMA	3.85	4.07	3.86	-0.35	84.91	0.92

SD*: Standard Deviation; r*: Pearson correlation is calculated at the 0.01 significance level; Accuracy is defined as the percentage of achieving the error difference with ± 5 bpm.

highest accuracy with clinically accepted error difference was also achieved by U-LMA.

2) Motion Scenario: The video samples from the UBFC-rPPG database were used to assess the effect of rigid and non-rigid motions on HR estimation. Table IV presents the performance metrics for all the compared methods. Overall, all the methods performed well under the motion scenario due to uncompressed videos. Similar to the constrained scenario, BCG performed the worst for the motion scenario as well. BCG method's performance was suboptimal due to the presence of rigid and non-periodic head movements [52]. An improved performance of the GREEN method indicates that the method is effective for uncompressed videos and is also data-driven. ICA-Poh performed relatively well due to the accurate selection of the BVP signal since there was no loss of information from the videos. Interestingly, the statistical independence among the components suffered due to the similarity of motion and pulse spectra under motion scenario, which led to the almost similar performance of ICA-Poh and FastICA. Furthermore, KernelICA and U-neg also demonstrated a similar performance for different reasons; KernelICA uses motion intolerant chrominance signals followed by KDICA, whereas U-neg uses a negentropy based function for the unmixing matrix estimation using undercomplete ICA, combining PPG information from all color channels. Although the RMSE, MAPE, mean error, and error standard deviation of U-neg was reduced, the accuracy was degraded in the motion scenario, as was to be expected. CHROM and POS performed relatively better than all the methods except U-LMA. This is due to their ability to perform well under motion scenarios due to extraction motion resistant signals followed by the alpha tuning procedure. Nevertheless, the proposed U-LMA outperformed all the methods, reporting the minimum value of errors, highest accuracy, and Pearson correlation, justifying its exceptional performance and clinical relevance.

3) Illumination Variation Scenario: The effect of illumination variations on the methods used for this study was evaluated using the COHFACE database. The GREEN method has shown a negative correlation for this scenario, indicating their susceptibility for uneven illumination distribution as it is susceptible to the illumination variation artifacts due to varying light intensity distribution [4]. ICA-Poh did not perform well due to low frame rate of the videos, as explained in the study conducted by Poh et al. [2]. The POS performance was suboptimal due to heterogeneous illumination conditions as a result of its assumption of independent intensity variations [14]. While, CHROM and BCG performed better than these three methods in terms of accuracy. However, these methods could not perform adequately due to the susceptibility of BCG for illumination variations [52] and due to considerably larger differences

TABLE IV

PERFORMANCE METRICS OF THE METHODS UNDER RIGID AND NON-RIGID MOTION SCENARIO

Methods	RMSE (bpm)	MAPE (%)	SD* (bpm)	Mean (bpm)	Accuracy (%)	r*
Green	28.08	20.26	25.20	12.90	44	0.34
ICA - Poh	20.49	13.34	19.56	6.72	58	0.54
CHROM	14.08	9.00	13.16	-5.35	66	0.70
POS	14.27	9.37	13.51	-4.98	62	0.71
BCG	36.08	33.75	16.90	31.97	8	0.03
KernelICA	20.40	14.67	18.90	8.12	46	0.59
FastICA	20.36	14.92	19.63	6.10	46	0.56
U-neg	14.97	12.86	11.27	9.98	22	0.59
U-LMA	4.57	4.00	4.22	1.84	78	0.94

SD*: Standard Deviation; r*: Pearson correlation is calculated at the 0.01 significance level; Accuracy is defined as the percentage of achieving the error difference with ± 5 bpm.

TABLE V

PERFORMANCE METRICS OF THE METHODS UNDER ILLUMINATION VARIATIONS SCENARIO.

Methods	RMSE (bpm)	MAPE (%)	SD* (bpm)	Mean (bpm)	Accuracy (%)	r*
Green	22.76	26.93	20.39	10.61	22.50	-0.07
ICA - Poh	24.88	28.52	17.84	17.56	35.00	0.06
CHROM	19.49	22.48	10.54	16.47	35.00	0.25
POS	26.28	30.71	12.63	23.14	22.50	0.05
BCG	25.53	27.00	15.72	20.27	2.50	0.19
KernelICA	14.50	12.61	11.60	8.90	50	0.36
FastICA	15.76	14.89	10.68	11.71	45	0.41
U-neg	11.13	13.03	10.76	3.33	25	0.60
U-LMA	4.48	5.16	4.45	-0.85	80	0.92

SD*: Standard Deviation; r*: Pearson correlation is calculated at the 0.01 significance level; Accuracy is defined as the percentage of achieving the error difference with ± 5 bpm.

between actual and estimated specular distortions in the video for CHROM [10]. Furthermore, the other ICA based methods KernelICA and FastICA performed relatively better than the methods mentioned above. However, the degraded performance of the KernelICA is due to the same reason as the CHROM method, along with the inability of the kernel density based ICA, when performing under a higher degree of illumination distortion. Specifically, the KDICA used laplacian Kernel, which failed to work due to illumination and pulse spectra overlapping. FastICA performed better than KernelICA due to the implication of a statistically better optimization function to separate specular and PPG information. On the other hand, U-neg achieved better results than other state-of-the-art methods demonstrating the significance of undercomplete ICA and negentropy. Furthermore, U-LMA achieved the lowest error values and highest accuracy and correlation values proving its superiority for HR estimation under illumination variations scenario. Table V presents the performance of these methods under illumination variations scenario.

4) RMSE analysis: The RMSE for the HR estimation has been predominantly analyzed in most studies conducted so far. It is calculated as the square root of the averaged squared error differences among different samples which provides the overall distribution of errors. Fig. 11 and Fig. 12 depicts the box and whisker plot of RMSE for analyzing the RMSE distribution among methods based on databases and vice-versa. These RMSE plots provide a deep insight into the performance of state-of-the-art and proposed methods for all the databases used in the study.

Fig. 11 shows that the UBFC -rPPG database was challenging for all methods used in the study due to its realistic conditions considered during video acquisition, whereas the performance with VIPL-HR is the best due to constrained conditions. The COHFACE

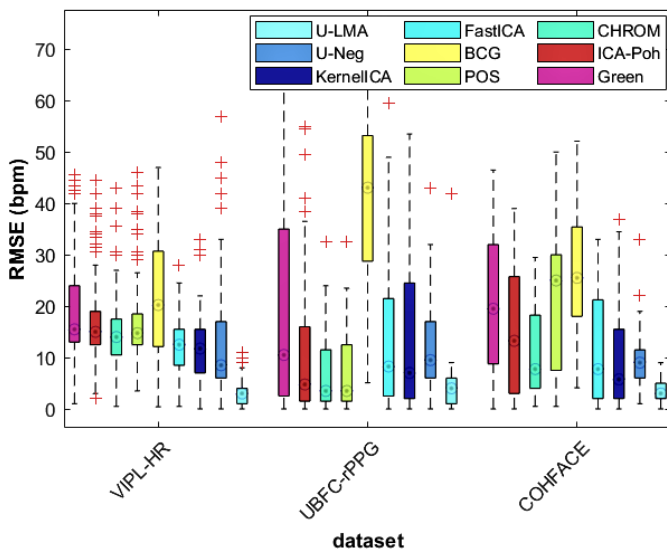


Fig. 11. Database wise RMSE comparison for the methods.

database is also challenging in terms of illumination variations throughout the samples. All methods, except BCG performed better under the motion scenario. Specifically, ICA-Poh, CHROM, POS, and U-neg illustrated their ability to deal with different types of motions. As mentioned before, the worst performing method is BCG, as depicted in the RMSE plots in Fig. 12. BCG was unable to adjust with the inevitable color distortions due to involuntary motions and illumination variations as mentioned in the original study [52]. The RMSE of the GREEN method was also very high since the study did not use any formulation for BVP signal extraction. The performance of all three ICA based state-of-the-art methods was similar, despite different objective functions used for unmixing matrix estimations. All of these methods suffered from the permutation problem, which makes it challenging to choose the appropriate BVP component and discarding other components simultaneously. Similar to ICA based methods, the color subspace transformation methods CHROM and POS also exhibited similar performance except for the illumination variations scenario in which the POS method failed to perform well. Since the videos were not recorded using an external source of light, causing serious illumination variations on different facial regions which produced an effect similar to multiple light sources (e.g., entirely black on one side and bright on the other side). U-neg performed relatively better than any state-of-the-art method, except for motion scenario, in which CHROM and POS achieved better performances than U-neg. Since both methods were able to extract the BVP signal due to uncompressed videos, that ensured extracting the detailed subtle color variations, which led to accurate BVP signal. However, higher RMSE of POS method for COHFACE database is due to heterogeneous illumination variations, which degraded its performance [14]. Furthermore, U-LMA performed far better than U-neg and other state-of-the-art methods either in terms of databases or state-of-the-art methods comparison. The proposed methods performed relatively better in all scenarios due to the proposed undercomplete ICA, which ensures better PPG information extraction from all three channels of RGB color space. Furthermore, despite U-neg used negentropy (differential entropy) for optimizing W with a standard ICA implementation proposed by Hyvärinen et al. [42], the experiments conducted during the study revealed that the entropy of CDF approximated by \tanh yielded better statistical independence than negentropy. Additionally, the lowest RMSE ranges by U-LMA when compared to U-neg in all scenarios were due to

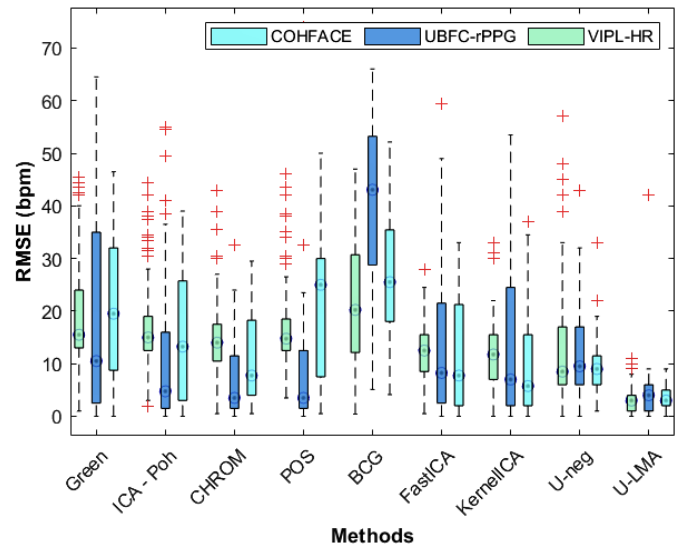


Fig. 12. Method wise RMSE comparison for the databases.

better optimization of an unmixing matrix W using customized LMA proposed in this work.

V. DISCUSSIONS

U-LMA outperformed all other methods in the various scenarios considered for the study due to its following components: Non-linear objective function (entropy of the CDF approximated by \tanh), and an efficient optimization algorithm (customized LMA). The non-linear function provided an advantage to counter the non-linearity associated with different types of motions and illumination variations. A customized LMA was able to find the global maxima for the entropy of the non-linear function in all samples with empirically evaluated damping parameter values. The study introduced U-neg which included undercomplete ICA with negentropy as an optimization function which allowed testing the effectiveness of the U-LMA components. Kurtosis was not used because negentropy exhibits better statistical properties than kurtosis [42]. Moreover, JADE used kurtosis, whose performance was surpassed by U-neg, as shown in the previous section. The performance comparison between both methods confirmed that processing utilizing higher-order statistics enhanced the performance of the proposed method. On the other hand, undercomplete ICA prevented the loss of BVP information by considering all the color channels, justifying the vitality of BVP information in red and blue channels for accurate HR estimation.

Following Bland-Altman and regression analyses, the mean bias of the illumination scenario is slightly further away from a zero error difference when compared to the constrained scenario, depicting the effect of illumination artifacts on HR estimation. As expected, the mean bias for the motion scenario is comparatively larger than the illumination and constrained scenario, respectively, due to the presence of rigid and non-rigid motions of the subjects in the video samples. Furthermore, the Pearson correlation value for the UBFC-rPPG database is higher by 0.2 bpm than other databases, which could result from the utilization of uncompressed videos for HR estimation.

The samples used under motion and constrained scenarios have the same sampling rate (30 fps), whereas the samples tested under illumination scenarios have a relatively lower frame rate (20 fps). It is worth mentioning pointing that a low frame rate results in missing blood volume variations, which can be captured using a higher sampling rate. The proposed U-LMA worked equally well for the videos captured with different frame rates. Hence, the blood volume

variations information loss due to compressed or low sampling rate videos did not significantly affect the proposed method for HR estimation.

Although the proposed method managed to confine most of the differences between the ground truth and the estimated HR values within the statistical limits, this did not justify its clinical relevance since none of the statistical limits were found to match with the clinically accepted error differences. Therefore, a new metric accuracy (error difference < 5 bpm) was defined to test the clinical relevance. The proposed U-LMA achieved sufficiently better accuracy justifying its clinical relevance as compared to other methods.

Nevertheless, U-LMA also possesses certain limitations that need to be addressed in the near future. First, although the effect of video compression and frame rates were tested, the effect of different camera-subject distances on the HR estimation was not addressed. The effect of different shooting distances could be tested on HR estimations by creating such databases followed by analyzing them. Second, although U-LMA was tested for the rigid and non-rigid motions, scenarios with periodic motions such as walking running, treadmill exercises, etc., or unconscious motions during sleep conditions were not addressed, which could be a future direction of the study. Third, the proposed method works well with various illumination variation conditions, but the effect of zero luminance or dark conditions was not addressed in this work, due to non-availability of reliable databases. The majority of the limitations for the proposed work are due to the limited or unavailability of sufficient benchmark databases. Finally, the proposed method only estimates HR, whereas other physiological parameters estimations like blood oxygen saturation, blood pressure, respiratory rate using BVP signals will be worth considering in future studies.

VI. CONCLUSION

This work addressed the BVP extraction as an undercomplete problem, while proposing the U-LMA. Taking into account the non-linearity due to motion and illumination artifacts, a novel entropy based non-linear function was proposed. The proposed non-linear function proved its effectiveness by addressing both type of artifacts. Furthermore, the non-linear function was optimized using the proposed customized LMA for entropy maximization, maximum statistical dependence due to better optimization of unmixing matrix W . The optimization using customized LMA, also aimed at reduced the effect of motion illumination artifacts. Additionally, the proposed method eradicates the need for IC selection and preserves the maximum possible BVP information from all channels of the RGB color space. A performance analysis for U-LMA was undertaken by comparing it under three scenarios: constrained, motion (rigid and non-rigid) and, illumination variations scenarios. The Bland-Altman analysis and regression plots proved the efficacy of the proposed method under all scenarios. Additionally, to check the influence of customized LMA and the entropy based CDF function, U-neg with a negentropy optimization function using standard ICA implementation was also tested. The results demonstrated the effectiveness of the proposed non-linear function and LMA by achieving comparatively better results than U-neg. In addition, U-LMA and U-neg, with other state-of-the-art methods, were also compared under the aforementioned scenarios using RMSE, MAPE, mean error, standard deviation, and accuracy (error difference $< \pm 5$ bpm), and Pearson correlation values. The results depicted the dominance of U-LMA in all scenarios by reporting comparatively the lowest error values, high accuracy and Pearson correlation values under 0.01 significance level. Higher accuracy values also suggest the applicability of the proposed method under clinical conditions.

VII. ACKNOWLEDGMENT

This work is supported by the LARSyS Project–UIDB/50009/2020, co-financed by Regional Development European Funds for the "Operational Programme Madeira 14–20"—Priority Axis 1 of the Autonomous Region of Madeira, number M1420-01-0145-FEDER-000002. The authors are also thankful to anonymous reviewers for their suggestions and recommendations to improve this manuscript for journal publication.

REFERENCES

- [1] S. Cook, M. Togni, M. C. Schaub, P. Wenaweser, and O. M. Hess, "High heart rate: a cardiovascular risk factor?," *European heart journal*, vol. 27, no. 20, pp. 2387–2393, 2006.
- [2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 7–11, 2010.
- [3] T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," in *2012 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 544–545, IEEE, 2012.
- [4] W. Verkrusye, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [5] Y. Yan, X. Ma, L. Yao, and J. Ouyang, "Noncontact measurement of heart rate using facial video illuminated under natural light and signal weighted analysis," *Bio-medical materials and engineering*, vol. 26, no. s1, pp. S903–S909, 2015.
- [6] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4264–4271, 2014.
- [7] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "Distanceppg: Robust non-contact vital signs monitoring using a camera," *Biomedical optics express*, vol. 6, no. 5, pp. 1565–1588, 2015.
- [8] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical signal processing and control*, vol. 13, pp. 102–112, 2014.
- [9] O. Gupta, D. McDuff, and R. Raskar, "Real-time physiological measurement and visualization using a synchronized multi-camera system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 46–53, 2016.
- [10] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [11] G. De Haan and A. Van Leest, "Improved motion robustness of remote-ppg by using the blood volume pulse signature," *Physiological measurement*, vol. 35, no. 9, p. 1913, 2014.
- [12] S. Kado, Y. Monno, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Spatial-spectral-temporal fusion for remote heart rate estimation," *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11688–11697, 2020.
- [13] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, 2018.
- [14] W. Wang, A. C. den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.
- [15] D.-Y. Chen, J.-J. Wang, K.-Y. Lin, H.-H. Chang, H.-K. Wu, Y.-S. Chen, and S.-Y. Lee, "Image sensor-based heart rate evaluation from face reflectance using hilbert–huang transform," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 618–627, 2014.
- [16] M. A. Hassan, G. S. Malik, N. Saad, B. Karasfi, Y. S. Ali, and D. Fofi, "Optimal source selection for image photoplethysmography," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1–5, IEEE, 2016.
- [17] T. Y. Abay and P. A. Kyriacou, "Reflectance photoplethysmography as noninvasive monitoring of tissue blood perfusion," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2187–2195, 2015.
- [18] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

- [19] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7411–7421, 2020.
- [20] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 568–574, 2013.
- [21] R. Song, S. Zhang, J. Cheng, C. Li, and X. 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*, vol. 116, p. 103535, 2020.
- [22] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1422–1433, 2016.
- [23] U. Bal, "Non-contact estimation of heart rate and oxygen saturation using ambient light," *Biomedical optics express*, vol. 6, no. 1, pp. 86–97, 2015.
- [24] M. Kedadouch, M. Thomas, and A. Tahan, "A comparative study between empirical wavelet transforms and empirical mode decomposition methods: Application to bearing defect diagnosis," *Mechanical Systems and Signal Processing*, vol. 81, pp. 88–107, 2016.
- [25] C. Wang, T. Pun, and G. Chanel, "A comparative survey of methods for remote heart rate detection from frontal face videos," *Frontiers in bioengineering and biotechnology*, vol. 6, p. 33, 2018.
- [26] C. Zhang, X. Wu, L. Zhang, X. He, and Z. Lv, "Simultaneous detection of blink and heart rate using multi-channel ica from smart phone videos," *Biomedical Signal Processing and Control*, vol. 33, pp. 189–200, 2017.
- [27] I. R. Tayibnapis, Y.-M. Yang, and K. M. Lim, "Blood volume pulse extraction for non-contact heart rate measurement by digital camera using singular value decomposition and burg algorithm," *Energies*, vol. 11, no. 5, p. 1076, 2018.
- [28] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics express*, vol. 18, no. 10, pp. 10762–10774, 2010.
- [29] P. Sahindrakar, G. de Haan, and I. Kirenko, "Improving motion robustness of contact-less monitoring of heart rate using video analysis," *Technische Universiteit Eindhoven, Department of Mathematics and Computer Science*, 2011.
- [30] J. Porrill and J. V. Stone, "Undercomplete independent component analysis for signal separation and dimension reduction," report, Citeseer, 1998.
- [31] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.
- [32] D. W. Marquardt, "An algorithm for least-squares estimation of non-linear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.
- [33] X. Niu, H. Han, S. Shan, and X. Chen, "Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video," in *Asian Conference on Computer Vision*, pp. 562–576, Springer, 2018.
- [34] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Un-supervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Letters*, vol. 124, pp. 82–90, 2019.
- [35] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *arXiv preprint arXiv:1709.00962*, 2017.
- [36] G. R. Tsouri, S. Kyal, S. A. Dianat, and L. K. Mestha, "Constrained independent component analysis approach to nonobtrusive pulse rate measurements," *Journal of biomedical optics*, vol. 17, no. 7, p. 077011, 2012.
- [37] R. Macwan, Y. Benezeth, and A. Mansouri, "Heart rate estimation using remote photoplethysmography with multi-objective optimization," *Biomedical Signal Processing and Control*, vol. 49, pp. 24–33, 2019.
- [38] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2593–2601, 2014.
- [39] P. Viola and M. 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.
- [40] T. M. Mahmoud, "A new fast skin color detection technique," *World Academy of Science, Engineering and Technology*, vol. 43, pp. 501–505, 2008.
- [41] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.
- [42] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [43] A. Tharwat, "Independent component analysis: An introduction," *Applied Computing and Informatics*, 2020.
- [44] M. K. Transtrum and J. P. Sethna, "Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization," *arXiv preprint arXiv:1201.5885*, 2012.
- [45] S. Amarai, A. Cichoki, and T. Chen, "A new learning algorithm for blind source separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.
- [46] J. J. Moré, *The Levenberg-Marquardt algorithm: implementation and theory*, pp. 105–116. Springer, 1978.
- [47] E. Fan, "Extended tanh-function method and its applications to nonlinear equations," *Physics Letters A*, vol. 277, no. 4-5, pp. 212–218, 2000.
- [48] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.
- [49] R. Artusi, P. Verderio, and E. Marubini, "Bravais-pearson and spearman correlation coefficients: meaning, test of hypothesis and confidence interval," *The International journal of biological markers*, vol. 17, no. 2, pp. 148–151, 2002.
- [50] K. Cai, H. Yue, B. Li, W. Chen, and W. Huang, "Combining chrominance features and fast ica for noncontact imaging photoplethysmography," *IEEE Access*, vol. 8, pp. 50171–50179, 2020.
- [51] K. Zheng, K. Ci, J. Cui, J. Kong, and J. Zhou, "Non-contact heart rate detection when face information is missing during online learning," *Sensors*, vol. 20, no. 24, p. 7021, 2020.
- [52] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3430–3437, 2013.
- [53] A. Chen, "Fast kernel density independent component analysis," in *International Conference on Independent Component Analysis and Signal Separation*, pp. 24–31, Springer, 2006.
- [54] D. McDuff and E. 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.