Robust remote heart rate estimation from multiple asynchronous noisy channels using autoregressive model with Kalman filter

Parham Nooralishahi⁎, Chu Kiong Loo, Liew Wei Shiung
Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

A R T I C L E   I N F O
Article history:
Received 28 December 2017
Received in revised form 3 August 2018
Accepted 3 September 2018
Available online 18 September 2018

Keywords:
Heart rate estimation
Blind source separation
Face tracking
Regression model
Interpolation
Search and rescue

A B S T R A C T
Remote heart rate measurement has many powerful applications, such as measuring stress in the workplace, and the analysis of the impact of cognitive tasks on breathing and heart rate variability (HRV). Although many methods are available to measure heart rate remotely from face videos, most of them only work well on stationary subjects under well-controlled conditions, and their performance significantly degrades under subject’s motions and illumination variation. We propose a novel algorithm to estimate heart rate. Also, it can differentiate between a photo of a human face and an actual human face meaning that it can detect false signals and skip them. The method obtains ROIs using facial landmarks, then it rectifies illumination based on Normalized Least Mean Square (NLMS) adaptive filter and eliminates non-rigid motions based on standard deviation of fixed length of the signal’s segments. The method employs the RADICAL technique to extract independent subcomponents. The heart rate measures for each subcomponent, are estimated by analysis of frequency signal to find the one with the highest magnitude. A two-steps data fusion method is also introduced to combine current and previous measured heart rates to calculate a more accurate result. In this paper, we explore the potential of our algorithm on two self-collected, and DEAP databases. The results of three experiments demonstrate that our algorithm substantially outperforms all previous methods. Moreover, we investigate the behavior of our algorithm under challenging conditions including the subject’s motions and illumination variation, which shows that our algorithm can reduce the influences of illumination interference and rigid motions significantly. Also, it indicates that our algorithm can be used for the online environment. Finally, the application of our algorithm in search and rescue scenarios using drones is considered and an experiment is conducted to investigate the algorithm’s potential to be embedded in drones.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Contactless physiological monitoring techniques open up many possibilities for easy and convenient continuous monitoring without hindering the daily-life routines of the monitored subjects. Traditional methods for physiological monitoring require sensors to be physically attached to the subject, such as photoplethysmography (PPG) finger-clip sensors, electrocardiogram (ECG) electrodes, and respiration (RSP) chest belts. Often they cause skin irritation and discomfort, in addition to hindering mobility in daily activities.

Many commercial mobile devices nowadays come with a built-in digital camera. The ubiquitousness of such devices allows ordinary commercial mobile devices such as smartphones, tablets, and laptops, to be employed for remote diagnosis, or telehealthcare. The main interest in developing remote diagnosis methods using commercial hardware is the low-cost due to not having to purchase specialized medical equipment and the convenience of having to use a readily available device. This is especially useful for telehealthcare in rural areas, where access to clinics or specialized medical equipment may be nonexistent. However, mobile devices such as smartphones can be used for self-diagnosis including heart rate monitoring [1], colorimetric tests [2], and diabetes management [3].

Identifying the underlying heart rate signal from a video recording of a subject is a challenging proposition. Under optimum conditions, the lighting and environment should ideally be static, and the video source should be recording using high-quality digital cameras. Practically, however, noise from video compression, body movements, and dynamic illumination introduces interference and may result in an inaccurate estimation of the heart rate signal [4]. Most computational estimation methods include frequency filtering to exclude signal frequencies that are above or below that of

⁎ Corresponding author.
E-mail addresses: parham.nooralishahi@siswa.um.edu.my (P. Nooralishahi), ckloo.um@um.edu.my (C.K. Loo), liew.wei.shiung@siswa.um.edu.my (L.W. Shiung).
heart rate signals but are unable to account for noise which is very similar to the heart rate signal itself. In addition, identifying the underlying heart rate signal using computational estimation methods \([5,6]\) typically selects only a single estimated outcome to represent the closest fit to the actual heart rate signal.

In this research, we introduce a contactless heart rate estimation schema that improves the accuracy of the measurement, compares to other methods. Additionally, the algorithm solves the challenges of unwanted signal influences including illumination variation, and rigid or non-rigid facial or environmental motions by using illumination rectification, non-rigid motion elimination, and few temporal filters. Despite the methods which are used to eliminate the effect of unwanted sources, some undetectable sources may sabotage the physiological signal in a realistic situation. Consequently, a data fusion algorithm is employed to adjust real-time measures based on the multiple measured heart rates and previous estimation.

The method employs various techniques to minimize the effect of rigid and non-rigid motions on the estimation and uses the RADICAL technique to extract independent components which outperform other ICA algorithms. Moreover, we proposed to estimate multiple heart rate measures from three different regions of the human face as well as three independent components resulted from applying ICA. So multiple measurements have been used in a proposed regression model which avoid sudden motions and noises and reduce the effect of rigid and non-rigid motions. The proposed technique increases the accuracy of the algorithm enormously as shown in the experiments.

The algorithm starts with extracting regions of interest using facial points. The physiological signals are calculated using the extracted regions. Then, an independent component analysis technique is adapted to extract subcomponents. The subcomponents are processed separately to obtain multiple heart rate measures which are later used in the data fusion method. The resulted components are transformed to the frequency domain using Fourier transformation. Afterward, peak points are obtained from the frequency signals with minimum peak distance of predefined length. Later, the frequency with the highest magnitude is chosen as the frequency of the heart rate for each signal. Finally, the heart rate measures are fused using the data fusion technique proposed in this paper, to calculate more accurate results.

In Section 2, we present a number of related works. In Section 3, we present the proposed methodology for heart rate estimation, including a method for tracking regions of interest (ROIs) of the subject’s face and an algorithm to compensate for motion. In Section 4, we present a proposed data fusion method to improve the accuracy of the heart rate estimation method using the current and previous heart rate measures. In Section 5, we present the results of three experiments that are designed to test and validate our methodology. In Section 6, we discuss the conclusions drawn from the experiment.

2. Related works

The innovation of using digital cameras in mobile devices for measuring heart rate (HR) is an extension of PPG technology. A pulse oximeter device for measuring PPG illuminates the skin and measures the variations in light absorption corresponding to a change in blood pressure from heartbeats. Early PPG sensors require LEDs to illuminate the skin and a photodiode at close range for measurement.

To extract the underlying physiological signals from the illumination of the skin, several methods can be applied. Linear filters are ineffective if background noise has the same frequency band as the desired physiological signal. Any given signal is assumed to be consisting of the desired physiological signals combined with other signal sources, or noise, such as from background illumination or if the frame of reference between the camera and the measured area changes. Blind source separation methods can be used to identify and extract the desired physiological signal. [7] designed a camera for capturing PPG information on several specific wavelengths simultaneously. Similarly, [8] and [9] utilized a multi-wavelength camera system for contactless measurement and calibration for estimating arterial oxygen saturation (SpO2).

When using only ambient light and commercial digital cameras, PPG information can be extracted using blind source separation by first separating the source images into the Red-Green-Blue (RGB) color channels [5,10–12]. Another study used Cyan and Orange color channels in addition to RGB [13]. Independent component analysis (ICA) was then applied to separate the image signals into source signals. A source signal is then selected as an approximation or estimation of the physiological of interest, typically using frequency analysis. For example, heart rate signals would have a strong signal in the frequency band corresponding to the human heart rate, around 0.75 Hz (45 bpm) to 4 Hz (240 bpm).

A method of heart rate estimation which is commonly used by smartphone apps requires only the user’s finger to be placed on the camera sensor, however, it is impractical and inconvenient for continuous, long-term measurement. Most contactless heart rate measurement methods observe the facial region, as it presents a wide surface area of the exposed skin particularly at the forehead or cheeks. This presents another dimension of difficulty as an automated physiological detection method would have to track the subject’s face and ROI. A popular face tracking algorithm based on [14] and [15] is used in several studies for ROI tracking [5,11,16].

3. Heart rate estimation

The proposed technique in this paper is composed of a number of steps as shown in Fig. 1. In the first step, ROIs on the subject’s face is located by using Constrained Local Model (CLM) presented in [17]. CLM algorithm extracts facial landmarks and generates a mask based on extracted points. Kanade-Lucas-Tomasi (KLT) algorithm [18] then is employed to track the location of featured landmarks in further image frames. The facial space is divided into three regions including the forehead, left and right cheeks based on facial landmarks. Forehead region is defined as the area between eyebrows and hairline, while left and right cheeks are defined as the areas below the eyes and above the lips.

After region extraction, a false signal detection algorithm [16] is applied to determine whether the extracted face belongs to an inanimate object and should be skipped or is an actual human face with
underlying physiological signal characteristics. Next, for each facial region in each video frame, the average of green, red, blue values of the region’s pixels are calculated as the three-dimensional raw pulse signals. The input image frames may not be captured in a constant frame rate, therefore Cubic Spline Interpolation [19] method is employed to form a piecewise continuous curve with a constant sampling rate. The resulted time series are normalized using its mean and standard deviation. The illumination rectification algorithm [20] is employed to reduce the influence of interference caused by illumination variation, camera vibration, etc. Moreover, a non-rigid motion elimination method [20] is also adopted to deal with non-rigid motions inside the regions. Afterward, Robust Accurate Direct Independent Component Analysis (RADICAL) method [21] is employed to extract a set of unobserved sources from available superposed mixtures.

All subcomponents are processed separately to obtain multiple heart rate measures which are later used in the data fusion algorithm. For each region, temporal filters are applied for eliminating unwanted signals and excluding powers of frequencies that are out of a predefined interval. Later, the signals are transformed to frequency domain using Fast Fourier Transformation (FFT) method [22]. The peak point detection algorithm is applied to indicate peak points. Eventually, the frequency with the highest magnitude is determined as the frequency of heart rate. The algorithm is detailed in Algorithm 1.

**Algorithm 1.** The proposed heart rate algorithm along with the data fusion method.

```plaintext
1: procedure HEART RATE ESTIMATION (frames)
2: for all images inside frames set do
3:   img = the current image from the frames and
4:   Perform face detection and facial landmarks extraction using
5:     [17] and [18].
6: regions = \{region1, ..., region_n\} ← extract regions of
7:     interest using facial landmarks. ▷ including forehead, left
8:     cheek, and right cheek.
9: for all regions inside the regions variable do
10:   reg = the next region from regions variable.
11:   state ← apply false signal detection algorithm on the reg
12:     data. ▷ the state is true, if the image is validated as real
13:     human face. Otherwise it sets to false.
14:   if state = true then
15:     \( (r, g, b) \) ← average all color channels of pixels of reg.
16:     sig_reg(end + 1) = (r, g, b)
17:   end if
18: end for
19: add all sig vectors to signals vector list.
```

**3.1. ROI detection and tracking**

Previous heart rate estimation methods mostly employ the Viola-Jones face detector [14] implemented in OpenCV [24] to detect faces in input images, but it only finds the bound of detected faces as rectangles, which is not enough for the proposed heart rate measurement since non-facial pixels at the ROI corners are also included which may have influence on estimation results. In this paper, CLM algorithm [17] extracts 66 facial landmarks as it is shown in Fig. 2, then Kanade-Lucas-Tomasi (KLT) algorithm presented in [18] is used to track the extracted landmarks.

In addition to the influence of non-facial pixels, facial body parts like eyebrow and lips as well as possible facial expression can have great influence on the estimation result. Therefore, we extract most stable regions as sampling areas. The regions are extracted based on the facial landmarks determined by CLM algorithm including,
(a) forehead which is the area between hairline and eyebrows, (b) left cheek which is the facial area below the left eye and above lips (c) and right cheek as shown in Fig. 2b. These regions are chosen, because they have the least possible movements, compares to other facial parts.

3.2. Physiological signal separation

According to [25], red, green, and blue channels all contain some level of plethysmographic signals. The green channel contains the strongest one among all three, because green light is better absorbed by oxy-hemoglobin than red light [26], and penetrates deeper into the skin to probe the vasculature as compared to blue light. However, in this paper, the three color channels are adopted because all of them have plethysmographic contents and can be used to separate the heart rate signals. Therefore the raw multidimensional time series are made of the mean values of all pixels of a region in each image frame, \( C_{\text{region}} = [c_1, c_2, \ldots, c_N] \) and \( C_i = [r, g, b] \), where \( n \) is the frame number, and \( r, g, b \) represents red, green, blue values.

3.3. False signal detection

An inanimate object, such as a drawing/picture of a human face or a dead human has no vital signs, therefore the estimated heart rate from this source of objects are false signals and detection of such cases from estimated results can be very difficult. As we experienced during this research, the existence of high illumination variation in the environment, subjects located out of reliable distance, or usage of low-quality cameras can create in-range heart rates for inanimate objects which cannot be eliminated after estimation. Therefore, by eliminating such falsify results, we can increase the reliability of the algorithm and in case of implementation for the real-environment, can decrease system load and a waste of system resources.

In this paper, we employ a false signal identification introduced in [16]. The probability distributions of the variation of peak power density ratio for the source signal before and after smooth filtering named as the ratio-variation PDs, where the peak power density ratio is the ratio of highest to total power within the power spectrum of the source signal. Moreover, the total power is calculated by adding the power within the frequency range of 0 Hz to 8 Hz. The variation of the peak power density ratio is defined in Eq. (1).

\[
\nu = \frac{\text{highest power before}}{\text{total power before}} - \frac{\text{highest power after}}{\text{total power after}}^{-1}
\]

\[
\times \left( \frac{\text{highest power total power before}}{\text{total power before}} \right)
\]

As it is determined in [16] the ratio-variation PDs of live humans and inanimate human-like objects are distinguishable. So it is introduced that the variation of peak power density ratio before and after smooth filtering for the true signals should be much smaller than for the false signals.

3.4. Interpolation

Although we require to assume that the camera has a fixed frame rate, capturing the frames in a consistent rate is almost impossible. Therefore an interpolation technique needs to be applied to interpolate the samples at a fixed rate. Previous researchers mostly used linear interpolation method because it is simple, easy to implement, but it fails to capture the curvature of the function which increases the local error. Although the accuracy can be improved by using more interpolating nodes, an important issue is the discontinuity of the first derivatives of the interpolating function at the sampling points. The goal of the cubic spline interpolation method beside getting a smoother interpolation function is to get interpolation formula that is continuous in both first and second derivatives within the intervals.

In this method the input time series named as \( f(x) \) is tabulated at the \( N + 1 \) points \( f_k = f(x_k), k = 0, 1, \ldots, N \). In each interval \( (x_k, x_{k+1}) \), a straight line can be fitted through the points \( (x_k, f_k) \) and \( (x_{k+1}, f_{k+1}) \) using Eq. (2) but the problem is that with a linear function, the first derivatives are not continuous at the boundary between two adjectives intervals, while the goal is that the second derivatives should be continuous even at the boundary. Therefore the interpolated function is used instead as shown in Eq. (2). The cubic spline interpolation technique improved the sampled signal by replacing the straight line connecting the data points with a third-degree polynomial.

\[
f = Af_k + Bf_{k+1} + Cf_k^r + Df_{k+1}^r
\]

\[
A = \frac{x_{k+1} - x}{x_{k+1} - x_k}
\]

\[
B = 1 - A
\]

\[
C = \frac{1}{6}(A^3 - A)(x_{k+1} - x_k)^2
\]

\[
D = \frac{1}{6}(B^3 - B)(x_{k+1} - x_k)^2
\]

Where \( A \) and \( B \) are linearly dependent on \( x \) and \( D \) also have cubic \( x \)-dependence through \( A \) and \( B \). The cubic spline interpolation generates data more similar to the vital signals. After the cubic spline interpolation model is formed, a signal with a 33 ms time span is generated based on the interpolation model.

3.5. Illumination rectification

In this section, The method introduced in [20] is employed to reduce the effect of the unwanted sources such as camera vibration, camera noise, illumination interference, etc. The mean green value of ROI is a function of time \( g_{\text{frame}} \). Two factors may affect the values of \( g_{\text{frame}} \), (a) the blood volume variations caused by cardiac pulse, (b) and the temporal environmental illumination variations. The method assumes the variations of \( g_{\text{frame}} \) caused by these two factors are as addictive as shown in Eq. (3).

\[
g_{\text{frame}} = s + y
\]

Where \( s \) denotes the green value variations caused by cardiac pulse, and \( y \) denotes the green value variations caused by illumination changes. Therefore, the goal would be to eliminate noise signal \( y \), and achieve \( s \) signal. In an ordinary environment, the lighting sources for the ROIs and the other objects in the environment are the same. Moreover, the imaging sensor’s vibration or noises also affect the whole image frame.

As described in [20], the method uses the background region \( g_{\text{bg}} = [g_1, g_2, \ldots, g_n] \) as a reference. We assume both the face ROI and the background share the same light sources and Lambertian models. The algorithm uses a linear function to estimate the correlation of \( y \) and \( g_{\text{bg}} \) as shown in Eq. (4).

\[
y = hg_{\text{bg}}
\]

\[
g_{\text{fr}} = g_{\text{frame}} - hg_{\text{bg}}
\]

\[
g_{\text{fr}} = s + (y - hg_{\text{bg}})
\]

Consequently, The algorithm minimizes the error which is part of \( (y - hg_{\text{bg}}) \) in (4). The optimal \( h \) can be found iteratively by using...
Normalized Least Mean Square (NLMS) adaptive filter. Let’s assume at each time point $j$, $h(j)$ is the currently estimated filter weight. The filter begins with an initial $h(0)$ and updates it after each step with a step-size $\mu$ as indicated in Eq. (5).

$$h(j + 1) = h(j) + \mu \tilde{g}(j) g_{bg}(j)$$

(5)

Least Mean Square filter is sensitive to the scaling of input signals, which can be deal with by normalizing the power of the input signals using Eq. (6).

$$h(j + 1) = h(j) + \frac{\mu \tilde{g}(j) g_{bg}(j)}{g_{bg}(j) g_{bg}(j)}$$

(6)

Where $g_{bg}(j)^H$ is the Hermitian transpose of $g_{bg}(j)$, and the $g_{bg}(j)^H g_{bg}(j)$ is the input signal energy. After applying the method to the input signal, the pulse signal reduces the illumination variations and becomes more visible. It should be noted that the optimal $h$ vary for different input videos since the distance of background and the face may vary and the reflectivity of the subjects’ skin also can be different.

3.6. Non-rigid motion elimination

The heart rate estimation algorithm is very sensitive to environmental changes expressly non-rigid motions. The rigid motions do not alter the shape or size of objects, while non-rigid motions like facial expressions, alter one or both of these features. The non-rigid motions may contaminate the data and cause falsely measures, particularly when the frequency of variations are close to the frequency of heart rate $HR_{freq}$.

In this paper, the employed method omits the data segments that may be affected by non-rigid motions. Based on the fact that the heart rate is estimated over fixed time span, the data is segmented into $m$ fixed-size segments $S = \{s_1, s_2, \ldots, s_m\}$. The standard deviation $\sigma$ of the segments are the key element to determine the affected segments. The $5\%$ of segments with the largest standard deviations are eliminated from the input signals. The algorithm is described in Algorithm 2.

Algorithm 2. The non-rigid motion elimination method.

1: procedure NON-RIGID ELIMINATION data \(\triangleright\) data is a one-channel time series.
2: \hspace{1cm} segsize \(\leftarrow\) 1 \times \text{length(data)}. \hspace{1cm} \triangleright m is the number of segments.
3: \hspace{1cm} \Sigma \leftarrow \{\}
4: \hspace{1cm} for $i \leftarrow 0$ to $m - 1$ do
5: \hspace{1cm} \hspace{1cm} $s_i \leftarrow \{\text{data}[i \times m], \ldots, \text{data}[(i + 1) \times m]\}$
6: \hspace{1cm} \hspace{1cm} $\Sigma(\text{end} + 1) \leftarrow \text{std}(s_i)$
7: \hspace{1cm} \hspace{1cm} end for
8: \hspace{1cm} Sort the elements of $\Sigma$ descendingly.
9: \hspace{1cm} $\Sigma \leftarrow$ the first $(0.5 \times m)$ elements.
10: \hspace{1cm} $S \leftarrow \{\}$ \hspace{1cm} $\triangleright$ The $S$ is the final data.
11: \hspace{1cm} for all $s$ in $\Sigma$ do
12: \hspace{1cm} \hspace{1cm} if $s \notin S$ then
13: \hspace{1cm} \hspace{1cm} \hspace{1cm} Append $\{\text{data}[i \times m], \ldots, \text{data}[(i + 1) \times m]\}$ to $S$
14: \hspace{1cm} \hspace{1cm} end if
15: \hspace{1cm} \hspace{1cm} end for
16: \hspace{1cm} return $S$.
17: end procedure

3.7. Independent component analysis

Although the algorithm has the highest performance in the environment where no unwanted source contaminates the signal, in a realistic environment, the signals may contain multiple sources combined together. Therefore an independent component analysis method is used to separate these sources. RADICAL algorithm [21] is a fast and accurate non-parametric ICA method which is robust to outliers also. It maximizes the statistical independence, rather than a surrogate for this measure. The RADICAL method uses entropy minimization at its core, also its marginal entropy estimates are functions of the order statistics of the signals. Generally, the algorithm applies to whiten on the input signal, then it minimizes the sum of the entropies of the mixture marginal under rotations. According to [21], RADICAL outperforms FastICA [27] and JADE [28]. The extracted subcomponents may or may not be relevant in terms of heart rate estimation. Therefore the signals are considered individually to estimate multiple heart rate measures which are later used in the data fusion method to calculate more accurate results.

3.8. Temporal filtering

Three filters are employed to exclude frequencies outside of interest range. In this paper, the frequency interval is [0.7, 4.0] Hz to cover the normal range of heart rate from 42 to 240 beats per minute. The real heart rate signal is usually non-stationary like slow linear or more complex trends which can cause distortion of frequency data [29]. The first filter is a detrending filter introduced in [23], which reduces the slow and non-stationary trend of the signal. The filter is based on smoothness priors approach and operates like a time-varying FIR high pass filter. A Hamming window-based finite impulse response bandpass filter with a cut-off frequency of [0.7, 4.0] Hz is employed as the third filter to remove all data without range frequencies.

3.9. Heart rate measurement

Signals are transformed to frequency domain using Fast Fourier Transformation (FFT) method [22]. Afterward, all peak points of frequency signals with a predefined length, are identified as frequency candidates. The frequency candidate with the highest magnitude is determined as the frequency of heart rate $freq_{HR}$. Eventually, the selected frequency is converted to beats per minute using Eq. (7), where $S$ indicates sampling rate, $L$ represents the number of FFT samples, and $c$ is the calibration value. The calibration constant is required to cover the constant lag obtained from preliminary studies on a separated data set.

$$HR = 60 \times \left( \frac{S \times freq_{HR}}{L} \right) + c,$$

(7)

4. Channel fusion and regression model

The heart rate estimation algorithm may face many challenges in a real situation such as illumination variation, shadow changes, and facial movements which can disturb the physiological signals and may increase estimation error. Many of the environmental changes are hard or impossible to eliminate, for example, microchanges, or vibrations that their frequency is close to heart rate frequency, are impossible to remove and have an enormous effect on the result. In this paper, a data fusion method is proposed to calculate more robust and reliable results by fusing the multiple heart rate measures based on independent sources and combining the current and previous heart rate measures. The algorithm also provides confidence factor values for heart rate measures based on the Gaussian distribution presented in Eq. (8).

According to [30], a normal resting heart rate for adults eighteen or older, is between 60 and 100 beats per minute (bpm), depending on the person’s physical condition and age, so the algorithm initializes the Gaussian distribution presented in Eq. (8), where the initial $HR_{init}$ and $\sigma_{init}$ are set to 80 and 30.
confidence factor \( P(H_{R_t}(HR_{t-1}, \sigma_{t-1}) = \sqrt{2\pi} \exp \left( -\frac{(x-HR_{t-1})^2}{2\sigma_{t-1}^2} \right) \) 

Where \( \sigma_t \) is the standard deviation of the estimated heart rates and \( HR_t \) is the heart rate estimation of time \( t \). Then, the measures that their confidence factors are lower than a predefined threshold \( \alpha \), are eliminated. Later, the mean value of \( m \) heart rates with the highest confidence values is calculated as the candidate heart rate measure. If the confidence factor of the resulted heart rate is lower than \( \alpha \), the result will be ignored. The final measure is hired to update the regression model used to estimate the heart rate.

Even though the described method can remove the influence of shadow changes, facial expressions on the signal, the changes like illumination variation, body movement, or etc. can contaminate the whole data and sabotage the final results. Hence, we modified the method introduced in [31] that combines an autoregression model and Kalman Filter algorithm to predict \( HR_{t+1} \) using \( p \) previous heart rate measures including the current heart rate measure. We adopted the AIC criterion to get the optimal order \( p \). The method uses the estimated heart rate to update the regression model, and predict the \( HR_{t+1} \) using the same model. In this method, the regression model predicts the heart rate measures while the estimated heart rate is only used to update the regression model. The proposed algorithm uses the first \( p \) estimations to initial the regression model, then it starts to predict the heart rate at time \( t = p + 1 \) at the same time as the estimated heart rates are adopted to correct the regression model in each iteration. The regression model is an incremental procedure which supposedly becomes more precise over time, so the algorithm may face an instability which will be solved over time.

Auto regression model of order \( p \) claims that an imperfectly predictable variable depends linearly on its own previous values. The auto regression model of order \( p \) is defined as Eq. (9),

\[
X_t = c + \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t
\]

where \( A_1, \ldots, A_p \) are the parameters of the model (coefficients), \( c \) is a constant value, and \( \epsilon_t \) is white noise. There are many ways to calculate the coefficients such as least squares technique, or method of moments [32]. In this paper, Kalman filter is adopted to estimate and update the coefficients. Kalman filter is an optimal estimator. It is incremental, so its estimation is improved over time. The introduced method addresses the problem of trying to estimate the state \( x \) that is governed by the linear stochastic difference formula as shown in Eq. (10) and a measurement \( z \) as shown in Eq. (11),

\[
x_k = A x_{k-1} + B u_k + w_{k-1}
\]

\[
z_k = H x_k + v_k
\]

where the random variables \( w_k \) and \( v_k \) represent the process and measurement noise. It should be noted that they are assumed to be independent, white, and with a normal probability distribution.

The \( x_k \) is estimated based on the weights between input vector \( x(t-1), \ldots, x(t-p) \) and parameter \( A_{i}(t) \) (shown in Eq. (9)). In this research, transition matrix \( B_{i,p} \) is a diagonal matrix that its diagonal elements are equivalent to one. Process noise covariance matrix \( Q_{i,p} \) is also a diagonal Gaussian covariance matrix that its mean and standard derivation are respectively 0.0 and 0.4 and the initial error matrix \( e_{i,p} \) is a matrix that all its elements are equivalent to one. The measurement matrix \( H_{i,p} \) is the last \( p \) elements of the input vector, and finally, the noise covariance matrix of measurement model is a Gaussian distribution that its mean is 0.0 and its standard derivation is 0.2. The estimated state in time \( t \) is employed as coefficient values of autoregression model in time \( t \).

5. Experiment

We evaluate our algorithm using three experiments. The first experiment demonstrates the performance of all components inside the algorithm. Moreover, it evaluates the algorithm in dynamic environments. We investigate how much facial expressions, head movements, and illumination variation can affect the results and discuss the efficiency of the employed methods to deal with the changes. The second experiment evaluates the algorithm using the DEAP benchmark database presented in [33]. The experiment demonstrates the potential of the proposed algorithm compares to other methods in a more challenging environment. The third experiment explores the potential of using heart rate estimation algorithm in search and rescue operations. Moreover, it investigates the ability of our algorithm to reduce the effect of rigid and non-rigid motions. Additionally, the experiment compares the recent methods with the introduced algorithm.

There are different kinds of statistics were adopted in previous researches for evaluating the robustness and efficiency of the heart rate measurement methods. In this paper, we use three different kinds of statistics that commonly used in previous researches as discussed in [20]. The first factor is the mean of estimation error denoted as \( Me_{ct} \); the second factor is the standard deviation of estimation error \( SD_{ct} \); the third factor is the mean of error-rate percentage \( Me_{Rate} \) as shown in Eq. (12), where \( N \) is the number of videos of the targeted database, \( HR_{error} \) is calculated by Eq. (13), and \( HR_{gt} \) is the heart rate measures that are recorded by a device.

\[
Me_{Rate} = \frac{1}{N} \sum_{t=1}^{N} \frac{|HR_{error}(v)|}{HR_{gt}(v)}
\]

\[
HR_{error} = HR_{video} - HR_{gt}
\]

This paper follows the experiment plan described in [20], the following methods are considered for evaluation including, (a) Poh2010 [5] which adopts an Independent Component Analysis (ICA) method to convert color brightness averaging signals to independent components. (b) Kwon2012 [34] which employs JADE ICA technique to extract independent subcomponents from color brightness averaging signals. Finally, it measures the heart rate using frequency analysis of extracted subcomponents. (c) Poh2011 [11] that uses JADE ICA algorithm to extract independent components, then the source signals are smoothed using a five-point moving average filter. A 128-point Hamming window bandpass filter to eliminate data outside of [0.7, 4.0] Hz interval. The heart rate measurement is performed by power density estimation using the Lomb periodogram. (d) LI2014 [20] which proposes an illumination rectification, and a non-rigid motion elimination algorithms to remove unwanted signals. Finally, it employs Welch’s power spectral density estimation method to estimate heart rate frequency. All of the techniques employed peak detection algorithms to find the location of each heartbeat for further heart rate variation analysis except LI2014. In both experiments, the proposed method compares with these methods.

5.1. Experiment 1: Evaluation of remote heart rate estimation in real situations

In this experiment, we collected 135 video samples from 27 different subjects while they are wearing Nonin Onyx II Model 9560. In order to evaluate the estimation result, the heart rate measurement has been collected continuously during the experiment and the average of every five seconds has been employed for validation.
Table 1
Four different video samples are collected from each subject. In the first two samples, it is assumed that no head movement or facial expression is occurred, the samples are recorded in two different lighting conditions. The third sample includes head movements, and the fourth sample includes facial expressions.

<table>
<thead>
<tr>
<th>Samples for each subject</th>
<th>Duration</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without any movements or facial expression</td>
<td>2 min</td>
<td>2</td>
<td>Two video samples in two different lighting conditions are recorded. The first sample is recorded with normal illumination intensity, and the second video is recorded with poor lighting condition. It is recommended to the subjects that they avoid sudden movement. It is told to the subjects that express random emotions.</td>
</tr>
<tr>
<td>With head movement (slow and continuous movements)</td>
<td>2 min</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>With facial expression</td>
<td>2 min</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. A photo of the first experiment setup. The user sits on a chair facing the screen. Also the camera placed at 100 cm distance.

purposes. The participants are asked to perform specific actions in each video session, such as head movements, and random facial expressions. The details of the samples are listed in Table 1. Frontal face videos were recorded for each subject in 24-bit RGB color format with the resolution of 600 x 800 pixels at 30fps and saved in MPEG format, while they were staring at the camera with 100 cm fixed distance as shown in Fig. 3. Heart rate measures were collected using the device fifty times per minute by Lenovo G50-70 containing a Core i5 @ 1.70 GHz x 4 processor and 4 GB RAM.

In this experiment, aside from evaluating all components of the algorithm, the potential of the proposed method in different conditions are tested. The experiment investigates how the algorithm behaves when it faces rigid or non-rigid motions, additionally it tests the algorithm in different illumination intensities to discuss the effect of illumination intensity on the efficiency of the estimation method.

The image frames are stored in a fixed-size queue at 30 fps. The size of the queue is equivalent to 160 frames which are about five seconds of recording. The three regions of interest are extracted using CLM method. Then physiological signals are calculated using a simple average technique. The generated data is interpolated using cubic spline interpolation method. In this research, the data is collected at 30 fps while the interpolated signals are sampled at 100 samples per second. Fig. 4 shows the generated signals of about 80 frames, and it indicates that the cubic spline interpolation method provides smoother results than linear interpolation method. Additionally, it generates the data more similar to the vital signals and minimizes the local error, compares to the linear interpolation technique employed in previous researches.

The illumination interferences and rigid motions like camera vibration, sabotage the physiological signals. Fig. 5a shows the signal of the green channel in three different illumination intensities, and also the signal resulted by applying illumination rectification method. The experiment demonstrates that physiological signal becomes too weak in poor lighting conditions. Fig. 5a indicates that, although the adopted method can eliminate the effect of illumination interference, it cannot boost the strength of the physiological signal in case of poor illumination intensity.

Based on our experiments, non-rigid motions like facial expressions inside the extracted regions can contaminate the signal. Fig. 5a indicates the signal that is collected when the user expresses surprise and laughter. The forehead shows extreme changes during surprise expression, also left and right cheeks represent significant variation onset of a laughter. The variations may contaminate the interpolated signal and its power spectral density distribution when it transforms to the frequency domain since they contain the major part of the power of the whole signal. Therefore eliminating these parts can improve the resulted estimation.

In this experiment, the interpolated signal is divided into 12 segments (m = 12). As shown in Fig. 5b, the standard deviations of all segments are calculated. The 5% of segments with highest standard deviations are chosen to be eliminated.

As discussed earlier, the environmental and subjective changes such as camera noise and vibration, head movement, and illumination variation may cause combining different frequencies with the targeted signals, hence we employed an ICA algorithm to extract these components. Here, the signal is passed to the RADICAL algorithm to extract subcomponents as shown in Fig. 6.

Fig. 4. Cubic and linear interpolation results based on brightness average of green channel for more than 80 frames.
experiment, each region includes three average brightness signals, so totally nine signals are passed to the ICA method to extract nine sources which may contain physiological data.

Three temporal filters are adopted to remove the frequencies outside of the defined frequency interval [0.7, 4.0]Hz. In Fig. 7, it shows the original signal and the results of applying detrending and FIR bandpass filter. It is clear that the final signals provide more clear physiological data.

The heart rate measures for an experiment on ninety image frames and the initialized Gaussian distribution is shown in Fig. 8a. The method demonstrates notable potential in reducing the effect of local changes and sudden movements. Moreover, the results which are shown in Fig. 8b indicates that the algorithm can minimize the effects of undetectable unwanted sources and unreliable results.

The algorithm is evaluated using all video samples available in the dataset. The results of the proposed algorithm are compared to the contact sensor measurements. Fig. 9 shows estimation error spectrograms which compare the camera measures with the contact measures in different conditions, despite the camera being placed 1 meter from the participant. The experiment indicates that the mean of estimation error $M_E$ is 1.42 with restriction of motion
Fig. 6. photo (a) shows the brightness color average of green channel, and photos (b)–(d) are extracted subcomponents by RADICAL technique.

Fig. 7. The figure represents the results of two temporal filters. The first part is the original signal, the second part is the detrended fluctuation, and the last part is the filtered signal using FIR bandpass filter.
and illumination and the $SD_e$ is equivalent to $1.637$. Also, it shows that the mean of estimation error increases while the illumination intensity of the environment is dropping.

The same experiment is conducted without restriction of motion and illumination changes, the algorithm has the mean of estimation error $M_e$ of $8.65$ and the $SD_e$ of $3.68$ in poor lighting condition. It also evaluated for rigid and non-rigid motions, which they have respectively the $M_e$ of $2.37$ and $3.64$ and the $SD_e$ of $2.57$ and $3.07$. The experiment demonstrates significant results in the environments with normal illumination intensity and further, it proves that the algorithm can tolerate rigid motions and reduce their effects. In Fig. 9, we show the error spectrograms of the estimated heart rate by comparing them with the measures collected by contact sensor. The vertical axis demonstrates the estimated measures and the horizontal axis shows the contact sensor measures. For each of the mentioned situations in the figure, a batch of measured and collected results has been shown that can be used to analyze the behavior of the algorithm. As shown in Fig. 9, in the environment with normal illumination intensity, the collected an estimated heart rate measures are close to each other which means the estimation error is minimized. On the other hand, in the environment with weak illumination intensity, the estimation error has been increased and the reliability of the algorithm elevated adequately, because of the degradation of vital information in the input signal. Moreover, the algorithm demonstrates trustable behavior facing rigid motions since the average error is in an acceptable range and no fluctuated results have been detected during the experiment. In case of non-rigid motions, untrustable estimation has been detected. Although the average estimation error is in an acceptable range, the variety of fluctuated results decrease the reliability of estimation. Therefore, it can be implied that heart rate estimation shows instability in the case of weak illumination intensity and non-rigid motions.

Finally, the algorithm is evaluated with the other methods. All video samples of 27 subjects recorded with restriction of motion and illumination are used for evaluation. From Table 2, we can see that all methods work perfectly with $M_{eRate}$ lower than $3\%$ which means all methods including our own algorithm got almost perfect results. The results of the database are almost ideal because of restrictions of illumination variation and motions. As shown in Table 2, the proposed data fusion technique has a significant impact on the algorithm’s results which means the fusion technique makes the algorithm more robust and stable. In realistic situations, many challenges may occur, so to test the reliability and robustness of the algorithm in a more difficult situation, we carry out an experiment over a more difficult database in the next experiment.

Fig. 8. The photo (a) shows the confidence factor distribution and estimated heart rates of 90 image frames. Photo (b) indicates the 38 heart rate measures and the final heart rates which is predicted by the introduced regression model.
Fig. 9. Estimation error spectrograms calculated using contact and camera measures. The figures show the estimation errors of four different experiment conditions, (a) normal illumination intensity, (b) weak illumination intensity, (c) rigid motions, and (d) non-rigid motions. The horizontal and vertical axes indicate respectively the contact measures as well as the camera measures. The videos of 5 subjects are used and the heart rate estimations are obtained using a 120 window frames.

<table>
<thead>
<tr>
<th>Method</th>
<th>$M_e$(SD) (bpm)</th>
<th>Mean of error-rate ($M_{\text{data}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poh2010</td>
<td>1.77(1.69)</td>
<td>1.87%</td>
</tr>
<tr>
<td>Kwon2012</td>
<td>-1.93(2.23)</td>
<td>2.14%</td>
</tr>
<tr>
<td>Poh2011</td>
<td>1.78(1.98)</td>
<td>2.24%</td>
</tr>
<tr>
<td>Li2014</td>
<td>2.65(2.93)</td>
<td>2.87%</td>
</tr>
<tr>
<td>Our algorithm</td>
<td>1.42(1.637)</td>
<td>1.49%</td>
</tr>
<tr>
<td>Ours without data fusion</td>
<td>2.99(2.84)</td>
<td>2.91%</td>
</tr>
</tbody>
</table>

5.2. Experiment 2: DEAP database

In this experiment, we test the proposed algorithm on the DEAP database. We demonstrate the potential of our algorithm by comparing the estimated results with collected heart rate data, and we investigate the accuracy of the algorithm, compares to other algorithms. The database is considered as a difficult database since both illumination variation and motions are involved. Details regarding the setup environment are explained in [33].

DEAP database is a multi-modal dataset for the analysis of human affective states [33], the electroencephalogram (EEG) and physiological signals of 32 subjects were recorded while they were watching 40 one-minute long music videos. Participants rated each video in terms of arousal, valence, like/dislike, dominance, and familiarity. Frontal face videos were recorded for twenty-two of thirty-two participants. Table 3 shows more details about the database.
The results on DEAP database are shown in Table 4. All the calculation is done the same way as we did for the first experiment. The performance of all the methods dropped significantly facing real situation challenges. Li2014 method performs better than the other three methods because it adopts several temporal filters which we also used it in our algorithm. But the $M_{Rate}$ of 6.48% shows that the method is not as robust as our algorithm to make reliable estimations about the true heart rates. Our algorithm represents the best results among the other methods with the mean errors of about 3 bpm. The standard deviation measure demonstrates the stability of the algorithms, therefore our algorithm shows more stability and robustness compares to other algorithms.

5.3. Experiment 3: Using remote heart rate estimation on search and rescue drones

Without the restriction of motion and illumination changes, the method can be applied to the mobile robots like a Quadrotor. In this experiment, we explored the potential of applying Heart rate estimation algorithm on unmanned aerial vehicles (UAVs) for search and rescue operations. Moreover, it conducted to determine the reliability of the heart rate estimation algorithm in a different operational environment. The experiment has two parts, the first part includes the videos that are recorded in Advanced Robotic Lab., Wisma R&D, University of Malaya, and the second part involves the videos which are recorded in an indoor sport court of Residential College 2, University of Malaya. For both of parts, the experiment is carried out in low and high illumination intensity. The second part has more illumination variation and the drone flies at higher altitude. So the heart rate estimation algorithms are evaluated in two different environments.

This experiment featured three subjects, different ages (18–30), and multiple skin colors as partly demonstrated in Fig. 10. One of the subjects were wearing glasses and another one had facial hairs. During the experiment, participants were asked to lie down and face the drone’s camera while their video was recorded. Moreover, the heart rate measures are recorded using a certified wearable sensor Nonin Onyx II Model 9560, at 30 samples per second. The heart rate measure has been collected from selected participants during the experiment while the drone was hovering above the target. Then the recorded time estimations are used for evaluation.

Both of the videos and data are aligned using timestamps simultaneously.

Basically, the heart rate estimation method measures the micro-changes of the human face during a specific amount of time. So movements, illumination variation, and other micro-changes can disturb the result enormously, particularly the motions with a frequency near the heart rate can make the estimation unreliable. Therefore reducing these variations can increase the accuracy of the algorithm. The Parrot Bebop drone is equipped with anti-vibration bumpers embedded inside its structure, additionally, the drone employs video stabilizer to take stable and clear footage regardless of the drone movement. The heart rate estimation is evaluated using five different video footages of three persons pretend to be injured and lie down on the ground in two different indoor environments as shown in Fig. 11. The heart rate sensor is utilized to record his/her heart rate during the operation. We adopted the mean of error-rate percentage $M_{Rate}$ as shown in Equation (12) to evaluate the robustness and efficiency of the recent heart rate estimation.

In this experiment, the following methods are considered for evaluation including, (a) Poh2010 [5], (b) Kwon2012 [34], (c) Poh2011 [11], (d) and Li2014 [20]. All of the techniques including our novel algorithm employed peak detection algorithms to find the location of each heart beat for further heart rate variation analysis except Li2014.

Our algorithm is tested against the mentioned techniques using the collected data. The results of the comparison are indicated in Table 5. All averages of $M_{Rate}$ results for the algorithm are higher than 10% which indicates that the environment is very complex and many unwanted sources have influenced on the data. Gener
ally the recorded videos demonstrate high instability of the drone and additionally, they indicate that Env. 2-1 and Env. 2-2 conditions have lower complexity compare to the other two conditions. Even though our algorithm and Li2014 have the lowest error rate percentage, our algorithm simply provides more robust measurement. Furthermore, Poh2011 shows poor performance in Env. 1-1 and Env. 1-2 compare to Li2014, because Poh2011 has lack of rigid and non-rigid motion elimination abilities. Poh2010 and Kwon2012 methods have the highest error-rate percentage since they cannot deal with unwanted signals and the environmental changes that may disturb the data. The introduced method shows significant results in environments with normal illuminations, and acceptable in poor lighting condition. Also, the result of our method without data fusion indicates the importance and the effect of the proposed technique for improving estimation results.

Fig. 12 demonstrates the result of applying the proposed heart rate estimator on six seconds of video footage recorded in Advanced Robotic Lab. in normal lighting condition. It indicates that the data fusion technique has remarkable improvement on the estimated heart rates.

### 6. Conclusion

Previous algorithms can measure heart rate from face videos with high accuracy under conditions with high restriction of motions and illumination variation. The methods are not able to deal with environmental changes in real situations which prevents them to be utilized in real life scenarios. We proposed a novel heart rate estimation algorithm that can reduce environmental interferences. A method based on Normalized Least Mean Square adaptive filter is employed to reduce the effect of rigid motions. Additionally, non-rigid motions are eliminated based on standard deviation of fixed length of the signal’s segments. Also, two temporal filters are adopted to smoothen the signal and remove all frequencies outside of [0.7, 4.0] interval. The RADICAL algorithm is adopted to extract independent subcomponents, further, the components are processed separately to measure multiple heart rates. Furthermore, a two-steps data fusion technique based on autoregression method combined with Kalman filter is introduced to calculate the more accurate result using current and previous heart rate measures.

The algorithm is evaluated using three different databases including two self-collected datasets, and DEAP database. The first dataset is used to investigate the potential of the proposed method. It explores the algorithm’s behavior in environments with and without restrictions of subject’s motions and illumination variation. Moreover, the experiment analyzes all components of the algorithm to prove their usability. DEAP database is utilized since all interferences are involved in the videos. The third experiment investigates its potentials to be embedded in a search and rescue drone. All experiments demonstrate that our algorithm substantially outperformed the previous methods. Moreover, our technique shows significant results facing illumination interference and rigid motions.

In future work, we will work on designing a more efficient algorithm for detecting the peak of each heartbeat, so more realistic and sophisticated analysis about heartbeat variation can be determined which can further use to extract more information about the subject’s physiological status. Moreover, we plan to improve the non-rigid motion elimination method.

### Acknowledgment

This work is sponsored by the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. UM, grand challenge’s project number is GC003A-14HTM.

### References


