Accurate Heart Rate Estimation From Camera Recording Via MUSIC Algorithm

Seyyed Hamed Fouladi1, Ilangko Balasingham1,2, Tor Audun Ramstad1 and Kimmo Kansanen1

Abstract—In this paper, we propose an algorithm to extract heart rate frequency from video camera using the Multiple Signal Classification (MUSIC) algorithm. This leads to improved accuracy of the estimated heart rate frequency in comparison with traditional methods based on ViolaJones (VJ) approach [8] and the KanadeLucasTomasi (KLT) algorithm [9]. A better tracking performance is limited by the number of samples and frame rate. Monitoring vital signs remotely can be exploited for both non-contact physiological and psychological diagnosis. The color variation recorded by ordinary cameras is used for heart rate monitoring. The orthogonality between signal space and noise space is used to find more accurate heart rate frequency in comparison with traditional methods. It is shown via experimental results that the limitation of previous methods can be overcome by using subspace methods.

I. INTRODUCTION

Pulsatile tissue volume can be measured directly by photoplethysmography. The total sum of volume changes is measured in the blood vessel by conventional photoplethysmography. Photoplethysmography (PPG) is related to, but not the same as traditional photoplethysmography. Photoplethysmography is an electro-optic scheme for measuring the tissue blood volume pulses (BVP) [1]–[3]. The vital signs such as heart rate (HR), respiratory rate, and arterial oxygen saturation, can be obtained by extracting PPG.

In non-contact PPG, variations in ambient light and adjusted brightness in camera are some parameters which have a significant influence on the DC part of the signal. They can be mentioned such as the variations in ambient light and adjusted brightness in camera, and physiologic parameters like variations in capillary density and in venous volume fluctuations. The blood hemoglobin absorbs light in comparison with other tissues. When arterial blood volume varies during the cardiac cycles, light absorption of the human skin fluctuates consequently [4]. Vital physical signs, such as heart and breath rate, and arterial oxygen saturation can be estimated by recording the tiny color variation along with minor light intensity variation on the skin. Non-contact PPG scheme is an innovative method which needs a region of interest (ROI), selected from the human face, to obtain cardiac pulsations using ordinary low-cost cameras of laptops or smartphones. Pixel values of region of interest are summed up at each frame. The AC part of the resulting one-dimensional signal contains the heart rate information. However, this observation is contaminated by noises and motion artifacts. Several methods have been proposed to remove undesired signals. In [5], the authors supposed that the PPG signal can be represented in small number of coefficients in wavelet domain and the undesired signals and the PPG signal are distinguishable in this domain. In [6], it has been assumed that PPG signal is non-Gaussian and independent from motion artifacts and noise. Thus, the PPG signal is extracted by maximizing the non-Gaussianity.

Low number of samples are available for short time intervals in ordinary cameras typically 25-30 frame per second (fps). As a result, the accuracy of estimation decreases in the case a low number of samples are available. By increasing the time interval, more unwanted signals and motion artifacts are added to the desired signal as well as having non-stationary behavior of PPG signal. In this paper, it is intended to overcome the availability of low number of samples by high resolution spectrum estimation method. This method, results in more accurate in comparison to the conventional methods which used traditional approaches like Fast Fourier Transformation (FFT) and periodogram for finding the HR frequency [6], [7]. The new algorithm proposed to estimate HR frequency is based on Multiple Signal Classification (MUSIC) algorithm, which can distinguish between harmonic components better than FFT based approaches as used in the state-of-the-art methods. To our best knowledge, we are the first to use a subspace method, MUSIC algorithm, to estimate HR frequency for such an application.

The paper is organized as follow. In Section II, high resolution spectrum estimation, theory of MUSIC algorithm and proposed algorithm are described. In section III, simulation results are presented. Section IV concludes.

II. ESTIMATION OF HEART RATE FREQUENCY USING MUSIC ALGORITHM

The following general steps should be done to estimate the heart rate frequency. Fig.1 shows how HR frequency is extracted, generally.

A. ROI Tracking

At the first step, ROI should be tracked to find the proper location in all the recorded frames. There are face tracking methods based on ViolaJones (VJ) approach [8] and the KanadeLucasTomasi (KLT) algorithm [9]. A better tracking results by using the KLT approach in comparison with the VJ algorithm, since the KLT algorithm exploits face tracking by
tracking feature points [7], [10]. After ROI tracking, the data is preprocessed before noise and motion artifacts removal.

B. Motion Artifacts and Noise Removal

Beside the face tracking, it is necessary to remove the noise and unwanted signals. Generally, a digital camera sensor has red, green, and blue (RGB) color channels. In [5], [8], [11], the color difference method in the color channels is used to estimate the non-contact PPG signal through a linear mixture of the RGB channels. In this method, it is supposed that the sources including noise, motion artifacts and desired signal have a non-Gaussian distribution. Since a linear mixture of the sources is available at each channel, the probability density function of observation tends to Gaussianity due to central limit theorem. Therefore, the mixture coefficients and sources can be estimated by maximizing non-Gaussianity of the output components. By having the mixture coefficients, the signal contains HR frequency, motion artifacts and noises can be separated.

C. Multiple Signal Classification algorithm

The MUSIC algorithm is exploited in order to achieve a better frequency resolution as one of the subspace methods for spectral analysis. The energy of the PPG signal containing HR frequency information is concentrated in the frequency domain. This property can be used in MUSIC algorithm for acquiring HR frequency.

Assume that we have the sum of signal, \( x(n) \), and additive noise, \( e(n) \), with variance \( \sigma^2_e \) at the observation, \( y(n) \),

\[
y(n) = x(n) + e(n).
\]

\( x(n) \) contains \( p \) complex sinusoids, \( s_k(n) \), including heart rate frequency, \( f_{HR} \), and undesired frequencies like motion artifacts frequencies.

\[
x(n) = \sum_{k=1}^{p} w_k e^{j2\pi f_k n} = \sum_{k=1}^{p} w_k s_k(n)
\]

where \( w_k \) and \( f_k \) are coefficient and frequency of the \( k \)-th \( (k = 1, \cdots, p) \) complex sinusoid. Accordingly, the signal autocovariance function, the inverse Fourier transform of the Power Spectral Density (PSD), becomes,

\[
r_s(\tau) = \sum_{k=1}^{p} |w_k|^2 e^{j2\pi f_k \tau} = \sum_{k=1}^{p} P_k e^{j2\pi f_k \tau},
\]

where \( \tau \) and \( P_k \) are the lag and power respectively. If the signal dimension is \( N, N > p \), \( r_s(\tau) \) can be calculated for the lags \( \tau = 0, \cdots, N-1 \). Therefore, the related autocovariance matrix can be obtained by:

\[
C_y = \begin{bmatrix}
    r_s(0) & r_s(1) & r_s(2) & \cdots & r_s(N-1) \\
r_s(1) & r_s(0) & r_s(1) & \cdots & r_s(N-2) \\
r_s(2) & r_s(1) & r_s(0) & \cdots & r_s(N-3) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_s(N-1) & r_s(N-2) & r_s(N-3) & \cdots & r_s(0)
\end{bmatrix}
\]

(4)

With the Carathéodory-Fejér-Pisarenko result for Toeplitz matrices, the autocovariance matrix can be written as [12], [13]:

\[
C_y = C_x + \sigma^2_e I
\]

\[
C_x = \sum_{k=1}^{p} P_k s_k s_k^H
\]

(5)

where \( C_y, I \) and the superscript \( H \) are autocovariance matrix for \( y \), identity matrix and the Hermitian transpose respectively and the vector \( s_k \) is:

\[
s_k = \left[ 1, e^{j2\pi f_k}, e^{j4\pi f_k}, \cdots, e^{j2(N-1)\pi f_k} \right]^T
\]

(6)

The MUSIC method relies on the eigendecomposition where the eigenvalues of \( C_y \) are ordered as \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \). In the noise free case, the eigenvalues \( \lambda_1, \lambda_2, \cdots, \lambda_p \) are non-zero and the eigenvalues \( \lambda_{p+1}, \lambda_{p+2}, \cdots, \lambda_N \) are zero. The eigendecomposition of autocovariance matrix of \( y \) can be written as:

\[
C_y = QAQ^H
\]

\[
\Lambda = \begin{bmatrix}
    \lambda_1 & 0 & \cdots & 0 \\
    0 & \lambda_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & \lambda_N
\end{bmatrix}, \quad Q = [q_1, q_2, \cdots, q_N]
\]

(7)

By equations (5) and (7), we have

\[
C_x = \sum_{k=1}^{p} \lambda_k q_k q_k^H
\]

(8)

and also:

\[
\sigma^2_e I = \sum_{k=1}^{p} \sigma^2_e q_k q_k^H
\]

(9)

Therefore,

\[
C_y = \sum_{k=1}^{p} (\lambda_k + \sigma^2_e) q_k q_k^H + \sum_{k=p+1}^{N} \sigma^2_e q_k q_k^H
\]

(10)

There are two groups of eigenvectors. The signal subspace is \( \mathcal{S}_x = \text{span} \{q_1, q_2, \cdots, q_p\} \) and the noise subspace (that is orthogonal to signal subspace) is \( \mathcal{S}_{\text{noise}} = \text{span} \{q_{p+1}, q_{p+2}, \cdots, q_N\} \).
span \( \{ q_{p+1}, q_{p+2}, \ldots, q_N \} \), accordingly, \( S_y = S_x \oplus S_{\text{noise}} \). It can be written as:
\[
S_y^H q_{p+1} = 0. \tag{11}
\]

The spectral MUSIC is written as:
\[
S_{\text{MUSIC}}(f) = \frac{1}{s_f^H Q_{m+1:n} Q_{m+1:n}^H s_f}. \tag{12}
\]

The \( p \) local maxima of \( S_{\text{MUSIC}}(f) \) of are related to the available frequencies, \( f_k \), in \( x \). It means that the denominator of the fraction in equation (12) tends to zero when \( f \) equals to one of the signal component due to orthogonality the signal space with the noise space.

The values of \( f_k \) are the \( p \) closest roots to the unit circle with the amplitude less than 1 of the following equation:
\[
s_c^H Q_{m+1:n} Q_{m+1:n}^H s_c = 0,
\]
\[
s_c = [1 \ z^1 \ z^2 \ \ldots \ z^{N-1}]^T. \tag{13}
\]

In order to use MUSIC algorithm, it is needed to obtain the model order \( p \) that can be adjusted by experiments or estimated by the proposed method in [14]. After applying the MUSIC method, a process should be done to find the heart rate frequency. The following subsections are allocated to model order selection and finding the heart rate frequency.

1) Selection of Order \( p \): In [14], a criterion was proposed for selecting the model order \( p \) by minimizing the Minimum Description Length (MDL) criterion:
\[
MDL(p) = -N \log \left( \frac{G(p)}{A(p)} \right) + E(p)
\]
\[
G(p) = \prod_{k=p+1}^{N} \lambda_k
\]
\[
A(p) = \left( \frac{1}{N-p} \sum_{k=p+1}^{N} \lambda_k \right)^{N-p}
\]
\[
E(p) = \frac{1}{2} p (2N - p) \log N
\]

The number of sinusoid can be computed by this algorithm.

2) Finding HR Frequency: First find the available frequencies \( f_k \) in the signal component \( x \). It is probable that the motion artifacts frequencies exist in \( f_k \). The most powerful harmonic between 0.5 Hz - 4 Hz is selected as HR frequency. It should be mentioned that \( S_{\text{MUSIC}}(f) \) in equation (12) is the pseudospectrum estimate of the signal. The powers can be found by mapping \( s_f \) related to the estimated frequencies into signal space \( Q_{1:m}^H \).

In Table I, the steps of the proposed algorithm are summarized.

### III. Experimental Results

In the experimental results, a camera is used with 640 \times 480 pixels and frame rate 30 fps. The spectrum by MUSIC algorithm is compared with traditional algorithms. The amount of ambient sunlight is varying entering through windows and is the only source of illumination while the experiments were done indoors. The distance between the participants and the laptop web-cam is approximately 0.5 m. It should be mentioned that the first 300 frames (10 seconds) are not used in the experiments due to automatic camera adjustment. During the experiment, the HR is measured by a Pulse Oximeter as a reference to check the methods accuracy. Since denoising or face tracking is not the aim of the proposed method and the main goal is improving the HR estimation quality with low number of samples, the participants were asked to be static during the recording.

In the first experiment, the MUSIC spectra and spectra found by FFT algorithm are evaluated with a low number of samples (60 samples or 2 seconds). In both methods, the signal is denoised by ICA (JADE algorithm [15]). In Fig.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>THE STEPS OF HR FREQUENCY EXTRACTION USING MUSIC ALGORITHM</strong></td>
</tr>
</tbody>
</table>

| 1) Preprocessing to find ROI |
| 2) ROI tracking |
| 3) Processing to find one dimensional signal from each channel |
| 4) Applying a denoising method like ICA to find \( y(t) \) in equation (1) |
| 5) Applying MUSIC Algorithm |
| a) Setting model order \( p \) |
| b) Finding covariance matrix \( C_t \) by equation (4) |
| c) Obtaining eigendecomposition of \( C_t \) as equation (7) |
| d) Finding eigenvectors \( q_k \) in which \( k = p + 1, \ldots, N \) |
| Namely, \( S_{\text{noise}} = \text{span} \{ q_{p+1}, q_{p+2}, \ldots, q_N \} \) for forming \( Q_{m+1:n}^H \) |
| e) Finding spectral MUSIC by: |
| \( S_{\text{MUSIC}}(f) = \frac{1}{s_f^H Q_{m+1:n} Q_{m+1:n}^H s_f} \) |
| 6) Obtaining most powerful component in frequency range 0.5 Hz - 4 Hz |

![Fig. 2. MUSIC and FFT based spectrum for 60 samples (2 seconds)](image-url)
2, the estimated spectrum by MUSIC and FFT are shown. The HR was 70 bpm recorded by the Pulse Oximeter during the experiment time. The MUSIC and FFT based spectrum peaked at 1.187 (71.22 bpm) and 1.475 (88.50 bpm) respectively. The result of MUSIC is more precise when a low number of samples are available, the MUSIC algorithm can distinguish between the harmonics in signals and provide high resolution spectrum which is not possible by FFT-based approaches. In other words, two close harmonic components merge with each other by using FFT-based approached.

In the second experiment, it is intended to evaluate the spectrum estimation methods in terms of number of samples. The number of samples is increased from 60 to 210 samples. For each case, we apply ICA to remove the noises and motion artifacts and extract HR frequency by FFT and MUSIC spectra. The 120-second recordings of eight participants are used in this experiment. The estimated frequency by both methods \( \hat{f}_{HR} \) is compared with the measured HR by using Pulse Oximeter (\( f_{HR} \)). The error is calculated simply as 
\[
e = |\hat{f}_{HR} - f_{HR}|.
\]

In Fig. 3, mean and standard deviation of the error for MUSIC and FFT-based spectrum are shown. As it can be seen, the MUSIC algorithm outperforms in the low number of samples case.

IV. Conclusion

In this paper, we proposed a new non-contact method for estimating heart rate. The proposed method is also useful when a low number of samples is available whereas FFT-based method does not work well in this case. We used the orthogonality of noise subspace and signal subspace to represent the signal spectrum. Using the MUSIC approach, we could distinguish between the available harmonics in the signal. As a result, the HR frequency was more accurately estimated by the proposed method. By this approach, the problem of previous FFT based methods was resolved.

V. Acknowledgment

The research work was performed, in part, of the research project Medical sensing, localization and communications using ultra wideband technology (MELODY) project no. 25122502.

REFERENCES