REMOTE PHOTOPLETHYSMOGRAPHY USING NONLINEAR MODE DECOMPOSITION

Halil Demirezen, Cigdem Eroglu Erdem

Marmara University
Department of Computer Engineering,
Goztepe, Istanbul, Turkey

ABSTRACT

Remote Photoplethysmography (rPPG) is a contactless non-invasive method for measuring physiological signals such as the heart rate (HR) using the light reflected from the facial tissue. Signal decomposition approaches are used to extract the heart rate signal from the subtle changes in the skin color. In this paper, we show that a recently proposed signal decomposition method, namely nonlinear mode decomposition (NMD), is quite successful in estimating the heart rate signal from face videos in the presence of subject motion. Experimental results on the PureDL dataset show that NMD based HR estimation gives better results as compared to well-known methods in the literature, which use Independent Component Analysis (ICA) for signal decomposition.

Index Terms— photoplethysmography, NMD, PPG, heart rate estimation

1. INTRODUCTION

Photoplethysmography (PPG) is a non-invasive method used both in contact and non-contact approaches for measuring physiological signals such as the heart rate (HR), or the respiratory rate (RR) using the light reflected from or transmitted through the tissue. Recent years have witnessed a boom in non-contact PPG research, thanks to successful developments which allow the PPG signal to be extracted from face images under ambient illumination using cheap cameras [1].

Photoplethysmography etymologically is a compound term where the "plethysmo" means "enlargement" in Greek. It was first mentioned by Hertzman et al. in the 1930s [2]. PPG is based on the fact that the light which penetrates the microvascular layer in the skin is absorbed by the oxy-hemoglobin in the blood directly proportional to the blood volume. Thus, the changes in the reflected or transmitted light provides us an estimate of the heart activity in terms of the HR (heart rate) and HRV (Heart Rate Variability), which is the heart rate change with respect to the time providing valuable information about the metabolism.

Photoplethysmography methods can be categorized into two groups: contact and non-contact methods. An example of contact photoplethysmography is the fingertip pulse oximeter which is widely used in medical centers today [3].

In non-contact photoplethysmography, the physiological parameters are extracted using cameras from the subtle color changes on the face caused by the heart beat. Much of the emitted light from a light source is reflected from the surface of the skin. Whereas, a small amount is refracted and travels to deeper layers such as the epidermis, and the dermis. The light is refracted and reflected while crossing each skin layer, resulting in a very small amount being absorbed by the oxy-hemoglobin in the blood. Within every heart beat cycle, the amount of blood in the tissue changes periodically, so does the amount of light absorbed by the oxy-hemoglobin. Although the amount of light that reaches the capillary region is very small, the periodic changes in the reflected light from this region can be measured in terms of color changes in the skin. Since it is low in power, measurement of the cardiac cycle is prone to be spoiled by noise. Other estimation artifacts are caused by head and facial motion and illumination changes.

In this work, we present a method for non-contact heart estimation using facial videos based on a recently proposed signal decomposition method, namely the nonlinear mode decomposition (NMD) [4], which is shown to be extremely noise-robust as compared to other signal decomposition methods in the literature. Therefore, it is very suitable for the signal decomposition task in rPPG, especially when there is subject motion, which introduces noise to the skin color information.

The organization of the paper is as follows. In Section 2, we give an overview of the related work on rPPG and point out potential areas for research. In Section 3, we provide an overview of the nonlinear mode decomposition method. In Section 4, we describe how we adopted the NMD method for the problem of remote heart rate estimation. In Section 5, we give our experimental results and show that NMD gives promising results as compared to other popular signal decomposition methods in the literature, especially when there is noise due to subject motion. Finally, in Section 6 we provide concluding remarks.

2. RELATED WORK

Most of the noncontact photoplethysmography methods in the literature follow similar basic steps, which are shown in Fig. 1. After the video frames are captured, a region of interest (ROI) is selected on the face at each frame in which the PPG signal is to be estimated. In order to obtain a high quality PPG signal, selection of the ROI is important. The ROI should include all pixels that contains the PPG signal, whereas it should not include the pixels that lack the PPG signal (due to occlusions, background etc.). Although ROI selection is challenging, efficient selection of the ROI dramatically decreases the false estimations. There are a number of studies in the literature that mainly focus on the ROI selection [5], [6], [7].

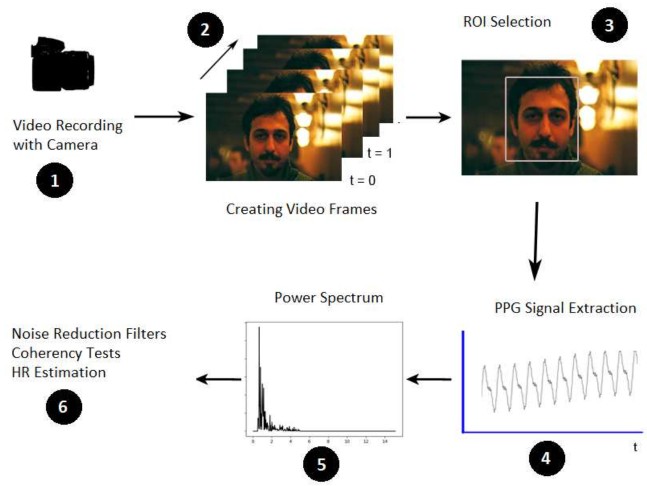

Fig. 1. Basic steps of a noncontact PPG based heart rate estimation algorithm

After the selection of the ROI, the color values of all pixels within the ROI are averaged. While a group of the studies used only the green channel [8], [9], [10], a significant number of methods used all three of the RGB color component values [3], [11], [12]. One of the first methods that use all three color components is based on the independent component analysis (ICA) [13]. Three signals constructed from average R, G, B values in the ROI are linearly separated into three independent components using ICA. One of the resulting components after ICA is expected to be the HR signal [3], [11].

The next step is peak detection either in time domain or frequency domain to estimate the heart pulse rate. The peak locations in the time domain signal can be estimated, which show the start of systole phase of the cardiac cycle. The number of peaks within a minute represents the heart pulse rate. While detection of the peak locations in time is responsive to instant heart rate changes, it is easily affected by noise [1].

The frequency domain analysis of the decomposed components requires sufficient number of samples, such as a 30-second long sliding window with 1 second shifts, leading to a delay in estimation of the changes in the heart pulse rate. Selecting a small temporal window size results in noisy estimates. Thus, there is a tradeoff between robustness and responsiveness [1].

In the last step, additional constraints are applied, such as discarding the frequencies that lie out of a certain range (e.g. by band-pass filtering the [30-250] beats per minute range) or applying coherency constraints.

Rigid and non-rigid motion of the subject, rapid illumination changes in the environment, and occlusions are the main factors which lead to errors in estimation of the HR. In order to improve the estimation accuracy, different ROI selection [9], signal decompostion methods [3], [11], illumination models [10], [7], multispectral aproaches [14], and multicamera setups [15] have been proposed and compared. However, these challenges still require further research, since the accuracy of rPPG is still not high enough to be used in clinical settings.

In this work, we present a nonlinear mode decomposition based heart rate estimation method, which is robust to noise, adaptively parametrized, and which results in physically meaningful modes.

3. NONLINEAR MODE DECOMPOSITION

Nonlinear Mode Decomposition (NMD) is a recently proposed signal decomposition method, which separates a signal into its physically meaningful oscillations, while removing noise [4]. Since real oscillations are rarely purely sinusoidal, it is assumed that a given signal $s(t)$ is composed of nonlinear modes $c_i(t)$ and some additive noise $\eta(t)$:

$$s(t) = \sum_i c_i(t) + \eta(t). \tag{1}$$

The components $c_i(t)$ are of the following form:

$$c(t) = A(t)v(\phi(t)) = A(t) \sum_h a_h \cos(h\phi(t) + \psi_h) \tag{2}$$

where $v(\phi(t)) = v(\phi(t) + 2\pi)$ is a periodic function of the phase and can be expressed using Fourier series expansion as shown by the summation in (2). The goal of nonlinear mode decomposition is to determine the characteristics of the NMs in (1) by estimating the amplitudes $A(t)$, phases $\phi(t)$, the amplitude scaling factors a_h and phase shifts ψ_h of the harmonics indicated in (2).

NMD is based on time-frequency representations (TFR) of the signal $s(t)$. The TFR used is either the windowed-Fourier Transform (WFT) or the wavelet transform (WT). There are four basic steps of NMD, which are given are as follows:

(i) Extracting the fundamental harmonic of an NM from the TFR representation.

(ii) Finding all possible harmonics of the fundamental harmonic.

(iii) Selecting the true harmonics of the same NM.

(iv) Building the nonlinear mode from fundamental and true harmonics, subtracting it from the signal and reapplying the same steps starting from (i) to estimate the remaining modes from the residue.

For details of the above steps the reader is referred to [4]. The advantages of NMD over other signal decomposition methods in the literature, such as empirical mode decomposition (EMD) [16] and independent component analysis (ICA) [13] have been stated as follows [4]:

- NMD is extremely noise-robust.
- The parameters of the algorithm are adaptively chosen.
- NMD returns modes which are physically meaningful since if an individual mode has a non-sinusoidal waveform, NMD will not decompose it into a few oscillations with simpler waveforms.

As an example of NMD, let $s(t)$ be a signal with two NMs and a white noise $0.5\eta(t)$:

$$s(t) = (10 - 0.03t)cos\phi_1(t) + 10cos\phi_2(t) + \\ +0.5\eta(t), \tag{3}$$

where

$$\phi_1(t) = 10\pi t,$$
$$\phi_2(t) = 14\pi t + 2\pi sin(2\pi t/10).$$

A plot of the TFR of the signal in (3) is given in Fig. 2 (a). This TFR is obtained by WFT. The NMD decomposition of the signal is given in Fig. 2 (b), which shows that the two oscillation modes have been successfully separated.

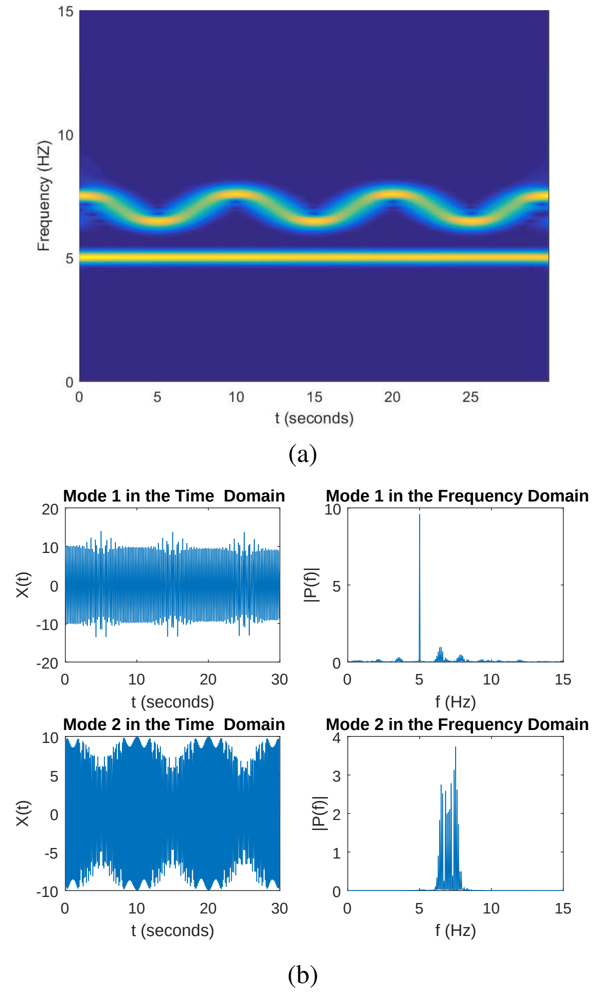

Fig. 2. (a) TFR representation of signal s(t). (b) Recovered Modes of s(t) given in (3).

4. HEART RATE ESTIMATION USING NMD

We have adopted the nonlinear mode decomposition method to solve the problem of remote heart rate estimation from face videos (step 4 in Fig. 1).

At each frame of a given video, the face of the subject is first detected using the Viola-Jones face detector [17], which returns the location of the face using a rectangular window. Since some non-face regions are also included, we use a subregion within this window as the region of interest (ROI). The ROI rectangle is obtained by decreasing the width of the window by 60%. We keep the ROI selection process simple in this work, to show the advantages of NMD over other signal decomposition methods under noisy conditions.

Then, the green (G) color component of each pixel within the ROI is averaged at each frame and concatenated, giving a 1D signal in time, $G(t)$. The first heart rate estimate is obtained after a signal of length 30 sec is obtained from the averaged G values of 900 frames (30 sec x 30 fps = 900). In order to obtain the next heart rate estimation, the window of 900 frames is shifted by 30 frames, providing a 96.7% overlap between two successive windows.

Before applying NMD to each 30 sec length window $G(t)$, we normalize them as follows:

$$\tilde{G}(t) = \frac{G(t) - \mu_G}{\sigma_G} \tag{4}$$

where μ_G and σ_G denote the mean and standard deviation of the average G values in the window, respectively. The NMD algorithm is then applied to the normalized signal $\tilde{G}(t)$, which decomposes the signal obtained from the face pixels into its modes. In our implementation, we run the NMD algorithm to decompose the signal into five modes. We expect the number of modes to be five, each component corresponding to the heart rate, the respiratory rate, the rigid head motion, the non-rigid facial motion and illumination changes, if any.

Each estimated component signal $c_i(t), i = 1, \ldots, 5,$

is then converted to the frequency domain via FFT (fast Fourier transform) to obtain $C_i(f)$, where f denotes the frequency. Then, $C_i(f)$ are band-pass filtered with a passband of [0.75Hz - 2.5Hz] (45bpm - 150bpm). The component which represents the heart rate signal is selected as the one which contains the "clearest" peak in the passband [0.75Hz - 2.5Hz]. The "clarity" of the peak is assessed using the max-median difference approach (MMD), which is defined as:

$$i_* = \arg\max_i \left[max(|C_i(f)|) - median(|C_i(f)|) \right] \quad (5)$$

where f takes values in the range [0.75Hz - 2.5Hz].

Given a face video, the frequency with the highest power in $|C_{i^*}(f)|$ (obtained from the first 900 frames) is selected to be the first estimated HR. The subsequent estimates are restricted to lie within ± 0.2Hz (± 12bpm) around the previous estimate HR for consistency. If none of the modes have a frequency that satisfies this restriction, the previous estimate of the HR will be used as the next estimate.

5. EXPERIMENTAL RESULTS

We used the PureDL dataset during the experiments [9], [18]. PureDL dataset contains face videos of 8 subjects under 6 different experimental setups:

(i) Steady: the subjects look at the camera from a frontal view and avoid any head motion.

(ii) Talking: the subjects are requested to talk without any additional head motion.

(iii) Slow translation: the subjects move their heads parallel to the camera plane with a speed of about 7% of the head height in pixels.

(iv) Fast translation: the subjects move their heads parallel to the camera plane, with the speed doubled.

(v) Slow rotation: the subjects look at the targets sequentially, which are placed with 35 cm intervals around the camera.

(vi) Fast rotation: the subjects look at the targets sequentially, which are placed with 70 cm intervals around the camera.

The videos are recorded with a rate of 30 fps with 640x480. The ground truth HR data have been captured with a pulse oxiometer with a sampling rate of 60 Hz. Each video in the dataset is decomposed into image files named with the timestamp of the frame and ground truth heart rate values measured during the video recording. For simplicity, we have not resampled the PPG signal from 30 Hz to 60 Hz by applying spline functions to align with the ground truth signal. Instead, we used a mapping to the nearest timestamp. Since we know the timestamp of the last frame of each 1 second boundaries, we have used its timestamp to search for the closest timestamp value in ground truth data.

SETUP	NMD-HR		ICA [3]	
	$\|\mu\|$	σ	$\|\mu\|$	σ
Steady	9.59	23.27	2.71	10.20
Talking	17.41	34.41	40.10	34.05
Slow Translation	5.71	25.22	25.85	36.66
Fast Translation	2.34	9.90	26.76	32.38
Slow Rotation	5.50	25.58	36.83	39.95
Medium Rotation	11.55	26.46	12.10	32.26

Table 1. Experimental results of the proposed NMD-HR method and the ICA based method [3]. The absolute mean $|\mu|$ error and the standard deviation σ of the estimation errors are given in beats per minute.

In the implementation of the ICA-based method, almost all the steps are similar to the steps of the NMD method except, three signals obtained by averaging R, G, and B color components are input to the decomposition step [3]. The ICA method takes these three input signals and estimates three linearly independent components. Then, the [0.75 2.5] Hz intervals in the power spectrum of the three components are evaluated to determine the heart rate estimate.

The experimental results are grouped under the 6 experimental setups described above. Thus, each setup has video recordings of 8 subjects consisting of 30 heart rate estimates leading to 240 estimates in each setup. We analyse the differences between the estimated HR values and the corresponding ground truth values. We used the mean and standard deviation of the error values to evaluate the performance of the NMD method.

In Table 1, the results for the six different setups for the proposed NMD-HR method have been given as well as the results obtained by the ICA-based method [3], [11]. In both methods, the "Talking" setup has the highest average error as seen from the mean values ($|\mu|$) of 17.41 and 40.10 in the NMD and ICA-based methods, respectively. Nevertheless, the NMD-HR method performs better than the ICA method in most of the cases.

6. CONCLUSION

In this work, we presented a method for heart rate estimation from face videos, which uses nonlinear mode decomposition [4] to estimate the periodic cardiac signal. We showed via experimental results that the proposed NMD-HR method gives better results as compared to the independent component analysis method in the literature in terms of lower average absolute error, using the MMD approach, especially in cases containing subject motion, which introduces noise to the color information obtained from the face. A more accurate ROI selection approach, which discards non-skin pixels on the face is expected to improve the accuracy of the results.

7. REFERENCES

[1] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson, "Remote plethysmographic imaging using ambient light.," *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[2] Alrick B Hertzman, "Observations on the finger volume pulse recorded photoelectrically," *Am. J. Physiol.*, vol. 119, pp. 334–335, 1937.

[3] Ming-Zher Poh, Daniel J McDuff, and Rosalind 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.

[4] Dmytro Iatsenko, Peter VE McClintock, and Aneta Stefanovska, "Nonlinear mode decomposition: a noise-robust, adaptive decomposition method," *Physical Review E*, vol. 92, no. 3, pp. 032916, 2015.

[5] Ronny Stricker, Steffen Müller, and Horst-Michael Groß, "Comparison of face segmentation methods for non-contact video-based pulse rate measurement," .

[6] Litong Feng, Lai-Man Po, Xuyuan Xu, Yuming Li, and Ruiyi Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 879–891, 2015.

[7] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal, "Distanceppg: Robust non-contact vital signs monitoring using a camera," *Biomedical optics express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[8] Sungjun Kwon, Hyunseok Kim, and Kwang Suk Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 2174–2177.

[9] Ronny Stricker, Steffen Müller, and Horst-Michael Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 1056–1062.

[10] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4264–4271.

[11] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[12] Hamed Monkaresi, Rafael A Calvo, and Hong Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1153–1160, 2014.

[13] Jean-François Cardoso, "High-order contrasts for independent component analysis," *Neural computation*, vol. 11, no. 1, pp. 157–192, 1999.

[14] Mark van Gastel, Sander Stuijk, and Gerard de Haan, "Motion robust remote-ppg in infrared," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1425–1433, 2015.

[15] Ethan B Blackford and Justin R Estepp, "Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2015, pp. 94172D–94172D.

[16] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*. The Royal Society, 1998, vol. 454, pp. 903–995.

[17] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.

[18] "Pulse rate detection dataset," Oct. 2017, http://www.tu-ilmenau.de/neurob/data-sets/pulse/.