REMOTE PPG BASED VITAL SIGN MEASUREMENT USING ADAPTIVE FACIAL REGIONS

H. Emrah Tasli [1,2] *, Amogh Gudi* [1,2]*, Marten den Uyl* [1]

[1] Vicarious Perception Technologies
[2] Intelligent Systems Lab Amsterdam, University of Amsterdam, The Netherlands

ABSTRACT

This paper proposes a remote photo-plethysmography measurement technique where human skin color variations are analysed for observing human vital signs including but not limited to average heart rate and variation. Remote monitoring of the vital signs could be useful for non-contact physiological and psychological diagnosis. For this purpose, an off-the-self non-invasive video camera is used. Facial appearance modelling is performed for stabilizing color variations in the selected facial region during the signal acquisition stage. The proposed method offers a novel signal processing approach for extracting the periodic component of the raw color signal for the heart rate and variation estimation. To this end, we have collected a ground truth dataset using a PPG instrument attached to the skin of the subject under observation. Objective performance tests show strong correlation with the ground truth values for the estimated heart rate and variation.

***Index Terms*—** Remote PPG, facial expression, active appearance models, heart rate and variability

1. INTRODUCTION

Remote monitoring of vital signs via a conventional video camera by detecting the photo-plethysmographic (PPG) signals have gained attention due to the advantages it offers. The main advantage of being remote is to provide a method for non-invasive and passive monitoring of the vital signs. This could be useful for medical circumstances where physical contact with the patient is not preferred. Moreover, as the video camera could capture multiple persons during a shoot, vital sign monitoring of multiple persons could be possible with the same configuration. Moreover, the method could be applied to detect psychological disorders/anomalies using the long term changes in the vital signs.

PPG based techniques depends on the reflectance properties of human skin for measuring the changes of oxygen saturation in the blood. The underlying principle is that the variations in blood flow due to heart beats would change the volume and the oxygen saturation of the blood in the vessels, and hence the skin reflectance. Sensors that are attached to the body to measure these color changes have been widely available. Moreover, recent methods for remotely measuring these periodic changes have been proposed [1], [2]. Measurement and analysis of the vital signs attain a vital role in monitoring human physical and/or psychological health conditions. Such measurements could well be useful in the sense that such methods could make long time monitoring and predictive analysis possible and hence making the precautionary action be a part of the treatment. Moreover, such methods could be used in the cases where invasive instruments for vital sign analysis are difficult or even impossible to use for various reasons, e.g. attachment/detachment of a sensor to a patient under intensive care could violate the sustained hygiene environment and regulations. The conventional methods of vital sign measurements (including but not limited to heart rate (HR), heart rate variability (HRV), respiratory rate, etc.) require an installation that could violate the hygienic conditions and/or cause discomfort to the patient in need of assistance. Methods utilizing visible light spectrum where no active illumination is applied or required have been presented in order to overcome the aforementioned difficulties. Remote photo-plethysmography (RPPG) measures the small deviations in color values in order to extract the periodic components (heart and/or respiratory rate) of the color signal. By capturing the color information of the skin with a proper sampling rate, these variations could be recorded and analyzed. However, the process of remotely measuring skin color variations for the purpose of vital sign measurement requires the measured skin locations to either stay stable, or be accurately tracked. Since the minor changes in color are essential for accurate vital sign measurement, even the smallest movements could cause a major degradation in the final performance.

This project has been partly funded by STW Valorization Grant

2. PREVIOUS WORK

There has been some prior work dealing with camera based remote estimation of human vital signs. [3] is one of the early studies published to deal with remotely estimating HR from a video. This method comprises of tracking the average brightness of a (manually selected) region of interest (ROI) on the subject's face. The first order derivative of this signal is then passed through a low pass filter, and an auto regressive (AR) spectral analysis is conducted to find dominant frequencies corresponding to the HR. However, this method is only shown to estimate the average heart rates over a 30 seconds period. Similarly, the work presented in [2] estimates HR and respiration rate from the video by spatially averaging the R,G,B channels in a manually selected ROI. Frequency spectrum analysis is performed following the band pass filter operation. Green channel is mainly focused as having the strongest phelthysmographic signal.

On the other hand, the method presented in [4] utilizes a highly sensitive thermal camera to estimate HR. The employed signal processing involves performing fast Fourier transform (FFT) on the signal acquired from a rectangular ROI.

The Eulerian Video Magnification (EVM) study [5] presents a method to amplify subtle changes for obtaining video magnification. The images are spatially decomposed and temporal (band-pass) filtering has been performed in order to amplify changes lying within a pre-set frequency band. Moreover, the flow of blood to the face is visualized on the reconstructed video. Another recently proposed method [6] performs independent component analysis (ICA), a special case of blind source separation, over the three color channels of an RGB video. It has been observed that the second channel of the ICA contains strong plethysmographic information. A face detection method is also included for obtaining the facial region and hence selecting ROI appropriately depending on the detected rectangular face region.

An alternative study described in [7] presents a method for estimating the pulse-rate of a subject in a video by tracking positions of multiple features on the head and performing principal component analysis (PCA) over their trajectories. The pulse rate is then extracted from the component that best corresponds to the frequency of heartbeats on the frequency spectrum. No color information is utilized in this method. A major disadvantage with this method is that strong movements of the head could easily violate the assumptions and create erroneous measurements.

A recent study [8] presents a method of estimating the average pulse-rate by performing PCA on the average R,G,B channel signals from a manually selected ROI on the subject's face over a 30 seconds period. The method in [9] explores the use of color channel signal obtained from a ROI in a video to estimate additional physiological parameters like resperation rate and blood oxygen saturation (SpO2).

The prior methods mostly define a fixed ROI for the signal acquisition. Therefore, the observed face is expected to be kept stable and even small movements are not tolerated. However, with our proposed method, ROIs are adaptively updated for creating a more intuitive and practical scenario.

3. PROPOSED TECHNIQUE

The purpose of this study is to present a method for measuring human vital signs where free head movement is allowed. For the proposed free head movement, the active appearance model (AAM) technique [10] is used to detect facial landmark

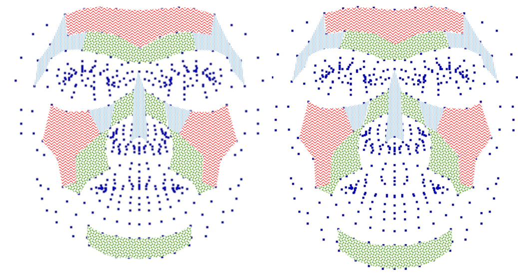

Fig. 1. ROI alternatives on neutral (left) & smiling (right) face

locations. The parametric representation of the facial appearance is computed using the FaceReader [11] framework. The proposed facial landmark localization gives the system the capability of tracking selected region of interest through the video and hence obtaining a robust signal acquisition.

The proposed technique is composed of three main steps. In the first step, face detection [12] and facial landmarks computations [11] are performed. The acquisition of color data is limited to the selected region of interest (ROI) defined by the computed landmark locations. The color signal is processed in the second step for obtaining the periodic component that relates to the vital sign information. In the final step, the required vital signs are generated using the noise free periodic signal. In this study we focus on the average HR and long term variability for illustration purposes. However, the estimated noise free signal could be further used to perform respiration rate interbeat interval (IBI) measurements.

3.1. ROI generation

Active appearance models have been introduced by [10] more than a decade ago for generating a parametric face model in relation to the annotated training dataset. Our facial analysis framework, commercially known as FaceReader [11], uses an improved version of this methodology for obtaining very accurate appearance models for 3D facial landmark detection. The detected landmarks are used to define regions of interest where the color data could be robustly extracted. Figure 1 presents some ROI alternatives and shows how the shapes change while turning from a neutral to smiling face. This region adaptation provides a stability in the signal acquisition phases and proves the advantage of the proposed technique. The regions colored with red are used in our experiments during the PPG signal acquisition.

Prior methods [2], [3], [6] mostly utilize a fixed ROI definition. Therefore the subject is expected to keep stable during the measurement. With our adaptive ROI generation this limitation is relaxed and the subject is free to move and talk during the data acquisition stage.

3.2. Periodic component extraction

In the second step of the proposed technique, collected color data is further processed to remove noise and extract the periodic signal component. For that purpose the green channel of the color signal is averaged on the selected ROI. All three RGB color channels contain plethysmographic information. However, green channel features the strongest component in regards to the amount of oxygen absorbed in the blood [2]. For that purpose, averaged green channel is used as the raw data for periodic component extraction.

We have introduced an intermediate step during the color averaging process where the outlier pixels in the selected ROI are removed from the mean calculation. The removal is performed using the following formula (1) where the pixels $p(x, y)$ in the selected region R with color value difference, $\Delta(x, y)$ with respect to the region average $\mu(R)$ is higher than 3 times the standard deviation $\sigma(R)$ of the selected ROI.

$$p(x, y) \text{ is an outlier, if } \Delta(x, y) \geq 3 * \sigma(R) \tag{1}$$

where pixel color difference $\Delta(x, y)$ is calculated in the 3 dimensional Euclidean space with each dimension corresponding to the RGB color values (2) .

$$\Delta(x, y) = \|p(x, y) - \mu(R)\|^2 \tag{2}$$

In order to utilize the plethysmographic information more efficiently, opponent and normalized color spaces are also investigated. Moreover, independent component analysis (ICA) of the RGB signal has also been experimented as proposed in [6]. However, no significant improvement has been observed under these variations. Therefore, the averaged green channel is utilized as the raw data in the rest of the analysis.

The averaged color values are accumulated for a fixed amount of time interval before conducting the temporal analysis. For a given time window of the sampled signal, the temporal mean of the averaged green color is subtracted in order to obtain a zero mean signal for further analysis. The obtained DC free average color signal, $s(t)$, ideally contains only the fluctuations due to the blood flow changes altering the skin reflectance properties. However, due to the inhomogeneity in the light and skin reflectance properties, the DC removed raw data $s(t)$ still contains strong trends and noise as observed in Figure 2 printed with green color.

The removal of the trend in the signal is vital for obtaining the periodic component and has been conducted in a three step approach. In the first step, the conventional detrending technique is applied to the raw signal [13]. In the second step of the proposed method, we intend to remove the high frequency (noise) component of the detrended signal as proposed in [6]. For that purpose, we subtract the detrended signal $D(s(t))$ from the raw signal $s(t)$ and obtain the low pass equivalent, $s_{LP}(t)$, of the raw signal $s(t)$. The low pass signal is observed to contain smooth periodic components as shown in Figure 2. Finally, another detrending operation is conducted to generate the detrended low pass signal $D(s_{LP}(t)$.

$$D(s_{LP}(t)) = D(s(t) - D(s(t))) \tag{3}$$

where $D(.)$ is the detrending operator.

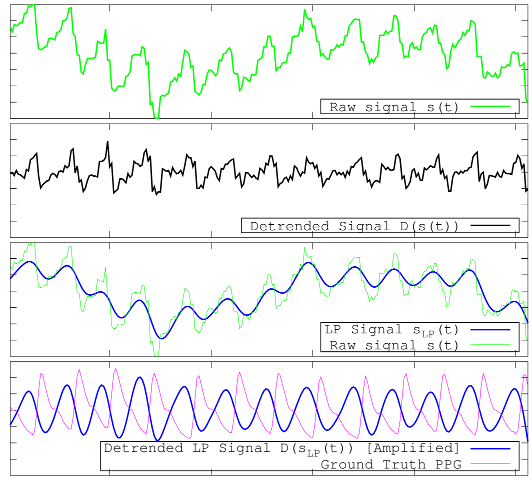

Fig. 2. Proposed three step approach for obtaining periodic component in the raw signal. Strong correlation (up to a phase difference) with the ground truth PPG signal is observed

The initial processing of the raw signal plays a vital role in the succeeding time and frequency domain computations. We have observed great increase in performance with the proposed initial processing of the noisy green color signal. The main improvement is observed by removing the detrended signal from the raw signal. We have used the detrending method proposed by [13] where a smoothness prior is presented to facilitate the estimation of HR variability (HRV).

The performance of the proposed signal processing could be visualized in Figure 2 where the obtained noise free periodic component is presented together with the ground truth PPG signal obtained through a PPG instrument attached to the fingertip of the subject under observation. The phase difference between the ground truth and estimated signal could be accounted for the time difference of the blood circulation in the face and fingertip.

3.3. Vital sign measurement

The body temperature, heart rate, blood pressure, and respiratory rate are considered as the vital signs and are crucial in determining one's physical and psychological state. The monitoring and variability in these signals could also be strong medical indications for diagnostic purposes. In this study, we focus on the heart rate and its long term variability.

The measurement of the HR is conducted using both time and frequency domain analysis where the HR is extracted from the average frequency of the PPG signal measured over a time window. The main periodic component could either be extracted using a peak detection over the noise free time signal or the frequency domain analysis could be used to find the strongest harmonic component. Both are evaluated and observed to agree in high signal to noise ratio cases. Figure 3 shows the estimated HR from the time analysis with respect to the ground truth PPG signal. In our measurements a strong correlation is observed between the estimated and real HR.

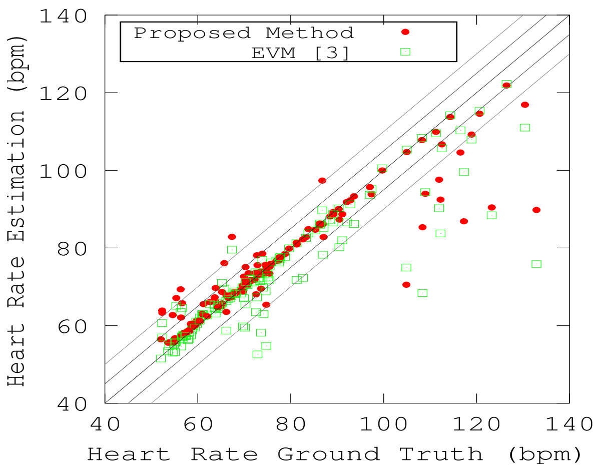

Fig. 3. Ground truth vs estimated hear rate measurements for the proposed and EVM method [5]. The advantage of the proposed method is observed under varying heart rates.

4. EXPERIMENTS

In order to validate the performance of our proposed method, we conducted experiments where the estimated remote PPG signal is compared with a ground truth PPG signal obtained from a contact PPG sensor.

To our best knowledge, there is no publicly available dataset with ground truth PPG measurements that could be used to verify the RPPG methods. Therefore, we have created a new dataset for evaluating RPPG measurements where the ground truth PPG is provided and synced with the captured video. We intend to release the dataset for public reach for research purposes.

A total of 10 subjects aged between 20-35 are included in the experiments. Two videos of resolution 720x1280 are recorded per subject at 30 fps for an average duration of 90 sec. In the first video, the subjects were simply instructed to sit naturally facing the camera. In the second video, the subjects were instructed to perform physical exercises (running) before sitting in front of the camera. Simultaneously, the subjects' pulse waveforms were recorded using a CMS-50 Pulse Oximeter placed on the subject's fingertip. In the first set of videos, the subjects' heart rates were stable in the range of 50-90 beats per minute (bpm). In the second set of videos, the heart rates were observed to be above 120 bmp, and slowly decrease to the subject's resting heart rate (as the subject's state of fatigue is alleviated). The results of the actual vs estimated HR measurements are presented in Figure 3. For each video in the dataset, we have segmented 6 non overlapping time windows over which the measurements are performed. For each subject, 12 HR measurements are presented in Figure 3 for the proposed and the EVM method [5]. The HR values for the EVM method is computed by finding average period in the selected window of the inversed FFT signal. The correlation between the HR obtained from the proposed method and the ground truth is shown in Figure 4.

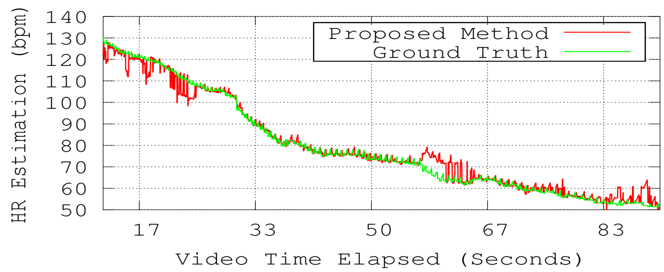

Fig. 4. Heart rate varies from 135 to 50 in a minute after a short running exercise. RPPG estimation follows the heart rate during the measurement.

It is observed that the accuracy of the estimated HR depends greatly on the smoothness prior of the detrending operation. A selection of a lower prior results in a more accurate measurement for the heart rates between 50-90 bpm, while the higher heart rates being underestimated. Similarly, the opposite holds for the higher valued priors. The results presented in Figure 3 and 4 were obtained for a fixed smoothness prior. The errors for the proposed/EVM method are computed as follows: (a) Average error for steady HR videos is 2.9/4.3 with a standard deviation of 4.4/6.8. (b) Average error for all videos is 4.2/5.6 with a standard deviation of 7.7/10.1.

Moreover, the proposed technique allows estimation of changes in the IBI during the HR measurement. This could be realized with the proposed time domain analysis after the detrending operation. This functionality is not possible with the prior art techniques where only average HR could be measured after the band pass filtering of the signal.

5. CONCLUSION

In this paper, we propose a novel RPPG method for average heart rate and variability measurements. The proposed facial landmark tracking using FaceReader makes the method robust under head pose and expression changes. This approach, combined with the preprocessing step where the local trend of the signal is extracted, provides accurate measurements. The IBI measurement capability is possible only within the proposed time domain approach. A ground truth heart rate measurement dataset, which seems to be missing in the research community, is collected for performance evaluation and will be made publicly available for research purposes.

The main limitation of the method is observed under poor lighting conditions. Head movements are mostly compensated with the help of FaceReader; however, poor lighting could deviate the performance of the proposed technique. As a future direction, the respiratory rate measurement using the variability in the IBI is planned. Additionally, a dynamic method for adapting the smoothness parameter of the detrending operation could be useful. Moreover, the effect of heart rate variability could further be analyzed for evaluating physical and psychological experiments.

6. REFERENCES

[1] Sijung Hu, Jia Zheng, V. Chouliaras, and R. Summers, "Feasibility of imaging photoplethysmography," in *International Conference on BioMedical Engineering and Informatics*, 2008.

[2] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson, "Remote plethysmographic imaging using ambient light.," *Optics express*, 2008.

[3] Chihiro Takano and Yuji Ohta, "Heart rate measurement based on a time-lapse image," *Medical Engineering and Physics*, 2007.

[4] M. Garbey, Nanfei Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE Transactions on Biomedical Engineering*, 2007.

[5] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2012.

[6] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, 2010.

[7] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013.

[8] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam; a non-contact method for evaluating cardiac activity," in *Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011.

[9] C. Scully, Jinseok Lee, J. Meyer, A.M. Gorbach, D. Granquist-Fraser, Y. Mendelson, and K.H. Chon, "Physiological parameter monitoring from optical recordings with a mobile phone," *Biomedical Engineering, IEEE Transactions on*, 2012.

[10] G.J. Edwards, C.J. Taylor, and T.F. Cootes, "Interpreting face images using active appearance models," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

[11] M. J. den Uyl and H. van Kuilenburg, "The facereader: Online facial expression recognition," in *International Conference on Methods and Techniques in Behavioural Research*, 2005.

[12] Paul Viola and Michael Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, 2001.

[13] M.P. Tarvainen, P.O. Ranta-aho, and P.A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, 2002.