This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2018.2842639, IEEE Signal Processing Letters

Remote Pulse Rate Measurement from Near-Infrared Videos

Sang Bae Park, Gyehyun Kim, Hyun Jae Baek, Jong Hee Han and Joon Ho Kim

***Abstract*—— In this study, a remote photoplethysmography (rPPG) using a near-infrared (NIR) camera image has been investigated to estimate the Pulse Rate in a dark environment. We propose a new face segmentation method using a skin boundary filter and pixel-based shape mask filter to improve the rPPG performance. The proposed algorithms were evaluated by applying them to the skin of the whole face and the skin of the lower part of the face. The mean error rates of the estimated Pulse Rate in the lower part of the face and the whole face area was approximately 2.9% and 9.9%, respectively. Additionally, new dynamic noise cancellation techniques were proposed using a noise cancellation filter, the Kalman filter, and an adaptive filter to remove motion-related noise. Consequently, the mean error rate for 12 subjects drops to 2.0% when the proposed methods are combined. The rPPG signal showed robust results in motion artifact.**

***Index Terms*— remote photoplethysmography, pulse rate, infrared ray, camera, recognition, facial skin**

I. Introduction

Pulse rate (PR) is one of basic indicators of health in human. Recently, the number of people who are self-managing their Pulse Rate is increasing rapidly [1]. During driving, PR monitoring is also important to prevent a secondary damage caused by angina and myocardial infarction in the driver. Several studies reported that Pulse Rate management could help in predicting cardiovascular diseases and preventing secondary accidents [2], [3]. The motor industry is attempting to adapt such a technology to a car for the drivers' safety. For example, a departed driver rescue and exit maneuver (DDREM) system moves a car to a safe area if the driver experiences health problems. In most cases of driving conditions at night, the illumination is very low. Therefore, in this study, we propose a method to measure the Pulse Rate of a user in a dark environment using a near-infrared (NIR) camera. The PR monitoring using a camera increases the degree of freedom of the user by measuring the PR in a non-contact manner. Previously, most studies on camera-based rPPG used a color camera without additional light sources because they assumed normal illumination conditions [4]-[6]. However, in a dark environment, an NIR camera with infrared source is required to increase the amount of light reflected from the face without interfering the eyes. Some studies have shown that PR estimation performance is improved by using a high-speed NIR camera in low-illumination conditions [8], [9]. Some studies used multiple infrared sources to improve accuracy in motion situation [10].

In this study, we propose a method to measure PR using a low-cost NIR camera, which is widely available, to improve the accuracy in a dark environment. Further, we propose an algorithm that can achieve the same or better results without using the additional light sources and sensors proposed in previous studies. The low-cost NIR camera is used because, if the light source is increased and the high-resolution camera is used instead, the development cost increases and the development of the miniaturization module is affected. In previous studies, PPG signals have been estimated from various regions of the face such as the nose and forehead. However, the area where the rPPG is well estimated is still unclear. In most studies using NIR, adaptive filters were applied to increase the accuracy in PR estimation, but it is still insufficient to be used in a real environment [13-17]. According to ANSI/AAMI EC13:2002 [18], the error in beat per minutes (bpm) should be less than 5bpm or 10%, whichever is greater. Therefore, we propose a new face segmentation method using a skin boundary filter and pixel-based shape mask filter to improve the rPPG performance. Further, we herein propose a new dynamic noise cancellation technique using a noise cancellation filter, the Kalman filter, and an adaptive filter. These two algorithms were applied to the skin of the whole face area and the lower part of the face.

In this study, the PR was estimated by measuring the changes in NIR light absorption due to blood flow in the face. Hemoglobin in blood has different optical responses depending on the amount of oxygen. Thus, the near-infrared rays are reflected on the face and the images are acquired continuously. PPG can be obtained from hemoglobin changes caused by the relaxation/contraction of blood vessels due to heartbeat, such that PR monitoring is possible [21], [22]. As shown in Fig. 2, this algorithm is designed in five steps.

1) Capture an image of a face from an infrared (IR) camera.
2) Set the lower part of the face as the area of interest and extract the skin of the lower part.
3) Calculate the absorbance of light reflected from the skin portion in the region of interest (ROI).
4) Estimate rPPG using the calculated results.
5) Estimate HR using FFT after removing motion noise from rPPG.

The research work disclosed in this publication is funded by Samsung Research. (Corresponding author: Joon Ho Kim). The authors are with the Samsung Electronics Seoul R&D Campus 33 Seongchon-gil, Seocho-gu, Seoul 06765, korea (e-mail: sangbae81@gmail.com;gyehyun.kim@samsung.com, hyunjae.baek@samsung.com;jh78.han@samsung.com; mythos.kim@samsung.com).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2018.2842639, IEEE Signal Processing Letters

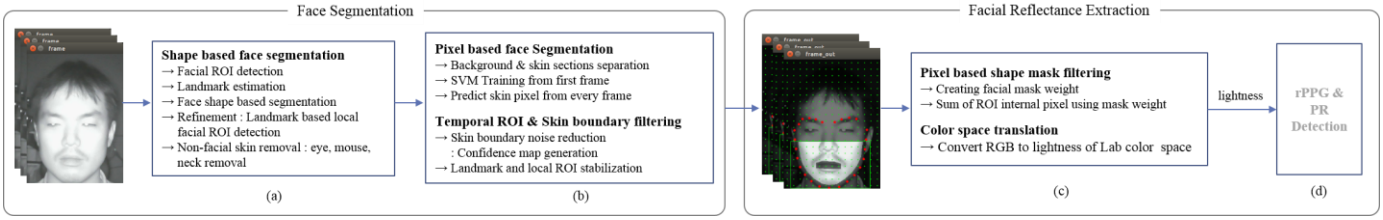

Fig. 2. Proposed architecture. Fig. 2, (a) and (b), there is disclosed a system and method of performing facial recognition and skin segmentation from IR image data. Fig. 2, (c) and (d), There is disclosed a system and method of performing facial reflectance estimation from pixel data. An image input in real time is performed every frame and estimates the change in the skin within the lower part of the face. There is a lot of noise at the boundary of the face. In this study, we applied the confidence map.

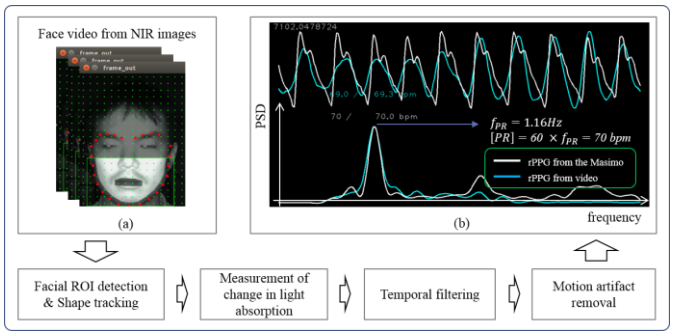

Fig. 1. Proposed framework for PR extraction from facial videos using near-infrared ray in realistic human computer interaction (HCI) situations.

II. Proposed Method

A. Equipment and Data Collection

As shown in Fig. 3, rPPG is detected from the face using the camera and the PR is estimated using FFT [24]. The length of signal for FFT operation is 512 and uses the moving window for real-time data. The estimated PR is compared to the reference PR. Subsequently, the reliability of the result is evaluated and the algorithm is analyzed.

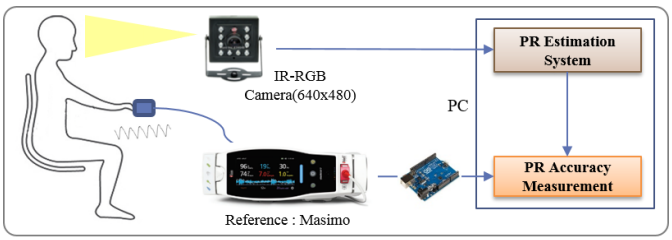

Fig. 3. Experiment setup of this study

"Masimo Radical-7" was used as a reference device (ground truth) to evaluate the reliability and accuracy of the estimated PR. The reference device records the PPG and measurement time at 100 Hz. The measurement device was the "ELP USB FHD04H-BL36IR" and measurement was performed at 30 Hz. And this model uses single wavelength of about 800~1100nm. All videos are recorded in 24-bit RGB color format at 30 frames per second using IR-cut filters removed from the lens [25]. Synchronization of the two instruments was based on measurement time. The images were obtained at 640 × 480 and 480 × 640 for 2 min. Further, stationary images were obtained from 12 subjects. Subjects were measured in a relaxed state with some facial muscles movable. Skin tone is all 'V' type and all subjects are healthy male without cardiovascular disease [26]. The photograph was taken in the dark laboratory and the NIR rays were reflected on the face and subsequently captured by the camera. Additionally, the subject gazed at the front without any motion. Further, the distance between the camera and the subject was approximately 80 cm. Hereafter, the lower part of the face area is called the local ROI.

B. Face Segmentation

As shown in Fig. 1, estimating the raw signal of the rPPG from the face consists of two parts. In the face segmentation phase, the NIR camera has R, G, and B color spaces with IR filters removed from the lens. In the facial reflectance extraction stage, we used the L channel in the Lab color model that synthesized the R, G and B channel. Since only single infrared source is used and R, G and B sensor data imply the same reflection patterns of the single source, the L channel exhibiting a luminance characteristic is used for processing. As shown in Fig. 1. (a) and (b) are designed to separate the face part from the background in the image. In other words, only the face part is extracted by separating the face and the background in the image. In this study, we used the Dlib library [27] for face detection, tracking, and landmark estimation. The landmark was used to remove the eyes, mouth, and neck from the facial image. It was better to remove each area when moving or changing facial expressions. Support vector machine was used to estimate the pixels of the skin region in the face [28] and separate the skin and the background in the ROI domain.

$$inside_{mask} = distance\_transform \quad (1)$$

$$dist_{max} = \log 10.5 - \log 0.5 \quad (2)$$

$$inside_{mask} = \log(inside_{mask} + 0.5) - \log 0.5 \quad (3)$$

$$inside_{mask} = \begin{cases} dist_{max}, & inside_{mask} > dist_{max} \\ inside_{mask}, & inside_{mask} \le dist_{max} \end{cases} \quad (4)$$

Further, we applied the confidence map to the skin pixels to reduce the edge noise of the estimated skin area. In the equations of "(1)"–"(9)", the confidence maps are created in order to apply lower weight to the pixels which are closer to the boundary. As shown in "(1)", the distance transform finds the shortest Euclidean distance from the pixels in the skin region to the boundary. Based on the distance, the first weight value ($inside_{mask}$) is calculated by using the log scale "(3)". The value of the log function in "(2)" and "(3)" is determined heuristically through experiments. We added the condition of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2018.2842639, IEEE Signal Processing Letters

"(4)" to saturate the $inside_{mask}$ if a pixel is enough far from the boundary. The confidence map is calculated within the skin area which is extracted from facial landmarks. The local ROI was determined by using landmarks as the green rectangle shown in Fig.2. Smoothing to the local ROI position was applied to reduce the local ROI shake. The $dist_{max}$ is a threshold for distance to saturate the confidence level if the distance of a pixel is greater or equals to it. Therefore, as shown in "5", we make a mask value by a ratio of the $inside_{mask}$ and $dist_{max}$. Equation "6" is a method for creating a confidence map which applied a recursive update model. When there is a big change in mask, we updated the confidence map with more weight value (0.5) as shown in "7" according to the difference between a current mask and the previous mask "8".

$$result_{mask} = {inside_{mask}}/{dist_{max}} \tag{5}$$

$$mask_w(n) = \begin{cases} result_{mask}, & n = 1 \\ mask_w(n-1) \times (1 - mask_{w_{rate}}) + \\ (result_{mask} \times mask_{w\_rate}, & n > 1 \end{cases} \tag{6}$$

$$mask_{w\_rate} = \begin{cases} 0.5, & maskDiff > 0.4 \\ 0.05, & maskDiff \le 0.4 \end{cases} \tag{7}$$

$$maskDiff = E[mask_w(n-1)] - E[result_{mask}] \tag{8}$$

C. Facial Reflectance Extraction

The reflectance is calculated only in the local ROI region. As shown in "(10)", new R, G and B value were calculated by combining the confidence map with the input image. This calculation process is performed at every frame. After then, the lab color projection was applied to estimate a facial reflectivity value, luminance value (L).

$$[R,\ G,\ B] = {\sum(frame \times mask_w)}/{\sum mask_w} \tag{10}$$

D. Motion Artifact Removal and Remote PR Extraction

As shown in Fig. 4, the estimated reflectivity data from the skin is stored in the stack.

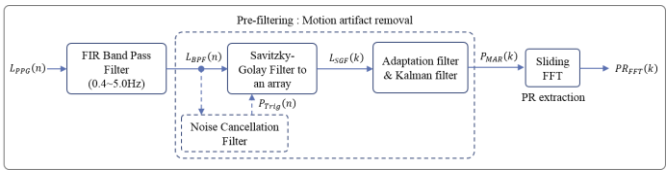

Fig. 4. Proposed rPPG and PR extraction

The values stored in the stack are estimated using the FIR bandpass filter (BPF, cut-off frequency of 0.4–5.0 Hz, 8th order) [21], [29]. Even small motion such as muscle movement can degrade the rPPG signal quality. As shown in Fig. 5, Case 1 corresponds to the case where only the facial muscle is slightly moved. Case 2 corresponds to the case where the face is turned 45 degrees to the left and right. Therefore, this study designed a motion canceling filter by combining the Kalman filter and the adaptive filter based motion artifacts filter (MAF) [15]. In addition, Noise cancellation filters was designed with a noise filter and smoothing filters to increase signal quality.

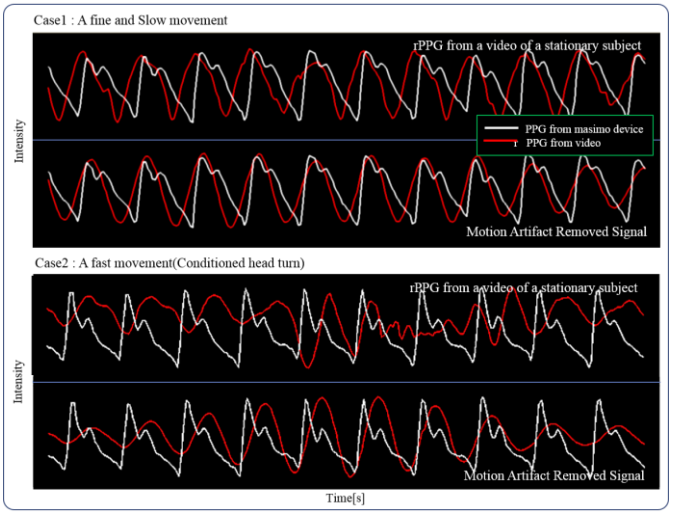

Fig. 6. Comparison between motion artifacts removed signal from the rPPG from video and the obtained general rPPG.

As shown in Fig. 6, a median filter was used to remove noise for rapidly changing signals. In addition, smoothed signals are generated using cubic b-spline and Savitzky–Golay filters (SGF) ($S_f(n)$). Subsequently, as shown in "(12)," the smoothing signal and the median filtering signal ($M_{MF}(n)$) are multiplied to obtain the wave-form characteristic ($P_f(n)$) [23]. Consequently, the primary component signal becomes larger ($P_B(n)$) and the signal generated by the motion becomes smaller ($P_A(n)$). Subsequently, the non-principal component signal is removed using "(13)." The noise signal in the high frequency of the principal component signal is smoothed using SGF after the noise cancellation filtering. The result of the signal midpoint is used as an input to the previously studied motion artifact removal logic [15], [18], [19].

$$P_f(n) = M_{MF}(n) \times S_f(n) \tag{12}$$

$$P_{Trig}(n) = \begin{cases} P_A(n), & P(n) \times 2 < 0.1 \\ P_B(n), & P(n)/2 \ \ge 2.0 \end{cases} \tag{13}$$

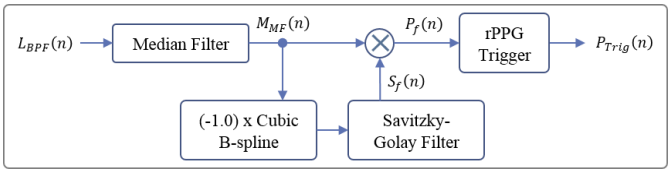

Fig. 5. Proposed noise cancellation filter

After motion artifact removal filtering, the pulse signal is converted to the frequency domain and its power spectral density (PSD) distribution is estimated using Welch's method [23]. The PSD estimates the signal's power distribution as a function of frequency. We use the frequency with the maximal power response as the PR frequency f_{PR} (Fig. 2. top-right); the average PR measured from the input video is computed [7], [8].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2018.2842639, IEEE Signal Processing Letters

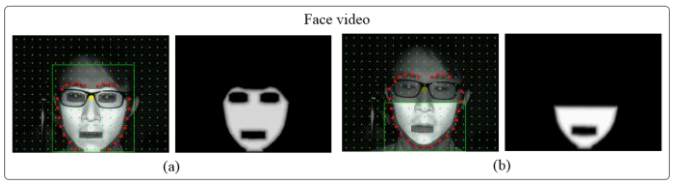

Fig. 7. Remote Pulse Rate measurement from face videos under realistic situations. Two regions of interest were tested for the stationary date-set. (a) full-face recognition, And (b), local ROI recognition. The right images of (a) and (b) are the result of applying the confidence map.

III. Result

A. Accuracy Measurement

The measurements are based on the comparison between the calculated PR (PR_{video}) and the reference PR (PR_{gt}). The measured error (PR_{error}) can be calculated as follows: $HR_{error} = PR_{video} - PR_{gt}$. Thereafter, different accuracy measurements can be obtained [14]:

1) Mean of the measured error: $M_e = \sum PR_{error} / n$
2) Standard deviation of the measured error:

$$SD_e = \sqrt{\sum [(M_e - PR_{error})^2] / (n-1)}$$

3) Root mean square error: $RMSE_e = \sqrt{\sum {PR_{error}}^2 / n}$
4) Mean of error-rate percentage:

$$M_{eRate} = \frac{1}{n} \sum PR_{error} / PR_{masimo}$$

B. Evaluation of Regions of Interest using Stationary Data

We analyzed the results after removing the eyes and mouth areas. We filled in inner position of them with '0' to get the effect from the skin image. As shown in Fig. 7, (a) analyzed the skin information of the entire face excluding the forehead area, and (b) analyzed the results using the skin information of the local ROI. Further, we evaluated the difference in the result of applying the proposed algorithm. As shown in TABLE 1, using the local ROI is better than using the whole facial skin. The numerical results show that the mean error rates of the PR estimated in the whole face area and in the local ROI were approximately 9.9% and 2.9%, respectively; i.e., the error rate decreased by approximately 3.41 times.

TABLE 1
Error measurement rates for different regions of interest, derived from the stationary dataset. Applying only BPF and SGF.

PR Extraction Method		M_e (bpm)	SD_e (bpm)	$RMSE$ (bpm)	M_{eRate} (%)
Facial skin	Average	7.8	11.3	14.9	9.9
	Standard Deviation	7.8	8.3	10.2	10.7
The lower part of the face	Average	2.3	5.5	6.2	2.9
	Standard Deviation	2.2	5.4	5.6	2.9

C. Comparison of Motion Artifact Removal Results

As shown in TABLE 2, applying the motion artifact removal algorithm proposed in this study reduces the average error rate of PR by approximately 0.9%. As shown in Fig. 5 the rPPG signal is improved when using the proposed algorithm. The evaluation data set include natural movements of facial muscle and slow head movement when the subject were in a relaxed state.

TABLE 2.
Error measurement rates for different motion artifact removal, derived from the stationary dataset.

PR Extraction Method (The lower part of the face)		M_e (bpm)	SD_e (bpm)	$RMSE$ (bpm)	M_{eRate} (%)
Does not apply	Average	2.3	5.5	6.2	2.9
	Standard Deviation	2.2	5.4	5.6	2.9
Apply	Average	1.7	3.6	4.0	2.0
	Standard Deviation	1.5	3.3	3.5	1.6

The correlation coefficient of this algorithm showed approximately 0.87 ($PR_{video}(bpm)$ vs. $PR_{gt}(bpm)$) [13]. The data for analyzing the correlation coefficient is the total PR data for 2 min detected in 12 subjects. As shown in Fig. 8, the circle is an error case that cannot be removed by the motion artifact removal algorithm.

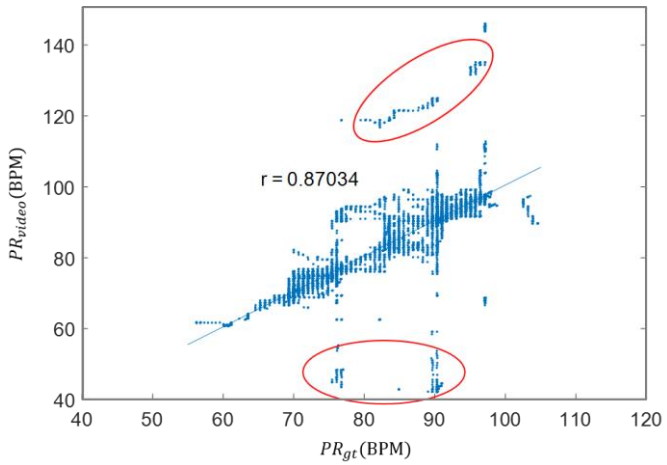

Fig. 8. Scatter plot comparing PR_{video} measured by our method with the ground truth (PR_{gt}) from PPG of Masimo device

IV. Conclusion & Future Work

This study was developed to measure PR in a non-contact manner in a dark environment. The measured results in the lower part showed better performance. In addition, we improved the rPPG signal using the existing method [15], [18], [19] and the noise cancellation filter for the noise signal generated by the user's motion artifact and change in facial expression. Further, the possibility of measuring PR using NIR light in dark situations (driver, sleeping, newborn, etc.) was confirmed. We are still processing the video recorded in the PC in real time; however, we are attempting to improve the algorithm to monitor in real time from a long-term viewpoint. Additionally, we plan to verify the algorithm in various environments and improve the dynamic noise canceling algorithm to the user's motion.

REFERENCES

[1] "mHealth (mobile health) industry market size projection from 2012 to 2020 (in billion U.S. dollars)", The Statistics Portal, https://www.statista.com/statistics/295771/mhealth-global-market-size/

[2] Je-Min Song, Gye-Hwan Jin, Sung-Bo Seo and Keun Ho Ryu, "Design and Implementation of a Prediction System for Cardiovascular Diseases using PPG," Journal of the Korean Society of Radiology., vol. 5, pp.19-25, Feb. 2011.

[3] Johanna A A G Damen., Hooft L., Schuit E., Debray TPA., Collins G., Tzoulaki I., Lassale CM., Siontis GCM., Chiocchia V., Cumming-Roberts C., Maia Schlussel M., Gerry S., Black J., Heus P., van der Schouw YT., Peelen LM. and Moons KGM., "Prediction models for cardiovascular disease risk in the general population: systematic review," Formerly the British Medical Journal., vol. 353, May 2016.

[4] Wenjin Wang, Sander Stuijk, and Gerard de Haan, "A novel algorithm for remote photoplethysmography : spatial subspace rotation," IEEE Trans. Biomed. Eng., vol. 63, no. 9, pp. 1974-1984, Sep. 2016.

[5] CPRistian Galea and Reuben A. Farrugia, "Forensic face photo-sketch recognition using a deep learning-based architecture," IEEE Signal Processing Letters, vol. 24, no. 11, pp. 1586 -1590, Nov. 2017.

[6] X. Li et al, "Remote Pulse Rate measurement from face videos under realistic situations," in Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014, pp. 4264–4271, June 2014.

[7] M. Garbey et al., "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," in Biomedical Engineering, pp. 1418–1426, Aug. 2007.

[8] Tomas McKelvey, "Remote Pulse Rate extraction from near infrared videos," Department of Signals and Systems Division of Signal Processing and Biomedical Engineering Chalmers University of Technology Gothenburg, Sweden 2016, Master's thesis 2016.

[9] Hu M., Zhai G., Li D., Fan Y., Duan H., Zhu W. and Yang X., "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and Pulse Rates under sleep situation," PLoS One, Jan. 2018.

[10] van Gastel, M., Stuijk, S., & de Haan, G., "Motion robust remote-PPG in infrared," IEEE Transactions on Biomedical Engineering, vol. 62, no. 5, pp.1425-1433, May 2015.

[11] Anna Lindqvist and Markus Lindelow, "Remote Pulse Rate extraction from near infrared videos," Master's thesis in Biomedical Engineering, Control and Mechatronics, Master's thesis 2016.

[12] Ilaria Bosi, Chiara Cogerino and Marco Bazzani, "Real-time monitoring of Pulse Rate by processing of near infrared generated," SMART 2017: The Sixth International Conference on Smart Cities, Systems, Devices and Technologies, June 2017.

[13] M. Poh, D. McDuff, and R. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Optics Express, vol. 18, pp. 10762-10774, May 2010.

[14] Byeong-ro Lee and Ju-woo Lee, "Design of motion artifacts of PPG Signal based on Kalman filter and adaptive filter," J. Korea Inst. Inf. Commun. Eng., vol. 18, no. 4, pp. 986-991, Apr. 2014.

[15] Hamed Azami, Karim Mohammadi, and Behzad Bozorgtabar, "An improved signal segmentation using moving average and Savitzky-Golay Filter," Journal of Signal and Information Processing, vol. 3, no. 1, pp. 39-44, Feb. 2012.

[16] Fernando Andreotti, Alexander Trumpp, Hagen Malberg, and Sebastian Zaunseder, "Improved Pulse Rate detection for camera-based photoplethysmography by means of Kalman hiltering," International Conference on Electronics and Nanotechnology, pp. 21-24, Apr. 2015

[17] Ram M.R., Madhav K.V. and Krishna E.H, "A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive Filter" IEEE Transactions on Instrumentation and Measurement, vol. 61, pp. 1445 - 1457, May 2012.

[18] Association for the Advancement of Medical Instrumentation, "Cardiac monitors, heart rate meters, and alarms", Approved 6 May 2002 by American National Standards Institute, Inc.

[19] Fernando Andreotti, Alexander Trumpp, Hagen Malberg and Sebastian Zaunseder, "Improved Pulse Rate detection for camera-based photoplethysmography by means of Kalman filtering," International Conference on Electronics and Nanotechnology, vol. 35, pp. 428-433, Apr. 2015

[20] Kwan Woo Lee, Hyung Gil Hong and Kang Ryoung Park, "Fuzzy system-based fear estimation based on the symmetrical characteristics of face and facial feature points," Expert Syst. Appl. vol. 81, pp. 108-133, June 2017.

[21] W. Verkruysse, L. Svaasand, and J. Nelson. "Remote plethysmographic imaging using ambient light," Optics Express, vol.16 (26), pp. 21434–21445, Dec. 2008.

[22] Mohammed Faisal Lutfi, "Effect of hemoglobin concentration on Pulse Rate variability," International Journal of Biological & Pharmaceutical Research, vol. 2(5), pp. 127-131, January 2011

[23] Xu Long and Nam-Ho Kim, "De-noising method using nonlinear filter algorithm in mixed noise environment," J. Korea Inst. Inf. Commun. Eng., vol. 18, no. 9, pp. 2265-2271, Sep. 2014

[24] P.Welch. "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," IEEE Trans. on Audio and Electroacoustics, vol. AU-15, pp. 70-73, June 1967.

[25] "1/3-inch CMOS Digital Image Sensor," AR0330CM, https://www.onsemi.cn/PowerSolutions/document/AR0330CM-D.PDF

[26] T. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," Archives Dermatology, vol. 124, no. 6, pp. 869–871, 1988.

[27] http://dlib.net/

[28] Shah, Parin M., "Face detection from images using support vector machine," Master's Projects. 321. http://scholarworks.sjsu.edu/etd_projects/321

[29] M. Tarvainen, P. Ranta-aho, and P. Karjalainen, "An advanced detrending method with application to PRV analysis," IEEE Trans. Biomed. Eng., vol. 49(2), pp. 172–175, Feb. 2002.