DYNAMIC ROI BASED ON K-MEANS FOR REMOTE PHOTOPLETHYSMOGRAPHY

Litong Feng[1], Lai-Man Po[1], Xuyuan Xu[1], Yuming Li[1], Chun-Ho Cheung[2], Kwok-Wai Cheung[3], Fang Yuan[1]

1. Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
2. Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China
3. Department of Computer Science, Chu Hai College of Higher Education, Hong Kong SAR, China

ABSTRACT

Remote imaging photoplethysmography (RIPPG) can achieve contactless human vital signs monitoring. Though the remote operation mode brings a great convenience for RIPPG applications, the RIPPG signal quality is limited by the remote nature. Improving the RIPPG signal quality becomes an essential task in the clinical application of RIPPG. Since the region of interest (ROI) of the RIPPG transforms from a point to an area, there is a new approach to improving the RIPPG signal quality through refining the ROI. In this paper, we propose a dynamic ROI for RIPPG, which can automatically select the skin regions corresponding to good quality RIPPG signals. First, a fixed ROI is divided into non-overlapped blocks. Then two features are proposed to perform no-reference quality assessment for RIPPG signals from different blocks. After that, K-means clustering operates in a two dimensional feature space. A dynamic ROI can be selected for a video segment based on the clustering result, updated every two seconds. Nineteen healthy subjects were enrolled to test the proposed ROI selection method on both the facial region and the palmar region. Experimental results of heart rate measurement show that the proposed dynamic ROI method for RIPPG can effectively improve the RIPPG signal quality, compared with the state-of-the-art ROI methods for RIPPG.

Index Terms— No-reference quality assessment (NRQA), photoplethysmography (PPG), region of interest (ROI), remote imaging.

1. INTRODUCTION

Photoplethysmography (PPG) is an electro-optic technique which can non-invasively measure the tissue blood volume pulse (BVP) underneath the skin [1]. The conventional contact PPG (CPPG) needs dedicated light sources and contact probes to detect the reflection/transmission light intensity variation from the tissue due to BVP. Recently, the emerging remote imaging PPG (RIPPG) technique can get rid of contact probes and dedicated light sources. Fig.1 illustrates the working principles of the CPPG and the RIPPG.
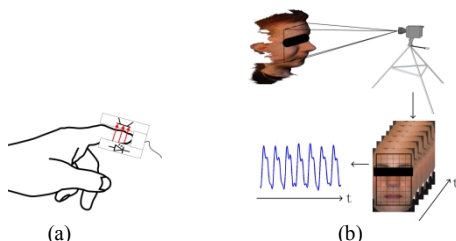

Fig. 1. The illustration of the working principle of PPG: (a) CPPG, (b) RIPPG.

The cardiac pulsation information can be accessed remotely by the RIPPG, and only a digital camera and ambient light are needed [2], [3]. For RIPPG, ambient light is used as the light source and the digital camera acts as a photoelectric converter. The digital camera focuses on a region of interest (ROI) of the human skin. A sequence of the skin images is recorded in the video format. Hemoglobin in blood can absorb light. And the hemoglobin concentration varies quasi-periodically with BVP. Hence BVP underneath the skin surface can modulate the light absorption by the skin along with cardiac cycles, expressing as tiny color variation of the skin. This tiny color variation can be detected by digital cameras. From the frame sequence of the video, the PPG signal can be accessed by video/signal processing techniques

The development from the CPPG to the RIPPG brings new challenges. First, the signal-to-noise ratio (SNR) of a RIPPG signal is much lower than that of a CPPG signal. Without the aid of contact probes and dedicated light sources, the ambient light is used as the light source for the RIPPG. Actually, ambient light is regarded as a noise source in the CPPG application. Besides, there is a certain distance between the digital camera and the subject. These changes from the CPPG to the RIPPG limit the SNR of a RIPPG signal. Second, the motion artifacts problem is much more serious for the RIPPG, compared with the CPPG [5]. When the subject has a motion, the ROI used by the RIPPG will also move correspondingly. Ambient light is not distributed with a spatially homogeneous intensity in most environments. The ambient light intensity and the digital camera's response on the ROI will vary with the movements of the ROI. Since the light intensity and the camera's response on the ROI act as the signal source for the RIPPG, the true RIPPG signal will be disturbed by the subject's motions. This is the reason why the motion artifacts problem is serious for the RIPPG.

The low SNR and high motion artifacts severely impair the RIPPG signal quality. Hence almost all researches on the RIPPG paid attention to improving the RIPPG signal quality. The approaches can be classified into three categories: the chrominance based method, the adaptive bandpass filter based method, and the ROI based method. The chrominance based method aims to extract the RIPPG signal through a linear combination of RIPPG signals from different color channels. The representative chrominance based methods consist of independent component analysis (ICA) [3] and the skin tone based method [4]. The adaptive bandpass filter based method focuses on refining RIPPG signals in the frequency domain by using RIPPG signal characteristics. Continuous wavelet transform is used to denoise and refine peaks for the RIPPG in [6]. In [7], auto-regressive model is utilized to perform the spectral analysis for the RIPPG.

The ROI based method targets on locating a good ROI on the skin region under measurement. Hence the selected ROI can provide a RIPPG signal with higher quality, compared with the whole facial or palmar region in the field of view. In order to improve the SNR of the RIPPG signal, all the pixels in a defined skin region will be averaged to calculate the RIPPG signal. Different regions of the skin carry RIPPG signals of different qualities, due to the difference in hair density, epidermis thickness,

capillary density, facial features, etc. Therefore, the regions which can provide relatively clear RIPPG signals should be selected as the ROI. While the regions which contain low quality RIPPG signals should be discarded. In [3], the whole face detected by the cascade classifier is chosen as the ROI. The forehead and cheeks regions can provide better RIPPG signals compared with other regions on the face, so the forehead or cheeks regions are chosen as the ROI in [2] and [8]. In [6] and [9], skin detection is utilized to mask the skin region and discard the non-skin region on the face in each frame because the non-skin region rarely contains the RIPPG information. And the dynamic skin mask is treated as a dynamic ROI for the RIPPG. In order to refine the ROI selection, the whole face region is divided into non-overlapped blocks in [10]. By far, the ROI selection for RIPPG still focuses on picking some regions which have good potential to provide good quality RIPPG signals, such as the forehead, the cheeks, and the skin region masked by skin detection. Though the block division method in [10] is more adaptive than the fixed ones, it still needs manual thresholding. There is a lack of an effective method to assess the RIPPG signal quality for the ROI selection.

This paper focuses on improving the RIPPG signal quality through a dynamic ROI (DROI) approach. The chrominance based method and the adaptive bandpass filter based method are not discussed in this paper. We want to investigate whether the RIPPG signal quality can be improved through refining ROI as a dynamic map. This dynamic map can indicate skin regions providing good quality RIPPG signals.

The rest of the paper is organized as follows. In section 2, we introduce two features, which can perform no-reference quality assessment of RIPPG wave segments. The DROI method is proposed in section 3. Experimental results are shown in section 4. Finally, a conclusion is drawn in section 5.

2. NO-REFERENCE QUALITY ASSESSMENT OF RIPPG SIGNALS

In order to determine a DROI, features for assessing the RIPPG signal quality are needed. There are three requirements for the feature selection. First, the features can directly index the RIPPG signal quality. Second, the quality assessment should be of no-reference type, because there is no ground truth. Third, the features can be extracted from a short RIPPG wave segment ($\leq$10 s), otherwise the ROI cannot be adaptively updated in time. According to above requirements, the cross correlation (CC) coefficient of consecutive wave segments and the SNR of a wave segment are utilized as features to assess the RIPPG signal quality.

2.1. Cross correlation of consecutive wave segments

A normal PPG signal is stable over time and consistent in the morphology. The PPG signal shape in the current heartbeat is quite similar to that of the previous heartbeat. However, a distorted or low-SNR PPG signal will not be consistent in morphology [11]. The shape of the distorted PPG wave varies from beat to beat. Hence the CC coefficient between consecutive wave segments of a clear PPG signal will be much higher than that of a distorted PPG signal. The CC coefficient is calculated as (1).

$$\gamma = \max_{u}\left(\frac{\sum_n[f(n)-\bar{f}_u][g(n-u)-\bar{g}_u]}{\sqrt{\sum_n[f(n)-\bar{f}_u]^2\sum_n[g(n-u)-\bar{g}_u]^2}}\right) \quad (1)$$

where $f(n)$ and $g(n)$ are two RIPPG signal segments, u is a shift between $f(n)$ and $g(n)$, γ is the CC coefficient, $\bar{f}_u$ is the mean of segment $f(n)$ in the overlap region during shifting, and $\bar{g}_u$ is the

mean of segment $g(n)$ in the overlap region during shifting. In this paper, the length of each RIPPG wave segment is 2 seconds. The shift is limited within ±0.5 seconds. The maximum CC coefficient at a certain shift u will be chosen as the CC coefficient between the two consecutive wave segments. There is a set of typical RIPPG signals shown in Fig.2. Fig. 2(a) is a RIPPG wave collected from a 5×5 pixels region on the subject's cheeks, which is a normal RIPPG signal. Fig. 2(b) is a RIPPG wave collected from a 5×5 pixels region on the subject's lips at the same time. Since the lip skin cannot provide a good RIPPG signal, there are distortions in the wave shown in Fig. 2(b). Fig. 2(c) shows the CC coefficients between consecutive wave segments (2s) for the normal RIPPG wave and the distorted RIPPG wave given above. The distorted RIPPG wave can be distinguished from the normal RIPPG wave by CC coefficients as shown in Fig. 2(c).
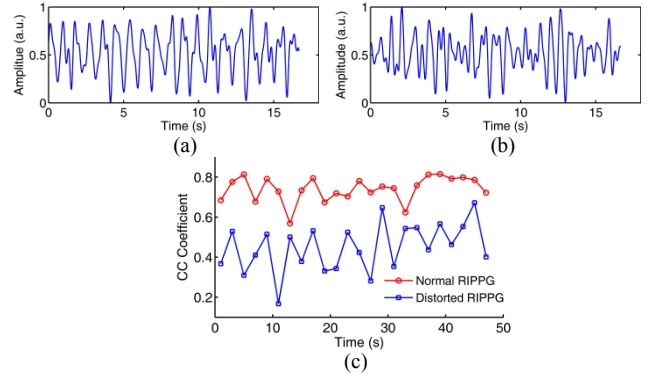

Fig. 2. Typical RIPPG waves and the CC coefficients of consecutive wave segments: (a) Normal RIPPG wave, (b) Distorted RIPPG wave, (c) CC coefficients of consecutive wave segments.

2.2. SNR of a wave segment

Since the normal PPG signal is quasi-periodical, a clear dominant peak and harmonic peaks will appear in the spectrum of the normal PPG signal, as shown in Fig. 3. A clear dominant peak in the spectrum means a high SNR of the RIPPG signal, which corresponds to a high RIPPG signal quality. When the RIPPG signal is unstable or distorted, its corresponding SNR will decrease. So we select the SNR of a RIPPG wave segment as another feature to distinguish between the normal RIPPG signals and the distorted ones. The SNR of a RIPPG wave segment used in this paper is defined as (2).

$$SNR = \frac{S(f_{HR})}{\sum_{f=0}^{f_{sampling}/2} S(f) - S(f_{HR})} \quad (2)$$

where $S(f)$ is the spectral density function of a RIPPG wave segment from a block, $S(f_{HR})$ is the spectral density at heart rate (HR) frequency f_{HR}, and $f_{sampling}$ is the sampling rate of the RIPPG signal. Actually, there is no reference CPPG signal to provide f_{HR}. Therefore, f_{HR} is estimated from the fixed ROI first, because averaging in the fixed ROI (the whole face/palm) can provide a stable estimation of f_{HR} in the frequency domain. Then the SNR for each block can be calculated using the estimated f_{HR} from the fixed ROI. In this paper, the window size for calculating the SNR of a RIPPG wave segment is 10 seconds, and the window will be shifted forward by 2 seconds every 2 seconds. Fig. 4 shows SNR values for the sample RIPPG waves used in Fig. 2. The SNR feature can distinguish the normal RIPPG wave segments from the distorted ones effectively.
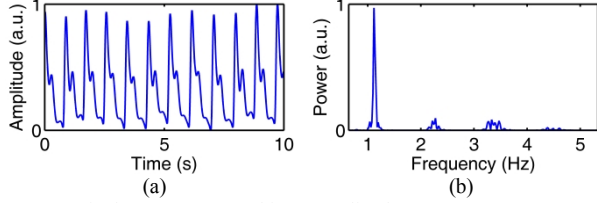
Fig. 3. A typical CPPG wave and its normalized spectrum: (a) CPPG wave, (b) Normalized spectrum.
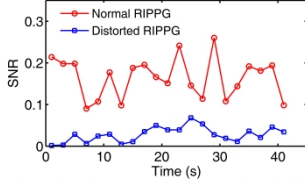

Fig. 4. The SNR values for the sample normal and distorted RIPPG waves.

3. METHODOLOGY

Based on no-reference quality assessment of the RIPPG signal, a DROI method for RIPPG is proposed, as illustrated in Fig. 5.
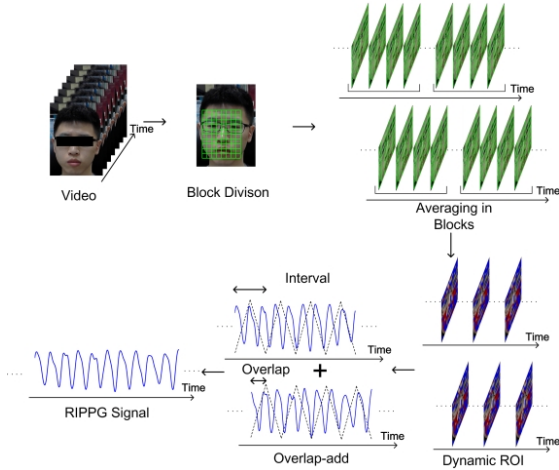

Fig. 5. The flow chart of the proposed dynamic ROI method for RIPPG.

A video clip for illustrating the proposed DROI method for RIPPG is provided at (http://youtu.be/lklryLa9xQQ). Detailed descriptions for each module are given in the following subsections.

3.1. Non-overlapped blocks division

First, a fixed ROI is selected on a skin region. Then the fixed ROI is divided into $M \times N$ non-overlapped blocks. Each block is defined as a square. The height of the fixed ROI will be divided into 25 equal parts, and the side length of each square block is 1/25 height of the fixed ROI. Averaging of pixel values in the green channel is performed in each of the $M \times N$ blocks along the time axis. Thus, $M \times N$ raw RIPPG signals can be obtained. After the averaging operation, the $M \times N$ raw RIPPG signals are processed using a bandpass filter with cutoff frequencies of 0.75-4 Hz. So $M \times N$ RIPPG signals can be accessed from the $M \times N$ blocks, respectively. In this paper, all RIPPG signals are extracted from the green channel of videos because the PPG signal gets the highest SNR in the green region of the light spectrum [12]. The cutoff frequencies of the bandpass filter corresponds to a HR range of 45-240 beats per minute (bpm).

Next, each RIPPG signal from different blocks is temporally segmented every 2 seconds along the time axis. The CC coefficients and the SNR values are calculated for the $M \times N$

RIPPG wave segments within a 2-second time window, and the two features are updated every 2 seconds. At the same time, motion detection is utilized to discard unreliable blocks within the same 2-second time window. The unreliable blocks usually contain hairs, lines or facial features. Motion detection is performed for each block within each temporal segment, as shown in (3).

$$D(i,j) = \max_{\substack{x,y \in Block \\ t \in Segment}} \left(\frac{dP(x,y,t)}{dt} \right), 1 \le i \le M, 1 \le j \le N \quad (3)$$

where $P(x,y,t)$ is the pixel value at the position of (x,y) in a block at the time t in a temporal segment, $D(i,j)$ is the maximum temporal derivative of pixel value in the pipeline defined by a spatial block and a temporal segment. Based on the motion detection result $D(i,j)$, a motion mask can be produced accompanying each feature map.

3.2. K-Means clustering on the feature map

The left $M_R \times N_R$ blocks belong to the reliable region, which correspond to the clear human skin. After reliable blocks are determined, a two dimensional feature map can be obtained for the fixed ROI in a 2-second frame sequence. Each point on the feature map consists of a pair of the CC coefficient and the SNR for a reliable block in this fixed ROI within a 2-second time window. Null values are assigned to the points corresponding to unreliable blocks. This feature map can describe the spatial distribution of RIPPG signal quality on a fixed ROI within a 2-second time window. K-means clustering is performed on each feature map [13]. The skin region providing good quality RIPPG signals, which are the blocks corresponding to the cluster with higher means of the CC coefficient and the SNR, will be mapped as the DROI as shown in Fig. 6. The DROI can be updated every 2 seconds. A RIPPG signal can be obtained by averaging the pixel values in the DROI sequence.
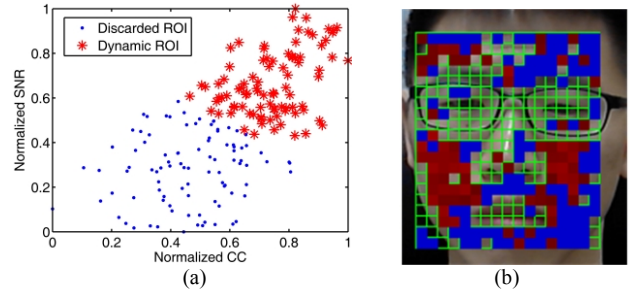

Fig. 6. K-means clustering on a feature map: (a) K-means clustering in a two dimensional feature space, (b) The clustering result mapped onto the fixed ROI (Red blocks - DROI, Blue blocks - Discarded ROI, Transparent blocks - Unreliable blocks).

3.3. Averaging in the DROI and overlap-add

Due to the update of the DROI every 2 seconds, there will be discontinuities in the RIPPG signal at transitions of the DROI. In order to remove these discontinuities, an overlap-add operation is adopted. The temporal segmentation of $M \times N$ RIPPG signals is shifted forward along the time axis by 1 second. A second DROI sequence is generated using the shifted temporal segmentation windows. A second RIPPG signal can be obtained by averaging the pixel values using the second DROI sequence. The two RIPPG signals have staggered discontinuities every 1 second. Each of the two RIPPG signals will be multiplied by repeated triangle windows, with an interval of 2 seconds. The final RIPPG signal results as the sum of these overlapping pieces, with an overlap of 1

second.

4. EXPERIMENTAL RESULTS

In total, 19 healthy volunteers (13 males, 6 females) in the ages of 23 to 34 years participated in the study. A low-cost webcam (Logitech C270) was used as the imaging device. Frame rate is 30 fps. All videos were recorded in color space (24-bit RGB) with a resolution of 640×480. A 1-minute long video was recorded for each subject. The subject was asked to raise the right hand, keep steady, and sit in front of the webcam at a distance of approximately 75cm during the video capture. A CPPG (HRS-06UF Heart Sensor) was contacted to the subject's index finger to record the BVP wave for reference. The proposed DROI method for RIPPG is tested on both the facial region and the palmar region. RIPPG signals are accessed from the two regions using the proposed DROI method and different benchmark methods [3], [6], [8], [10]. For the facial region, the complete fixed ROI (CFROI) is a fixed ROI located by Viola-Jones (VJ) face detector in the initial frame of a video as the outer rectangle on the face in Fig. 7 [14]. And the partial fixed ROI (PFROI) on the facial region consists of two fixed ROIs on the cheeks as the two dotted inner rectangles on the face in Fig. 7. For the palmar region, the CFROI is manually selected as a rectangle over the palm as shown in Fig. 7. There is no research about the optimal ROI on the palm for RIPPG, so the PFROI method was not tested on the palmar region. The ICA based RIPPG method [3], the HR map based ROI method [10], the skin mask based ROI method [6], and the proposed DROI method are all implemented by the use of the CFROIs on the face and the palm.
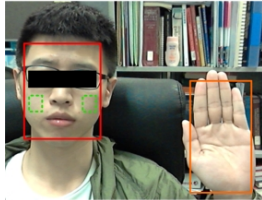
Fig. 7. Fixed ROIs on the facial region and the palmar region.

Then HR measurement was performed using these RIPPG signals and the reference CPPG signals of the 19 subjects. Bland–Altman analysis is utilized as the main method to analyze the agreement between the RIPPG and the CPPG [15]. The measurement bias and 95% limits of agreement between the RIPPG and the CPPG will be calculated to check the accuracy of the RIPPG. 95% limits of agreement is bias ±1.96 standard deviation of the differences, which shows how far apart HR measurements by the RIPPG are more likely to be for most individuals. Pearson's correlation coefficients (PCC) and the corresponding p-values are calculated to measure the linear dependency between the RIPPG and the CPPG [16]. In addition, root mean square error (RMSE) between HR values accessed by the CPPG and the RIPPG is also utilized to measure the accuracy of the RIPPG.

For the facial region, the statistic results using Bland–Altman analysis for different RIPPG methods are given in Table 1. The boxplot of RMSE of HR measurements using different RIPPG methods is shown in Fig. 8(a).

For the palmar region, the statistic results using Bland–Altman analysis for different RIPPG methods are given in Table 2. The boxplot of RMSE of HR measurements using different RIPPG methods is shown in Fig. 8(b).

It can be observed that the proposed DROI based RIPPG obtains the best performance on both the facial region and the palmar region, compared with the benchmark methods.
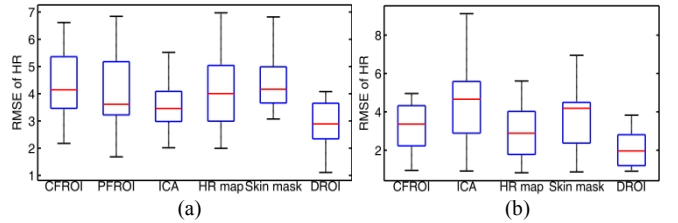
Fig. 8. The boxplot of RMSE of HR measurements using different RIPPG methods (a) on the facial region and (b) on the palmar region.

5. CONCLUSION

Due to the difference in the operation mode between the CPPG and the RIPPG, the ROI transforms from a point for the CPPG to an area for the RIPPG. Thus, ROI selection becomes a new problem for the RIPPG application. With the aid of two features, SNR and CC, for no-reference quality assessment of RIPPG signals, a dynamic ROI can be obtained based on K-means clustering. The experimental results approve the effectiveness of the proposed DROI method for RIPPG. In addition to improving the RIPPG signal quality, the proposed DROI method can also reveal the spatial distribution of the RIPPG signal quality on some skin region. Furthermore, this DROI method can be flexibly applied on different skin regions, including the facial region and the palmar region. Since the proposed DROI method for RIPPG is able to vary the ROI along with the subject's physical variation and the environmental variation, we believe that this DROI method can enhance the adaptability of the RIPPG in practice.

TABLE 1
THE PERFORMANCES OF DIFFERENT RIPPG METHODS IN HR MEASUREMENT ON THE FACIAL REGION

Statistic		CFROI	PFROI	ICA	HR map	Skin mask	DROI
Bias (bpm)		−0.7	−0.7	−0.8	−0.7	−0.9	−0.7
Limits of agreement	Upper limit (bpm)	8.2	7.4	6.2	7.9	8.2	5.0
	Lower limit (bpm)	−9.5	−8.8	−7.8	−9.4	−9.9	−6.4
PCC		0.924	0.934	0.951	0.928	0.921	0.967
p-value		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

TABLE 2
THE PERFORMANCES OF DIFFERENT RIPPG METHODS IN HR MEASUREMENT ON THE PALMAR REGION

Statistic		CFROI	ICA	HR map	Skin mask	DROI
Bias (bpm)		−0.7	−0.9	−0.6	−0.5	−0.6
Limits of agreement	Upper limit (bpm)	6.0	8.4	5.9	7.5	3.8
	Lower limit (bpm)	−7.4	−10.2	−7.0	−8.6	−4.9
PCC		0.954	0.915	0.957	0.933	0.980
p-value		<0.001	<0.001	<0.001	<0.001	<0.001

6. REFERENCES

[1] A. B. Hertzman, C. R. Spealman, "Observations on the finger volume pulse recorded photoelectrically," *Am. J. Physiol.*, vol. 119, pp.334−335, 1937.

[2] W. Verkruysse, L. O. Svaasand, J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16(26), pp.21434−21445, 2008.

[3] M. Poh, D. J. McDuff, R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18(10), pp.10762−10774, 2010.

[4] G. Haan, V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Bio-Med. Eng*, vol. 60(10), pp.2878−2886, 2013.

[5] L. Feng, L. M. Po, X. Xu, Y. Li, "Motion artifacts suppression for remote imaging photoplethysmography," in *Proc. DSP*, pp.18−23, Aug. 2014

[6] F. Bousefsaf, C. Maaoui, A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomedical Signal Processing and Control*, vol. 8(6), pp.568−574, 2014.

[7] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.* vol. 35(5), pp.807−831, 2014.

[8] G. Lempe, S. Zaunseder, T. Wirthgen, S. Zipser, H. Malberg, "ROI selection for remote photoplethysmography," *Bildverarbeitung für die Medizin*, pp.99−103, 2013.

[9] N. Kawasaki, "Improving motion robustness of contact-less monitoring of heart rate using video analysis," *M.S. thesis*, Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands, 2011.

[10] A. Rustand, "Ambient-light Photoplethysmography," *M.S. thesis*, Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway, 2012.

[11] W. Karlen1, K. Kobayashi, J. M. Ansermino, G. A. Dumont, "Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation," *Physiol. Meas.* vol. 33(10), pp.1617−1629, 2012.

[12] J. A. Crowe, D. Damianou, "The wavelength dependence of the photoplethysmogram and its application to pulse oximetry," in *Proc.* IEEE Annu. Int. Conf. Eng. Med. Biol. Soc., vol.6, pp.2423−2424, 1992.

[13] C. M. Bishop, "Mixture models and EM," in *Pattern Recognition and Machine Learning*, New York: Springer, 2006.

[14] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. CVPR*, vol. 1, pp.I-511−I-518, 2001.

[15] J. M. Bland, D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 1(8476), pp.307−310, 1986.

[16] B. Jacob, J. Chen, Y. Huang, I. Cohen, "Pearson correlation coefficient," In *Noise reduction in speech processing*, Springer Berlin Heidelberg, pp.1−4, 2009.