Optimization of Non-contact based Video Plethysmography Signal Acquisition for Detection of Cardiac Arrhythmia

Madan Kumar Lakshmanan
CSIR-Central Electronics Engineering Research Institute, Chennai, India

Email: mklakshmanan@ceeri.res.in

Abstract— Non-invasive detection of cardiac arrhythmia is essential for effective continuous monitoring of high risk patients suffering from congestive heart diseases or recovering from myocardial infarctions or in danger of being inflicted with arrhythmia induced cardiomyopathy. In this paper, a technique based on Video plethysmography (VPG) is applied for the accurate prediction of cardiac arrhythmia. VPG is an advanced signal processing technique which enables remote measurement of heart rate (HR) measures from which an anomaly in the subject's cardiac functioning can be deduced. In order to efficiently extract the HR measures, the factors which affect the capture of VPG signals are identified. This includes optimization of the intensity of illumination, use of polarizers to minimize specular reflection and customization of camera features such as automatic white-balance & exposure mode. Smart algorithms are used for facial recognition and identification of skin-pixels. Subsequently, advanced image processing algorithms are applied for real-time processing of video signals to extract video plethysmography signals. A database consisting of over 100 videos captured from over 20 subjects having different skin-tones and belonging to different demographic groups were used in the studies. The heart rate extracted from the VPG signals matches well (within $\pm$1.5 bpm) with the reference heart rate obtained using finger PPG device demonstrating the potential of the proposed approach as a screening tool for continuous and non-invasive cardiac arrhythmia monitoring.

I. INTRODUCTION

Monitoring cardiac arrhythmia is important in patients with high risks and can help in the detection of early markers for cardiovascular diseases, arrhythmias (mainly Atrial Fibrillation, Tachycardia, and Bradycardia), congestive heart failure, etc. It is also important to continuously monitor the cardiac state of patients who have undergone heart surgeries. The solutions which currently exist for continuous monitoring of heart rate are expensive, restricted to bed care and uncomfortable for the patients. In contrast, non-invasive techniques such as Video Plethysmography (VPG) offers an alternative approach for unobtrusive monitoring of patient's cardiac activity. This is accomplished by tracking the subtle variations in color of the skin surface induced by the pulsatile blood volume changes due to cardiac activity. The pulsatile blood propagating in the cardiovascular system alters the flow of blood volume within the skin tissues which in turn causes variations in the quantity of light absorbing hemoglobin molecules. Such changes in the absorption of light results in subtle variations in the skin tone. By the application of advanced signal and video processing techniques on videos captured by a normal camera, the variations in the skin-tone can be used to extract the pulsatile information even in ambient light. Such an application can be advantageously used in diverse environments including the non-contact based cardiac arrhythmia monitoring.

Recent literature on extracting pulse waveform from a human face includes Gerard de Haan and Arno van Leest's work [1] on improving motion robustness of Remote Photoplethysmography (r-PPG) using blood volume variations. The method suggested detects minute optical absorption changes caused by variations in blood volume of the skin. The different absorption spectra of arterial blood and bloodless skin are shown to cause variations in skin-color in a normalized RGB-space. The pulse signal is estimated from the time-sequential RGB pixel data averaged over the skin area. Wenjin Wang et.al [2] introduced a new algorithm for rPPG called spatial subspace rotation (2SR). The method estimates a spatial subspace of skin-pixels and measure its temporal rotation for pulse extraction. The advantages offered by this method are that it does not require prior information on the skin-tone and is robust to any motion induced artifacts. In [3] Gerard de Haan mathematically modeled the impact of motion on the PPG signals. Accordingly he presents various methods to combine the chrominance signals for noise cancellation, reduction of specular reflections, reduction of motion artefacts and efficient extraction of pulse signal. In [4] Xiaochuan He et.al, propose a novel method for the calculation of pulse transit time using Eulerian Video Magnification Framework (EVM). In the method the authors apply spatial decomposition and temporal filtering to the frames of the video, then the filtered signal is amplified to reveal the subtle changes in color caused by the blood pulsations.

In this paper, we present our work on optimizing the various parameters that influence the capturing of best videos for the efficient extraction of VPG signals that accurately map the heart rate of the subject. To this end over 100 videos from 20 subjects with different skin-tones, mean heart rates and demographic backgrounds were captured under different lighting conditions, different orientations of illumination arrangements, camera settings (white balancing & colour temperature). The best algorithm for extraction of good video plethysmography signals is also identified. The extracted heart rate from VPG signals are shown to match excellently well with the heart rate obtained from reference PPG signals demonstrating the potential of this approach as a valuable screening/diagnostic tool for early detection of cardiac arrhythmia.

II. PROCESS METHODOLOGY

The process flow to acquire videos and then obtain heart rate is elucidated in Figure 1. Videos were collected at 30 frames per second under stable and ambient lighting conditions with a consumer grade camera. A state-of-the-art detection framework, namely, Viola-Jones [5] was used to detect the face of the subject. Viola Jones algorithm is a robust, real-time and time-efficient algorithm that is used to find the Region-Of-Interest (ROI) in every frame of the video. The framework called *vision.CascadeObjectDetector* distinguishes facial regions from other features in the frame by applying a spatial filter. An empirically designed face-mask was then used to identify the skin-pixels. The skin-pixels provide information of pulsatile changes in the blood. To accurately detect the skin across subjects with different skin-tone the image is converted into a colorspace where the range of human skin tones are clustered to form a definitive shape. The RGB image is converted to YCbCr space where the illumination channel (Y) is separated from the two orthogonal Chrominance channels (Cb&Cr) [6]. This is useful because the detection of skin pixel is not affected by the illumination used. For detection of light, medium and dark skinned pixels, a range of possible skin pixel values are used for Cb and Cr color space (Cb>=98 & Cb<=142 & Cr>=133 & Cr<=177)) [6]. Two algorithms were applied to the images to extract VPG signals. Temporal averaging of green channel and Spatial Subspace Rotation.

Temporal Averaging of Green Channel Method is a simple mean of the green pixels found in the image captured. The ROI used in this image is Forehead/ Cheek area. The signals derived from Green channel averaging was more prominent than the signals derived from red and blue because the maximum absorption of oxyhemoglobin occurs in green light.

This method was found feasible during stable state conditions, but in real-time scenario, with movements in the person's face due to actions such as blinking, talking, etc. the method did not perform well.

As an alternative the Spatial Subspace Rotation described in Section IV was used. The derived VPG signals were then low-pass filtered to remove high frequency powerline noise. The heart rate is then deduced from the VPG signals.

A dataset of over 100 videos was recorded capturing variations in skin-tone, quality of illumination, camera settings and motion-induced artefacts. The process is depicted in Figure 1. To validate the results, reference data was collected from a fingertip Photo Plethysmography (PPG) Sensor (refer Figure 2).

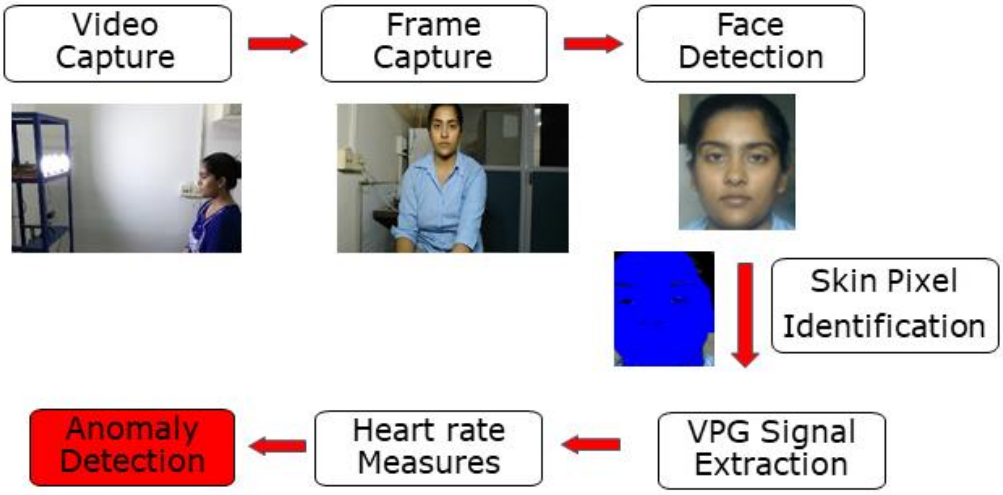


Figure 1: Process Flow – Contactless VPG extraction & anomaly detection

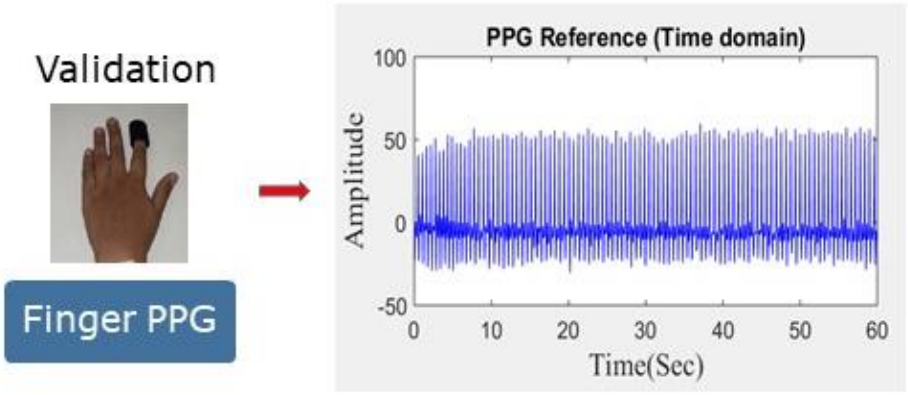


Figure 2: Validation of VPG signal with reference PPG

The setup used to capture the videos is shown in Figure 3. The various illumination arrangements used are depicted in Figure 4.

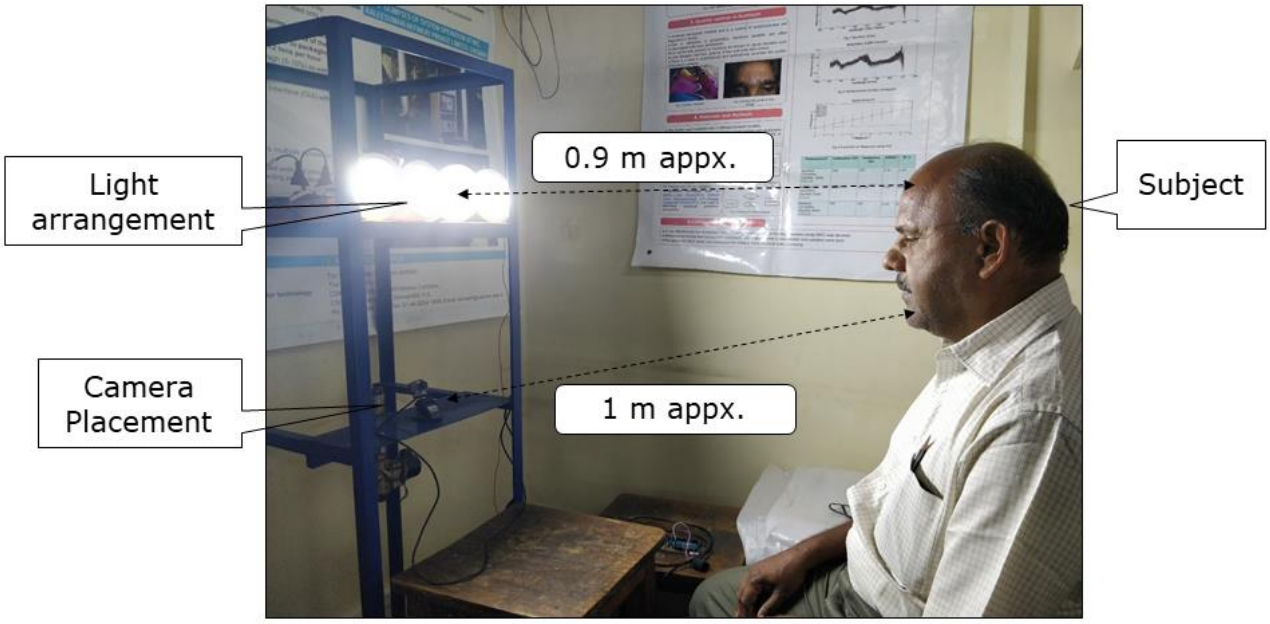


Figure 3: Typical setup used for capture of videos

Figure 4: Arrangement of illumination sources

III. Optimization of VPG Signal Acquisition

The following parameters were optimized to obtain the videos of best quality for the efficient extraction of VPG signals:

a. Quality of illumination
b. Orientation of illumination source
c. Camera configurations
d. Camera settings
e. Algorithms for efficient processing of videos to extract VPG signals of best quality
f. Methods to remove specular reflection during collection of videos
g. Algorithms for optimal identification of face
h. Methods for best identification of skin-pixels for subjects with varying skin-tones
i. Polarization of reflected light to minimise contribution from specular reflections

Four different cameras, namely, RasPiCam, Laptop camera, Intex consumer grade camera and Logitech C922 camera, were used to collect videos. The illumination was altered by using a standard light setup consisting of an array of one to four led lamps with different orientations for uniform coverage of light (refer Figure 4). Polarizers were also used to study the impact of specular reflection. Subjects from various

age groups, skin color, and facial aberrations/hair were selected to study the effects of the previously mentioned parameters on a video signal. Table 1 enlists the various configurations applied during the studies.

Based on the studies conducted, it was observed that the VPG signals of best quality were obtained when the subject is illuminated by 4 LED lamps (600 lux appx.) with the colour temperature of white balance set to 5000 K and use of a polarizer which is horizontally oriented is placed in front of the camera. Better VPG signals were extracted using 2SR method compared to the PPG signal obtained from RGB method. It may be mentioned that the results presented in this paper were obtained from videos captured using the above settings.

Table 1: Parameters for Optimization

Camera	Raspberry Pi	Laptop	Intex	Logitech C922
Illumination source	1 LED lamp	2 LED lamps	3 LED Lamps	4 LED lamps
White balance	2000 K	3000 K	4000 K	5000 K	6000 K
Orientation of Polarizer	No polarizer	Horizontal	45 degrees	Vertical
Orientation of Light Sources	Straight	Square	Rectangle
Algorithm for extraction of VPG signals	RGB	Spatial Subspace Rotation

IV. SPATIAL SUBSPACE ROTATION

In this section we describe the 2SR algorithm [2] which has been employed for the efficient extraction of VPG signals from the video frames. The methods has been shown to be robust against motion induced artifacts. Skin pixels are relatively homogenous in chromaticity. Assuming a uniform skin color in face, the distribution of skin-pixels in RGB-space is spherical. Due to intensity variations induced by shadows and specular reflections, the distribution is close to elliptical drawing the low intensity pixels close to RGB-space origin. The fluctuations in the hemoglobin content due to pulsatile blood volume change in skin tissues result in fluctuations in optical absorption across light spectrum in commensurate with the cardiac activity. We extract the principal component vector of this distribution and track the hue changes induced by pulsatile blood propagating in the skin tissues. Eigen vector analysis is used to extract principal component vector of RGB space of skin-pixels [2].

$$C = \frac{V^T.V}{N} \qquad (1)$$

V is an RGB pixel matrix with size $N \times 3$ where N is the number of skin-pixels and C is the correlation matrix of RGB w.r.t skin-pixels. Eigen values decomposition gives three Eigen vectors (u_1, u_2, u_3) and corresponding eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ where $\lambda_1 > \lambda_2, \lambda_3$ and $u_1$is the principal component vectors for skin-pixel in RGB-space.

Temporal relation between two subsequent frames is modeled using instantaneous rotation and scaling of principal Eigen vectors for corresponding RGB-subspaces [2]. We take a temporal stride of L frames to analyze subspace rotation. The subspace of first frame U_{ref} having eigenvectors of reference frame. Successive frames have subspaces $U_{t,t<L}$ and rotation between frames and reference frame is:

$$R = U_t^T . U_{ref} \qquad (2)$$

where R is the rotation matrix.
Rotation is measured against the $u_2^{ref} \times u_3^{ref}$ plane of reference. The rotation of u_1^t w.r.t this plane of reference will signify the pulsatile change of pixels.

$$R' = (u_1^{tT}) \cdot \left(u_2^{ref}\ u_3^{ref}\right) = \left(u_1^{tT} \cdot u_2^{ref}\ u_1^{tT} \cdot u_3^{ref}\right) \quad (3)$$

In addition to subspace rotation, eigenvalues corresponding to the variance/energy of the eigenvectors are also influenced by pulsatile blood. We will scale the rotation vectors with the eigen values

$$S^T = \left(\sqrt{\frac{\lambda_1^t}{\lambda_2^{ref}}} \quad \sqrt{\frac{\lambda_1^t}{\lambda_3^{ref}}}\right) \qquad (4)$$

S signifies the scale/energy change of the rotated subspace which is related to pulsatility of the skin. But skin reflections also contain motion distortion induced intensity changes. The component of S in the direction of rotation vector, which is $SR = S^T \cdot R^{'}$, is taken for pulse extraction. Since this rotation vector is in reference frame, it is back-projected to RGB plane in order to standardize the signal extraction for all the strides.

$$SR' = SR \cdot \begin{pmatrix} u_2^{ref\,T} \\ u_3^{ref\,T} \end{pmatrix} \qquad (5)$$

SR' is a 3×1 matrix having the subspace rotation component scaled and back-projected RGB space.
In a stride, SR' are concatenated to make a $3 \times L$ matrix. Three L-length traces obtained in SR' and first two are found to be in anti-phase. These two traces are combined to boost the pulse [2]. The pulses obtained across multiple strides are averaged and VPG signal is extracted [2].

V. RESULTS

Subjects numbering more than 20 with varying skin-tones and age-groups were used for the studies. The subjects selected had different mean heart rates which could be classified as normal (~60-90 bpm), Tachycardia (>100 bpm) and Bradycardia (<60 bpm). Figure 5 depicts a sample-set of the subjects from whom the videos were collected during the studies. The subjects were made to sit in front of the camera arrangement at a distance of about 1m as depicted in Figure 3. The parameters for capture of video are detailed in Section III. Each video was collected for a duration of 3 minutes. The extracted VPG signals and the HR measures for the some of the subjects are shown in Figure 6.

Figure 5: Sampleset of subjects used for studies. The subjects belonged to different age-groups and had different mean heart rate & skin-tones.

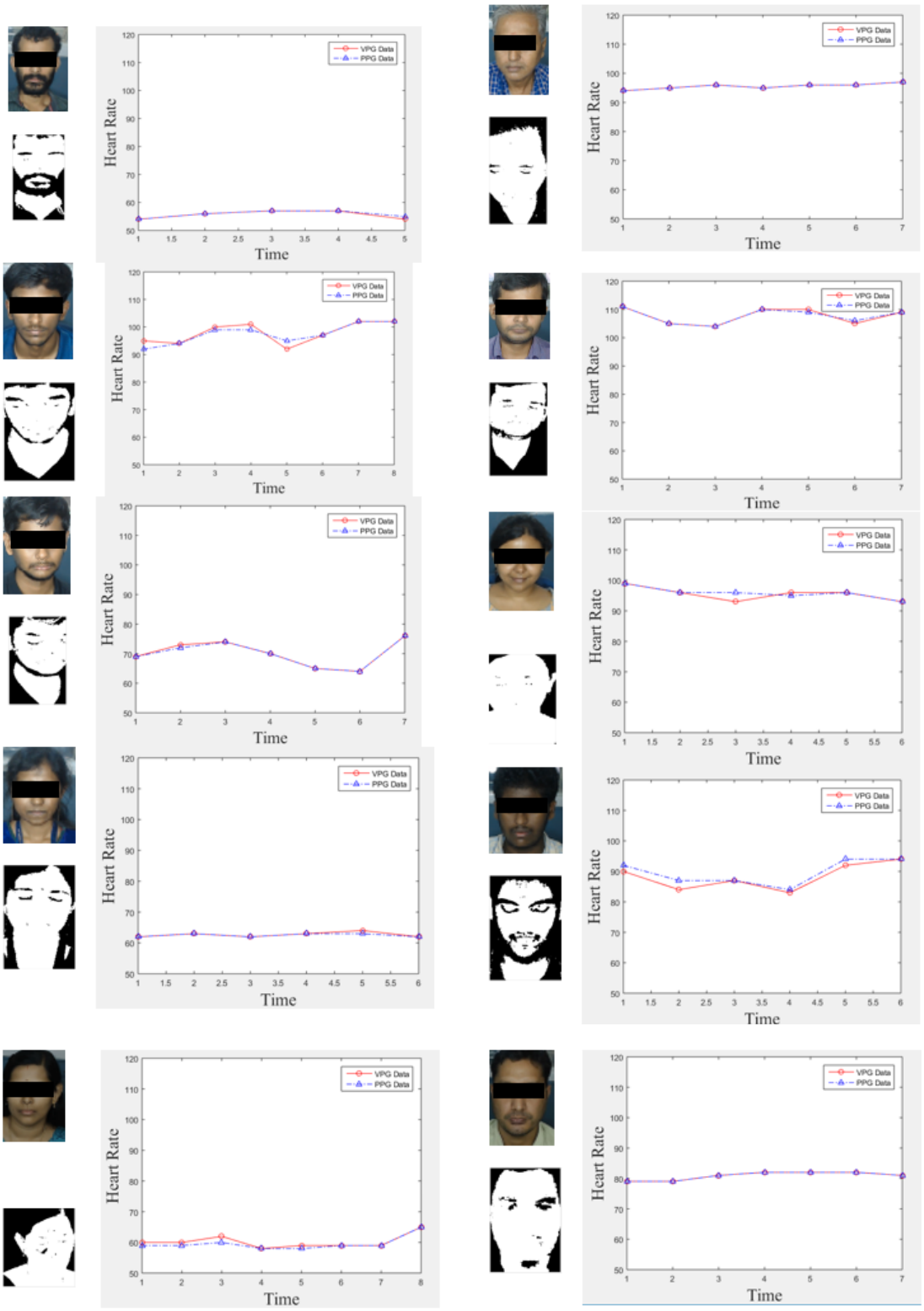


Figure 6: Heart rate plots for extracted VPG signals for different subjects. The extracted heart rate for all subjects show excellent correlation with reference PPG data collected from finger.

Table 1: Heart rate extracted from VPG signals – Subjects with Normal Heart rate (~ 60-90 bpm)

Subjects	Mean Heart Rate	RMSD (VPG vs Reference PPG)
Subject 1	92.00	1.73
Subject 2	80.00	0
Subject 3	67.00	0
Subject 4	70.00	0.45
Subject 5	75.60	0

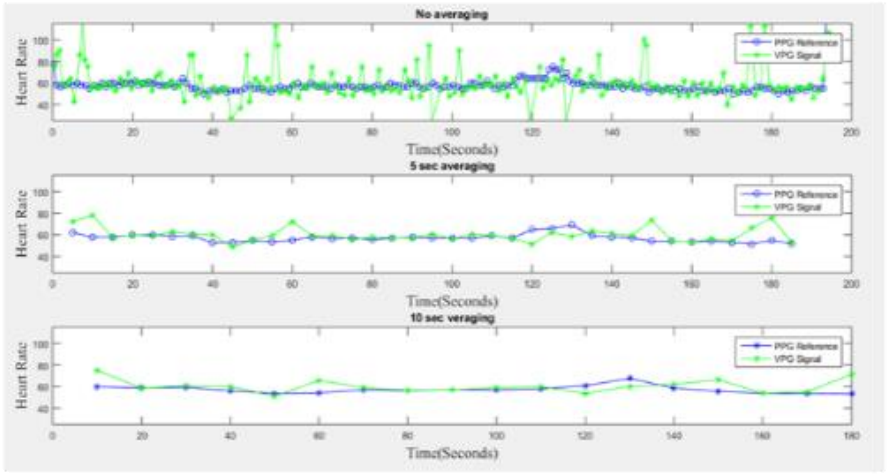
Table 2: Heart rate extracted from VPG signals – Subjects with Bradycardia (<60 bpm)

Subjects	Mean Heart Rate	RMSD (VPG vs Reference PPG)
Subject 1	56.00	0.44
Subject 2	61.70	1.29
Subject 3	62.49	0.38
Subject 4	60.00	1.56
Subject 5	59.60	0.93

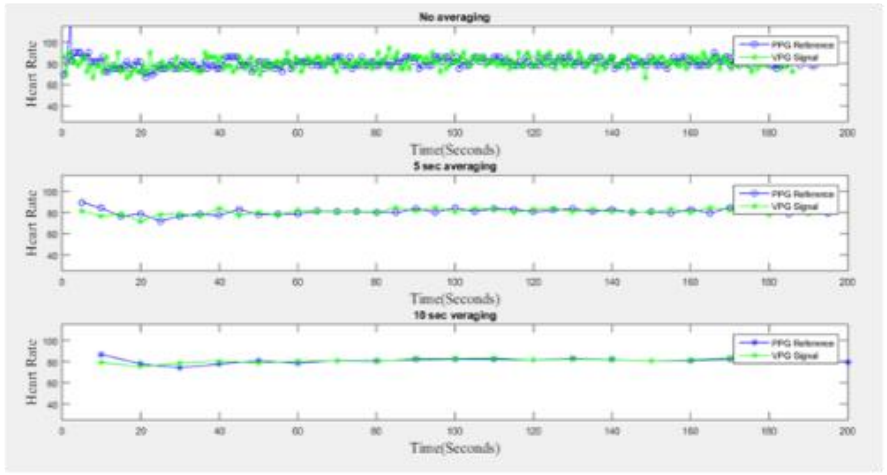
Table 3: Heart rate extracted from VPG signals – Subjects with Tachycardia (>100 bpm)

Subjects	Mean Heart Rate	RMSD (VPG vs Reference PPG)
Subject 1	95.57	0
Subject 2	109.00	0.52
Subject 3	95.83	1.29
Subject 4	97.50	1.69
Subject 5	107.50	0.38

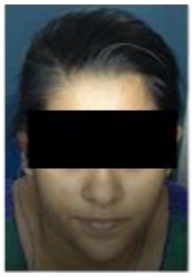
Subject 1
Bradycardia



Subject 2
Normal



Subject 3
Tachycardia



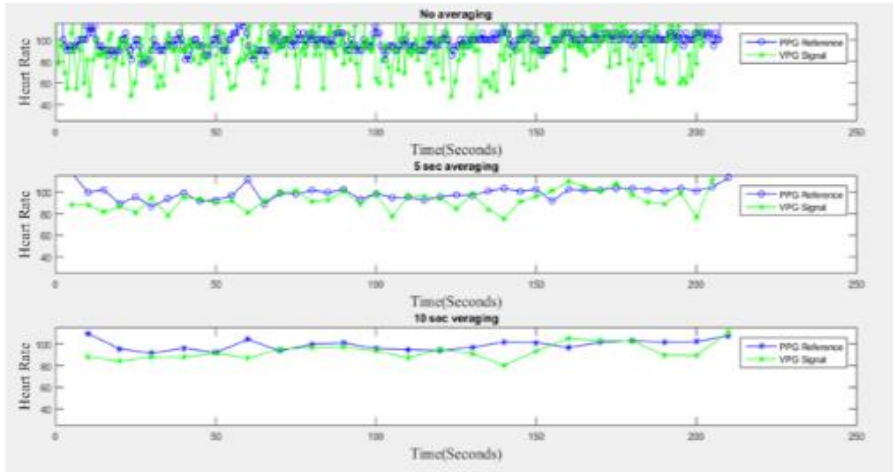


Figure 7: Heart rate graphs of Normal, Bradycardia and Tachycardia data – instantaneous, 5 seconds averaged & 10 seconds averaged. The derived heart rates match well with reference PPG (in finger) data demonstrating its potential as a vital screening tool for cardiac arrythmia.

Table 1-3 provide the mean heart rate extracted from VPG signals for subjects with normal, Bradycardia & Tachycardia conditions, respectively. The root-mean-square-difference of heart rates from the extracted VPG in relation to the heart rate derived from reference PPG from finger are also enlisted. The heart rate graphs of Normal, Bradycardia and Tachycardia data – instantaneous, 5 seconds averaged & 10 seconds averaged – are plotted in Figure 7. As evidenced from the tables and the figure the extracted heart rate from VPG signals matches very well with the heart rate obtained from reference PPG signals (< 1.5 bpm) demonstrating the potential of this approach for cardiac arrhythmia monitoring.

VI. CONCLUSION

A non-contact video plethysmography method is explored for continuous and un-obtrusive cardiac state monitoring. Detailed studies show that the optimization of various parameters such as lighting conditions, white balance, specular reflection and skin mask for identification of skin pixels are necessary to obtain good video plethysmography signals. The extracted heart rate from VPG signals matches very well with the heart rate obtained from reference PPG signals demonstrating the potential of this approach for cardiac arrhythmia screening. In the future efforts will be expended towards (1) validation of the presented studies with Holter ECG and (2) extension of methods to identify atrial fibrillation and flutter.

REFERENCES

[1] Gerard de Haan and Arno van Leest, “Motion-robust rPPG using the blood volume pulse signature”, Physiol Meas. 2014 Aug 27;35(9):1913-1926.
[2] Wenjin Wang, “A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation”, IEEE Transactions on Biomedical Engineering, Volume 63, No. 9, September 2016, pp- 1974-1984.
[3] G. de Haan and V. Jeanne, “Robust pulse rate from chrominance-based rPPG,” IEEE Trans. Biomed. Eng., vol. 60, no. 10, Oct. 2013, pp. 2878–2886.
[4] Xiaochuan He, Rafik A. Goubran and Xiaoping P. Liu, “Using Eulerian Video Magnification Framework to Measure Pulse Transit Time”, IEEE International Symposium on Medical Measurements and Applications, 2014, 978-1-4799-292,pp-1-4.
[5] Viola, P. & Jones, M.J. International Journal of Computer Vision (2004) 57: 137. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
[6] D. Chai K. Ngan "Face segmentation using skin-color map in videophone applications" Circuits and Systems for Video Technology IEEE Transactions vol. 9 pp. 551-564 Jun 1999.