Improvements in Remote Video based Estimation of Heart Rate Variability using the Welch FFT Method

MUNENORI FUKUNISHI*1, DANIEL McDUFF*2, NORIMICHI TSUMURA*1

[1]*Graduate School of Advanced Integration Science, Chiba University, 1-33 Yayoi-cho, Inage-Ku, Chiba 263-8522, JAPAN*
[2]*Microsoft Research, Redmond, United States*

Abstract: Non-contact heart rate and heart rate variability measurement has applications in healthcare and affective computing. Recently, a system utilizing a five-band camera (RGBCO: red, green, blue, cyan, orange) was proposed, and shown to improve both remote measurement of heart rate and heart rate variability over an RGB camera. In this paper, we propose an improved method for video-based estimation of heart rate variability. We introduce three advancements over previous work utilizing five-band cameras: (i) an adaptive non-rectangular region of interest identified using automatically detected facial feature points, (ii) improved peak detection within the blood volume pulse (BVP) signal, (iii) improved HRV calculation using the Welch periodogram. We apply our method to a test dataset of subjects at rest and under cognitive stress and show qualitative improvements in the stability of HRV spectrogram estimation. Although we evaluate our method using a five-band camera the method could be applied to video recorded with any camera.

Keywords: heart rate, heart rate variability, non-contact measurement, five-band cameras

1 INTRODUCTION

Non-contact video-based measurement of physiological signals is an active research area due to the numerous potential applications in health and wellness tracking, medical diagnosis, and affective computing. Recovering the blood volume pulse (BVP) is a challenge central to remote detection of physiological information and various techniques have been proposed for achieving this. Verkruysse et al. [1] demonstrated BVP measurement under ambient light was possible using spatially averaged measurements from green channel pixels in videos of the skin captured by a consumer camera. Poh et al. [2] developed a remote BVP measurement technique using a low-cost webcam, incorporating a blind source separation method, which can be used to calculate heart rate (HR), and the high- and low-frequency (HF and LF) components of heart rate variability (HRV) [3]. Haan et al. recently proposed a remote photoplethysmography (rPPG) method which is based on a simple skin optics model [15, 16].

Frequency analysis of heart rate variability via HRV spectrograms (HRVS) is a useful tool for non-invasive monitoring of the autonomic nervous system (ANS). The ANS controls involuntary body functions, such as breathing, blood pressure, and heart rate. The LF power within the HRV signal (0.05-0.15 Hz) is influenced by both sympathetic and parasympathetic activity and the power increases during sympathetic activation. For example, activation resulting from cognitive load [4]. The HF power within the HRV signal (0.15-0.40 Hz) is influenced by breathing and parasympathetic nervous system activity. The HF power increases when subjects are at rest. McDuff et al. [5] developed a remote HRVS measurement technique using a novel camera with a five-band sensor: red, green, blue, cyan, and orange (RGBCO) [6, 7]. They demonstrated that remote HRVS measurement could capture subtle changes in ANS activity between individuals at rest and under cognitive load.

In this paper, we propose three methodological improvements to previous work using five-band cameras for remote physiological measurement: (i) an adaptive non-rectangular region of interest (ROI) set using automatically detected facial feature points, (ii) improved BVP peak detection, (iii) improved calculation of the HRVS using the Welch periodogram [12]. The rest of this paper is organized as follows. First, we outline the method for remote extraction of the BVP and HRVS using a five-band camera which was proposed by McDuff et al. [5]. Second, we describe our proposed method for HRVS measurement and highlight the benefits over the method of McDuff et al. [5]. Third, we give details of the experimental setup, data collection and show the results of HRV detection with the proposed method using a five-band camera. Finally, we present our conclusions.

2 Existing Framework for Estimation of the Heart Rate Variability Spectrograms

Below we describe the method for remote extraction of the BVP and HRVS using the five-band camera which was proposed by McDuff et al. [5]. The method uses facial video sequences captured by a camera with five color

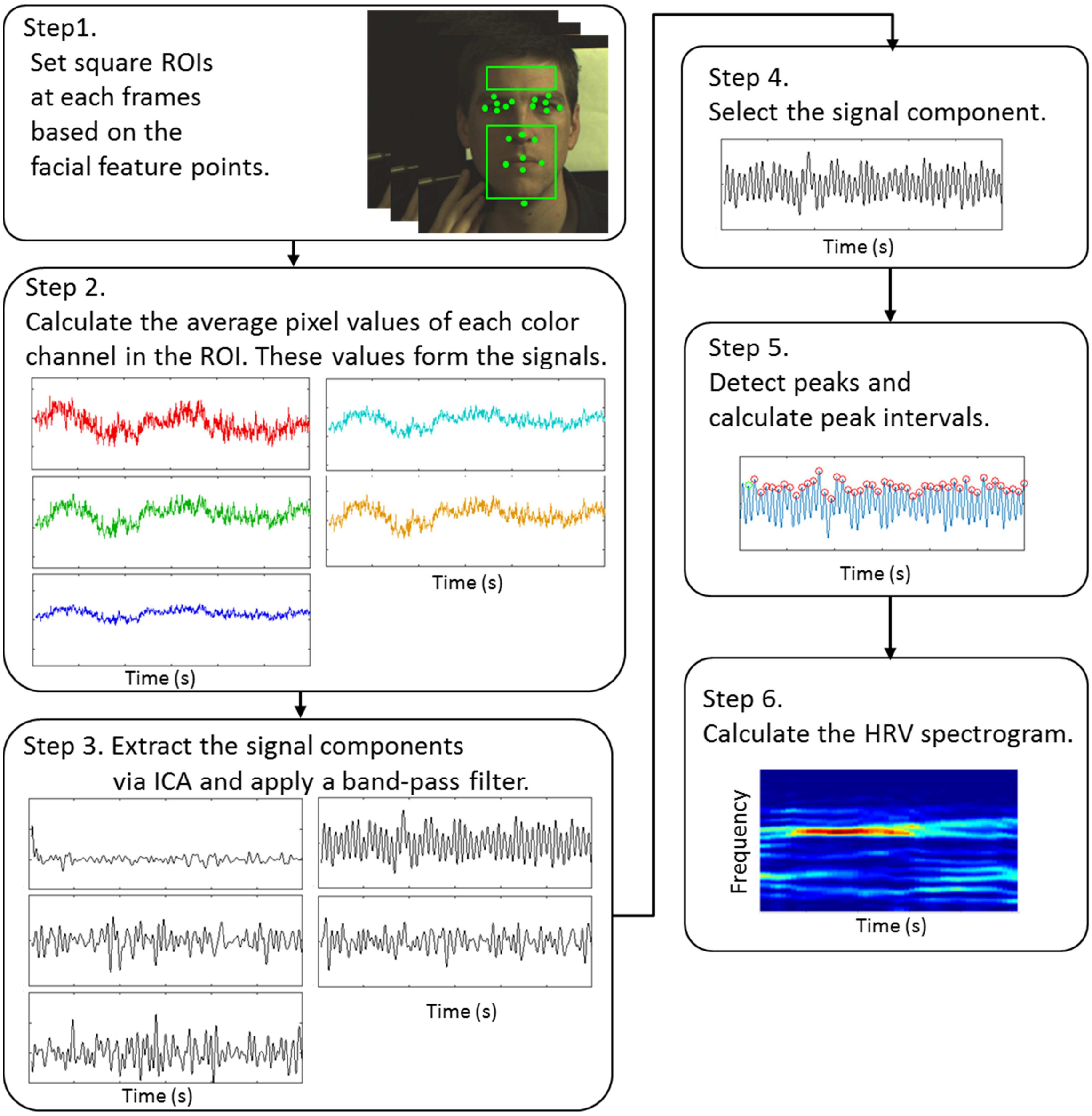

Fig. 1. Overview of the method to recover HRV spectrograms from videos. 1) Each frame is segmented using automatically detected facial feature points. 2) Spatial averages of the color channel signals are calculated over a time window to form observation signals. 3) ICA is performed to recover the underlying source signals. 4) The source containing the strongest BVP signal is selected. 5) Peaks are detected in the BVP waveform and inter-beat intervals extracted. 6) The HRV spectrogram is computed using the IBIs in successive time windows.

channels (12 bits/channel): red, green, blue, cyan, and orange (RGBCO) [6, 7].

Figure 1 shows an overview of the method for calculating the HRVS from a facial video sequence. First, the LEAR [13] facial landmark detector was applied on each frame to find the x- and y- coordinates of the facial feature points.

Using the feature points, a rectangular ROI was determined for each frame (Fig. 1, Step 1). Next, the average values of each color channel in the ROI were calculated. The resulting values form the temporal observation signals (Fig. 1, Step 2). Independent component analysis (ICA) was applied to the observed color channel signals and five source signals extracted (Fig. 1, Step 3). Each of the source signals was band-pass filtered using a Hamming window filter with low- and high frequency cut-offs at 45 beats-per-minute (bpm) (0.75Hz) and 180 bpm (3Hz) respectively. These cut-off frequencies were chosen to reflect the natural lower and upper limits in heart rate. The appropriate source signal was selected by calculating the normalized fast Fourier transform (FFT) of each source and choosing the source signal with the greatest

FFT peak within the range 45 - 180 bpm. This signal was designated as the recovered BVP waveform (Fig. 1, Step 4). Next, the individual pulse peaks within the BVP, and the time intervals between successive peaks (known as the inter-beat intervals (IBIs)), were detected from the selected and filtered BVP waveform (Fig. 1, Step 5). Finally, the HRV spectrogram was obtained by calculating the power spectral density of the IBIs using sequential overlapping windows (Fig. 1, Step 6).

Figure 2 shows the specific details of the peak detection (Fig. 1, Step 5). First, the source signal was divided into subsections in the time domain. The local maxima, and their respective indexes, were extracted from each of the time intervals. Next, pulse wave peaks were selected from the local maxima according to the following three criteria.

1) Local peak judgment: Each local maximum was compared with the previous and subsequent value in order to find peaks (the maximum had to be greater than both the previous and subsequent values in order to satisfy this criteria). See Figure 2 Step 5-1.

2) Peak value judgment: The local maxima were evaluated by comparing with a predefined threshold (each maximum had to be greater than the threshold to satisfy this criteria). See Figure 2 Step 5-2.

3) Time interval judgment: The candidates for peaks which satisfy the above two conditions were evaluated by comparing the time intervals between them with a predefined time threshold (the time interval had to be greater than the threshold to satisfy this criteria (i.e. the peaks could not be too close together)).

Peaks within the BVP signal were selected using the three criteria above. The time intervals between the peaks were calculated to form the inter-beat interval (IBI) signal which shows how heart beat intervals change over time. The time intervals of each peak are sampled unevenly as there are not regular IBIs. In order to apply a frequency analysis with unevenly sampled data the Lomb-Scargle periodogram [8-11] was used for the calculation of the power spectrum (Fig. 1, Step 6). The Lomb–Scargle periodogram (spectral power as a function of angular frequency $\omega \equiv 2\pi f > 0$) is defined by:

$$P_N(\omega) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_{j=1}^{N}(h(j)-\bar{h})\cos\omega(t_j-\tau)\right]^2}{\sum_{j=1}^{N}\cos^2\omega(t_j-\tau)} + \frac{\left[\sum_{j=1}^{N}(h(j)-\bar{h})\sin\omega(t_j-\tau)\right]^2}{\sum_{j=1}^{N}\sin^2\omega(t_j-\tau)}\right\},$$

where the time series of N data points $h(i) \equiv h(t_i)$, $i = 1, \ldots, N$.

$\bar{h}$ and σ^2 are the means and variance of the data def ined by following formulas:

$$\bar{h} \equiv \frac{1}{N}\sum_1^N h(i),\ \sigma^2 \equiv \frac{1}{N-1}\sum_1^N (h(i)-\bar{h})^2. \quad (2)$$

Here the parameter τ is defined by:

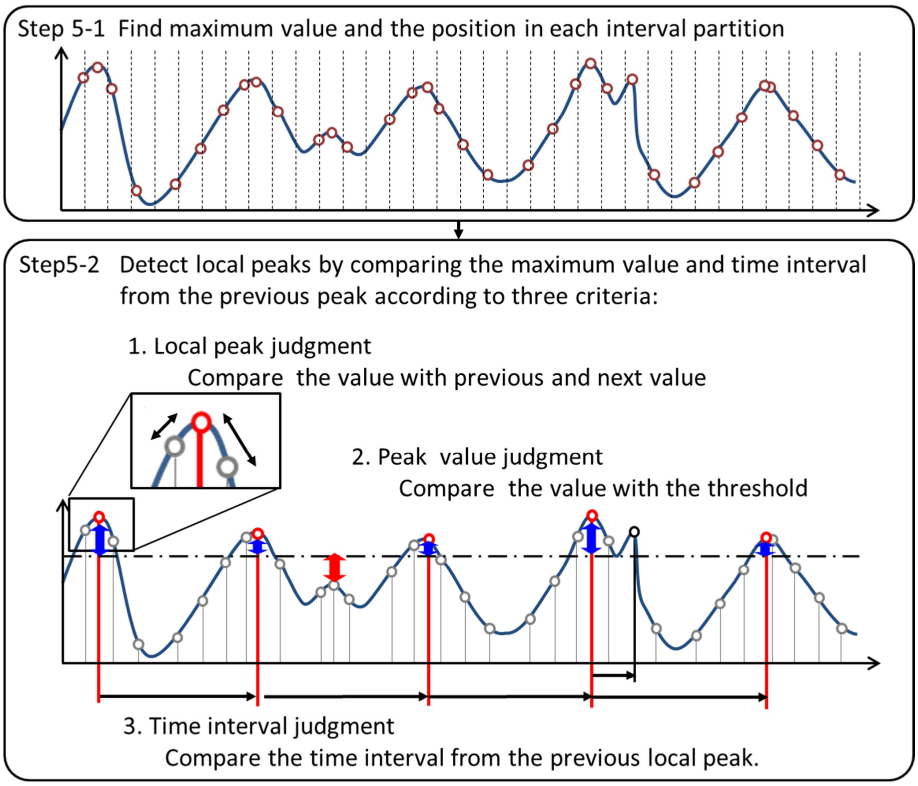

Fig. 2 The existing criteria used for locating peaks within the recovered BVP waveform [5].

$$\tau = \frac{1}{2\omega} tan^{-1}\left(\frac{\sum_{i=1}^{N} sin[2\omega_j t_i]}{\sum_{i=1}^{N} cos[2\omega_j t_i]}\right) . \qquad (3)$$

3 Our Improvements

Below we describe our proposed method and highlight how it offers improvements for heart rate variability measurement over the previous method.

3.1 ROI Setting

The previous method used rectangular ROIs based on the facial feature points. In that case, ROIs included non-flat and shaded facial areas and were particularly susceptible to artifacts due to facial movement. In addition, the

In this analysis, a 60-second window was used to calc ulate the power spectrum shifting the window with ste ps of one second across the sampling period.

rectangular regions may well capture some of the background when the head is turned. In order to help avoid these problems, we fit the ROI along the facial contours avoiding certain facial features, including the nose and mouth. For this purpose we used facial feature points detected using an alignment method based on regression trees [14] which can detect 64 feature points. Figure 3 shows an example of the ROI segmentation.

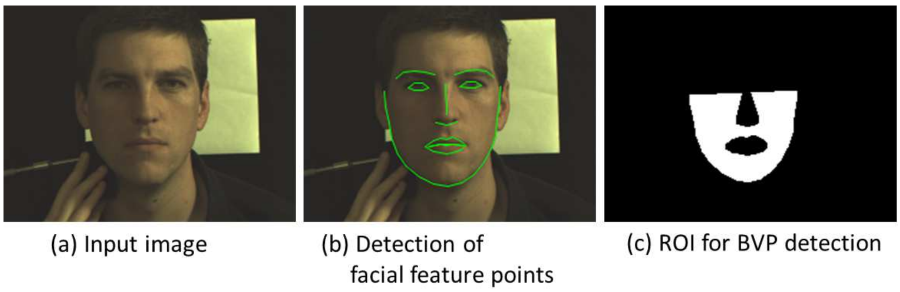

Fig. 3 ROI segmentation: a) the input frame/image, b) the detected facial feature points, c) the region of interest used for BVP detection.

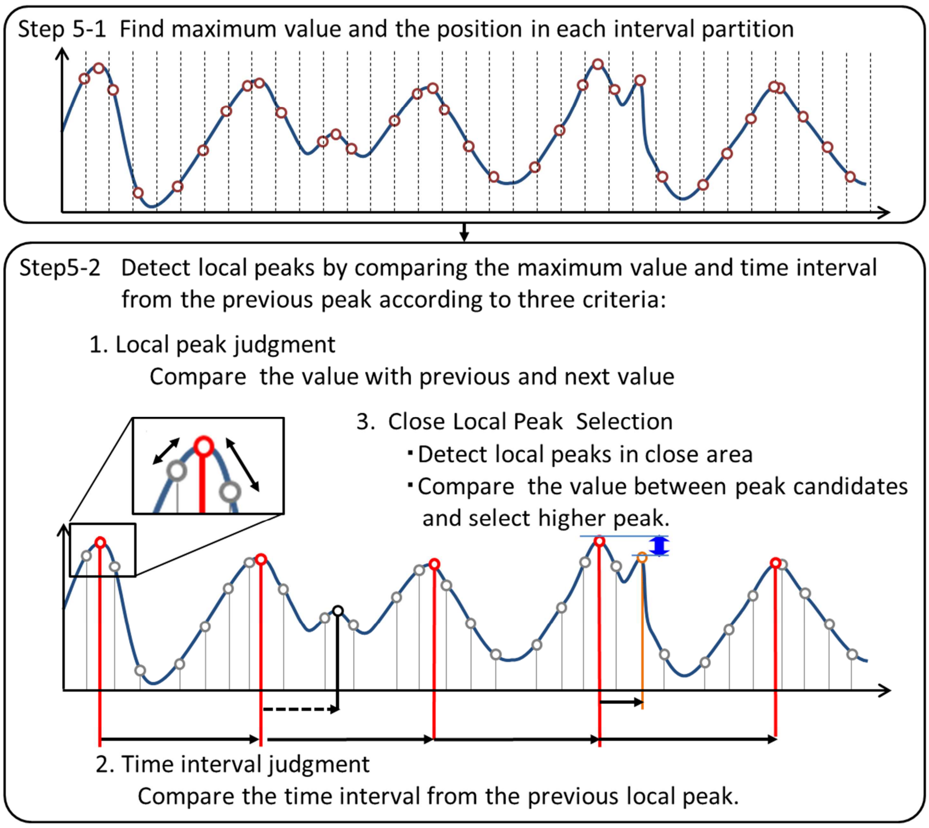

Fig. 4 We proposed a new set of criteria for BVP pulse peak detection. Using our proposed criteria, we compare the value of a set of candidate peaks negating the need for an amplitude threshold.

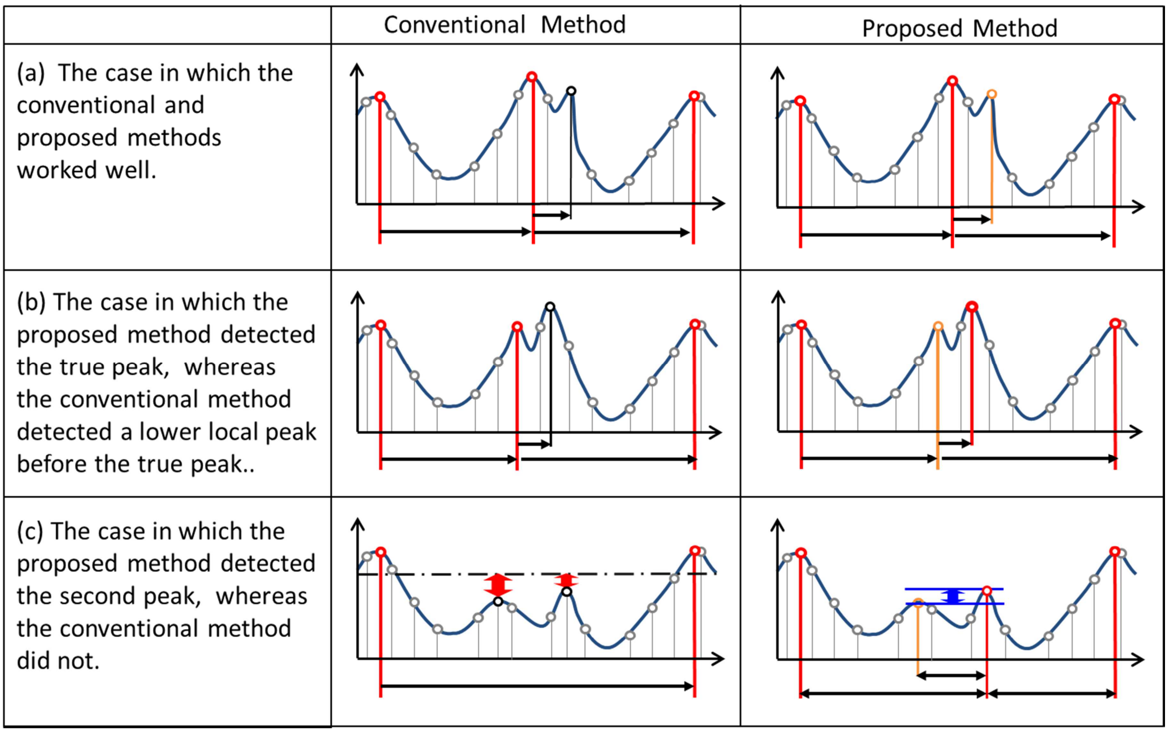

Fig. 5 Shortcoming of the conventional method [5] for peak detection and the improvements of the proposed method.

3.2 Peak Detection of Blood Volume Pulse

Figure 4 shows the criteria of the improved peak detection used to replace Step 5 in Figure 1. First, the source signal is divided into subsections in the time domain. The local maxima, and their respective indexes, are extracted from each of the time intervals (Step5-1). This step is the same as previous method. Next, pulse wave peaks are selected from the local maxima per the following three criteria.

1) Local peak judgment: Each local maximum is compared with the previous and subsequent value to find peaks (the maximum had to be greater than both the previous and subsequent values to satisfy the criteria). See Figure 4 Step 5-1. This criterion is the same as in the previous method.

2) Time interval judgment: The local maxima are evaluated by comparing the time intervals from previous local maxima to next local maxima with a predefined time threshold. The time interval has to be greater than the threshold to satisfy the criteria. This judgement and the next selection criteria are evaluated iteratively.

3) Close Local Peak Selection: We check whether there are other maxima which are close, within the time threshold, to each candidate peak. If there are a multiple maxima the local peak with the highest value is selected as the peak of heartbeat.

Figure 5 shows examples of the shortcoming of the previously published method [5] and the improved solution that our peak detection step provides. Figure 5 (a) shows that the previous method can detect local peaks correctly when the “local peak judgment”, “peak value judgment” and “time interval judgment” criteria act effectively, despite the existence of a noise artifact close to the second true peak. Figure 5 (b) shows that the previous method mistakenly detects the lower peak as the second pulse the since “time interval judgment” only checked the time interval of the candidate from the previous peak. On the other hand, our proposed method detects the higher peak even though there is a local maximum before it. Figure 5 (c) shows that the second and third local peaks cannot be detected, as those values are lower than the threshold for “peak value judgment”. Our proposed method can detect the second and third peaks, since we eliminate the “peak value judgment” criteria and apply the “Close Local Peak Selection” step to select the appropriate local peak by comparing the value of the candidates.

3.3 Heart Rate Variability Calculation

The previous method utilized the Lomb–Scargle periodogram to analyze the frequency domain characteristics of the heart rate variability (HRV). To enhance the robustness against misdetection of local peaks and the time intervals between them, we use the Welch periodogram for the analysis of HRV.

The Welch periodogram for a time series of N data points $h(n)$, $n = 1, \dots, N$ is defined by:

$$\hat{P}_W(\omega) = \frac{1}{K}\sum_{k=0}^{K-1} \hat{P}_M^{(i)}(\omega) \quad (4)$$

where $\hat{P}_M^{(i)}(\omega)$ is modified periodograms given by:

$$\hat{P}_M^{(i)}(\omega) = \frac{1}{LU}\left|\sum_{n=0}^{L-1} w(n)h_i(n)e^{-jn\omega}\right|^2 \quad (5)$$

$w(n)$ is data window, L is the window length, $\mathrm{U} = \frac{1}{L}\left|\sum_{n=0}^{L-1} w(n)\right|^2$ is a constant and $h_i(n)$is sampled data given by:

$$h_i(n) = h(n + iD) \quad (6)$$

where D is step size. We used hamming window for the data window.

In the Welch periodogram, the signal is divided into K fragments of length L (overlapping: 50% in our case) and the resulting periodograms from these segments are averaged. The averaging of modified periodograms tends to decrease the variance of the estimate relative to a single periodogram estimate of the entire data record.

In the process of HRV spectral analysis, a 30-second window was used to calculate the power spectrum calculated by the Welch periodogram shifting the window with step sizes of one second across the sampling period. Whilst HRV is typically calculated over longer signals (at least one minute) a 30 second window is sufficient here to demonstrate our method.

4 Experiments

4.1 Experimental Setup

Figure 6 shows the experimental apparatus we used to collect data. The video data of a subject's face were taken from a distance of 4 meters with a digital single-lens reflex (DSLR) camera with a five color channel CMOS sensor (RGBCO) (12 bits/channel) [7]. We performed experiments to measure the subjects at rest and under cognitive stress. The frame rate of the camera was 30 frames per second (fps). Each frame was 640 × 480 pixels. A standard Zuiko 50 mm lens was used in our experiment. Each frame was saved on a laptop PC (Dell Inc. Latitude E6530, 2.4 GHz, 3 MB cache). An artificial solar light was placed at a distance of 2.0 m from the subject.

In the experiments, we obtained videos from three participants (two Asian and one Caucasian). The experiments were conducted under two conditions for each subject. First, subjects were measured at rest (not under cognitive stress). Second, subjects were measured under cognitive stress; the subjects were required to keep doing mental arithmetic exercises. The duration of each experiment was 60 seconds.

4.2 Experimental Results

Figure 7 shows examples of the HRV spectrograms recovered from the camera measurements for one of the participants at rest. Each spectrogram is shown using a heat map. Red indicates high power, and blue indicates low

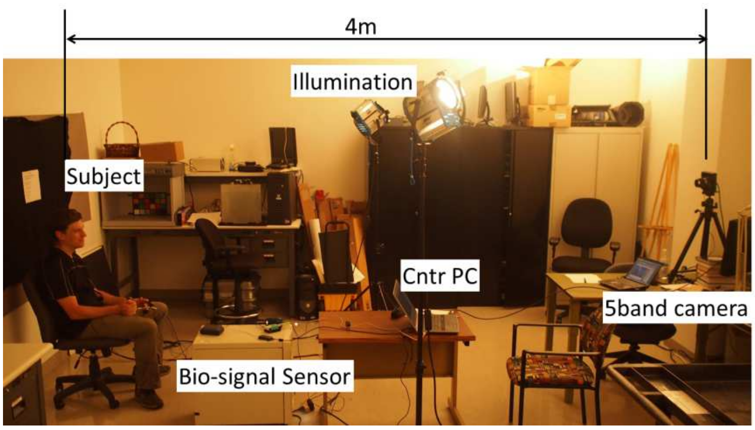

Fig. 6 Experimental apparatus used for data collection. The subject was seated four meters from the five-band digital single lens reflex camera. An artificial solar light was used for illumination.

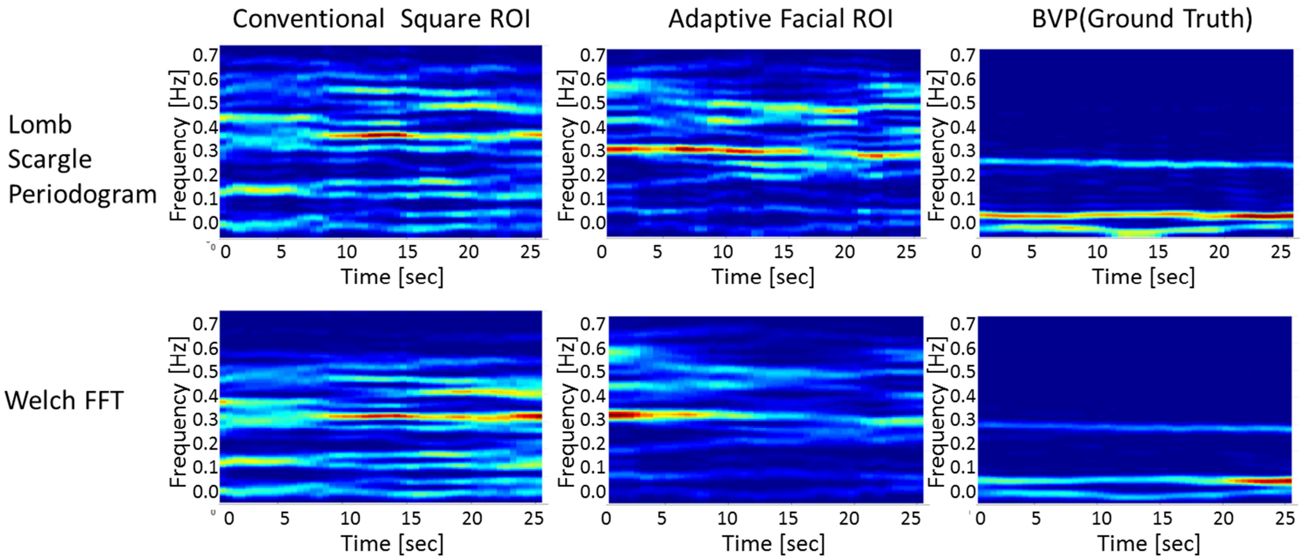

Fig. 7 Heart Rate Variability (HRV) spectrograms for a participant at rest.

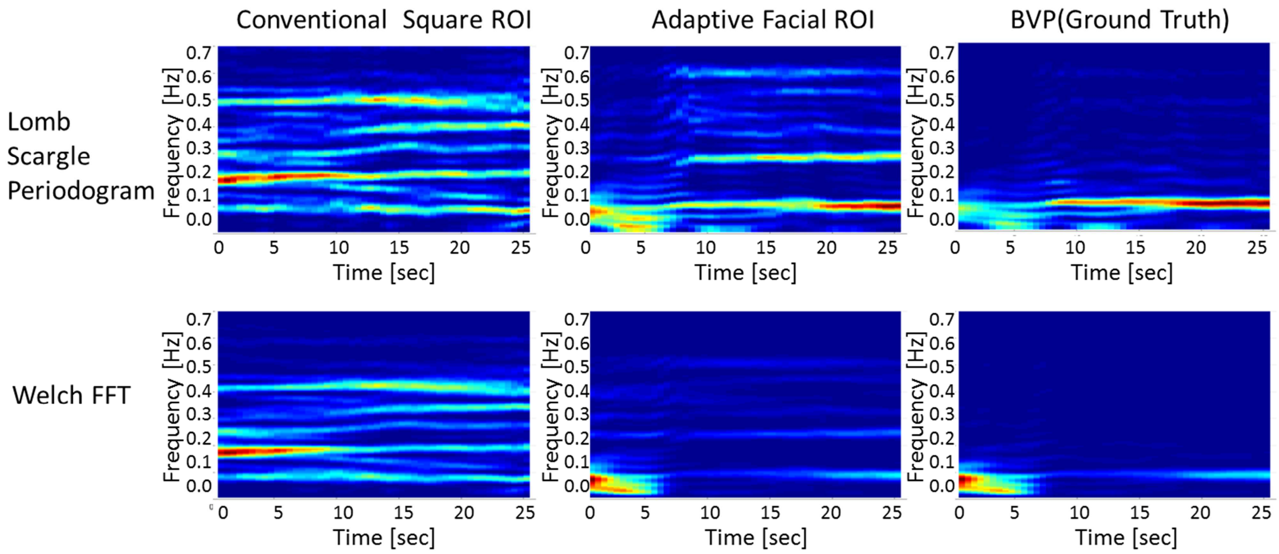

Fig. 8 Heart Rate Variability (HRV) spectrograms for a participant under cognitive stress.

power. The top row shows the results calculated using the conventional method [5], the bottom row shows the results calculated by the proposed method. The left column shows the results using a rectangular ROI, the middle column shows results of adaptive facial ROI and the right column shows the results calculated with a contact BVP sensor (the gold-standard PPG comparison). For the measurements at rest the results using the adaptive facial ROI show a clear peak in HF power (0.15-0.40 Hz) around 0.3 [Hz] which is also the case in the gold-standard comparison. Whereas the results using the fixed rectangular ROI show noisy frequency components in a wide range of frequency bands. Our proposed method (bottom middle) has fewer noisy frequency components than that of the conventional method (upper middle). In Figure 7 the difference between the gold-standard (bottom right) and our proposed method (bottom middle) is the absence of LF components detected using the proposed method. There is still some room for improvement for remote recovery of the HRV spectrogram.

Figure 8 shows examples of the HRV spectrograms recovered from the camera measurements for the participant under cognitive stress. The result for our proposed method (bottom middle) has fewer noisy frequency components than that of conventional method (upper middle). The result obtained using our proposed method (bottom middle) is very similar to the gold-standard measurement (bottom right).

5 Conclusion

We have proposed a method for video-based estimation of heart rate variability (HRV) spectrograms using a five-band camera (RGBCO: red, green, blue, cyan, orange). The method shows qualitative improvement in HRV measurement over previous work [5]. The proposed method consists of three advancements: (i) an adaptive non-rectangular region of interest selection using automatically detected facial feature points, (ii) improved peak detection from the blood volume pulse (BVP) signal, (iii) improved HRV calculation using the Welch FFT method.

We apply our proposed method to test data of subjects at rest and under cognitive stress. We observed that the proposed method could provide cleaner observations of HF components (0.15-0.40 Hz) and the results were similar to the HRV spectrograms calculated using a BVP contact sensor (the gold-standard comparison).

We acknowledge there are some limitations to this study. Our experiments were conducted in an environment with no changes in illumination and no large head motions of the subjects. Evaluation under a wider variety of conditions is necessary for practical use. We only evaluated results on a limited number of subjects in this experiment and performed a qualitative evaluation. Subsequent work will confirm the performance over a larger number of participants and more diverse conditions.

REFERENCES

[1] Verkruysse W, Svaasand LO, Nelson JS (2008) Remote plethysmographic imaging using ambient light. Opt Express 16(26):21434–21445

[2] Poh M-Z, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10762–10774

[3] M-Z. Poh, D. McDuff, R.W. Picard. "Advancements in noncontact, multiparameter physi-ological measurements using a webcam." IEEE Transactions on Biomedical Engineering vol.58, no.1, pp.7-11, (2011).

[4] Pagani M, Furlan R, Pizzinelli P, Crivellaro W, Cerutti S, Malliani A (1989) Spectral analysis of R–R and arterial pressure variabilities to assess sympatho-vagal interaction during mental stress in humans. J Hypertens 7(Suppl):S14–S15

[5] McDuff D, Gontarek S, Picard RW (2014) Improvements in remote cardio-pulmonary measurement using a five band digital camera. IEEE Trans Biomed Eng 61(10):2593–2601

[6] Monno Y, Tanaka M, Okutomi M (2012) Multispectral demosaicking using guided filter. In: IS&T/SPIE electronic imaging. International society for optics and photonics, pp 82 990O-82 990O

[7] MonnoY Kikuchi S, Tanaka M, Okutomi M (2015) A practical one-shot multispectral imaging system using a single image Sensor. IEEE Trans Image Process 24(10):3048-3059

[8] Lomb N , (1976) Least-squares frequency analysis of unequally spaced data. Astrophys. Space Sci., vol. 39, no. 2, pp. 447-462

[9] Scargle JD (1982) Studies in astronomical time series analysis II-statistical aspects of spectral analysis of unevenly spaced data. Astrophys J 1:835-853

[10] Press William H, Rybicki George B (1989) Fast Algorithm for Spectral Analysis of Unevenly Sampled Data. Astrophys J 338:277-280

[11] Schulz M and Stattegger K (1997) Spectrum: Spectral analysis of unevenly spaced paleoclimatic time series. Comput. Geosci., vol. 23, no. 9, pp.929-945

[12] Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio and Electroacoustics, 15(2):70-73

[13]Martinez B, Valstar M F, Binefa X and Pantic M (2013) Local evidence aggregation for re-gression-based facial point detection. IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 5, pp. 1149-1163

[14] Kazemi V and Sullivan J. (2014) One millisecond face alignment with an ensemble of regression trees. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1867-1874

[15] G. De Haan, V. Jeanne. (2013) Robust pulse rate from chrominancebased rPPG. IEEE Trans. Biomed. Engineering, vol.60, no.10, pp.2878–2886

[16] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, Gerard de Haan (2017) Algorithmic Principles of Remote PPG, IEEE Trans. Biomed. Engineering, Vol.64, Issue.7, pp.1479-1491