

Real-Time Approaches for Heart Rate Monitoring using Imaging Photoplethysmography

Sibylle Fallet¹, Leila Mirmohamadsadeghi¹, Virginie Moser², Fabian Braun^{1,2}, Jean-Marc Vesin¹

¹Swiss Federal Institute of Technology, Lausanne, Switzerland

²Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland

Abstract

Imaging photoplethysmography (iPPG) has gained a lot of popularity as a contactless heart rate (HR) monitoring technique. However, most of the existing approaches to estimate HR are based on block-wise processing schemes, which are not optimal for real-time applications. The aim of this study was to investigate robust HR estimation methods having a short estimation delay, which would be suitable for real-time HR monitoring applications using iPPG. The three following algorithms were evaluated: 1) an algorithm based on adaptive sliding-window singular value decomposition (SWASVD), 2) an adaptive band-pass filter (OSC-ANF-W), 3) a notch-filter bank (NFB) estimation method. The database used to evaluate these algorithms was composed of 46 records, acquired in the light using an RGB camera or the dark using an NIR camera. The subjects were asked to perform different tasks to induce HR fluctuations. For the visible/dark sequences, average absolute errors (AAE) of 3.42/5.25, 3.14/4.21 and 3.98/6.02 bpm were obtained for the SWASVD, the OSC-ANF-W and the NFB algorithms, respectively. The corresponding averaged estimation delays were 4 seconds for the SWASVD and OSC-ANF-W, and 3 seconds for the NFB.

1. Introduction

In recent years, imaging photoplethysmography (iPPG) has emerged as a very attractive technology to measure various cardiovascular parameters. This technique makes possible the remote sensing of blood volume changes occurring at each heartbeat from facial images of the subjects. Common PPG devices typically operate at defined wavelengths, but normal ambient light can be used as the illumination source for iPPG applications [1]. Various processing schemes have been proposed to extract the iPPG waveforms from the video sequences of the subjects. In the first step, the raw RGB traces are computed from the spatial averaging of the pixels within a chosen region of interest (ROI). This ROI, usually selected on the subject's

face, can be either fixed or constantly updated according to a tracking mechanism. The resulting traces consist of a mixture of different sources and further processing is required to extract clean iPPG signals. Some methods based on independent component analysis have been proposed to recover the blood volume pulse from the RGB traces [2, 3]. The oscillation of interest can also be isolated by using a band-pass filter [4]. In a second step, the HR is estimated from these iPPG signals. For instance, the HR can be calculated from the spectrum of the iPPG signal [1], from the local maxima of the iPPG waveform [2, 3] or by identifying the poles corresponding to the HR on an auto-regressive model [4]. Although promising results were obtained, it should be noted that the aforementioned strategies are based on block-wise implementations with considerable processing delays (from 7.5 seconds to one minute). For medical applications, such as monitoring in the intensive care unit, this delay should be as short as possible.

This study aims at reporting the performance of three different algorithms that could be employed for real time HR monitoring applications using the iPPG contactless technology. These algorithms are: 1) an algorithm based on adaptive sliding-window singular value decomposition (SWASVD), 2) an adaptive band-pass filter (OSC-ANF-W), 3) a notch-filter bank (NFB) estimation method. In order to assess the performance of these algorithms in challenging scenarios, variations of HR were induced in healthy subjects by an isometric handgrip exercise or by a protocol with modulated respiration. In addition, the possibility of performing HR monitoring in the dark using an NIR camera and appropriate illumination was also investigated.

2. Methods

2.1. Data

The database is composed of 46 4-minute records, from 12 healthy subjects. For each record, the following data were recorded simultaneously: one-lead ECG and video-

sequence of the upper body region. In order to induce HR variations, the subjects were asked to perform a handgrip isometric exercise or to modulate their respiration according to a given protocol. They were also asked to move the least possible. All the subjects gave informed consent. 24 sequences were recorded in the dark using an NIR camera with an infrared illumination system and 22 sequences were recorded in the artificial light using an RGB camera. The video sequences were sampled at 20 frames per second with a resolution of 1.3 megapixels. The raw traces were obtained by averaging the pixels within a manually determined and fixed rectangular region of interest (ROI) on the forehead. These raw iPPG signals were then band-pass filtered between 0.6 and 4 Hz using an 8th order Butterworth band-pass filter. After visual inspection of the iPPG waveforms derived from the video-sequences recorded in visible light, it was noted that the quality of the green iPPG waveform was generally higher compared to blue and red iPPG waveforms, which is in agreement with [1] (see example in Figure 1). For this reason, iPPG waveforms from red and blue channels were discarded.

The reference HR was derived from the ECG signal. A local maxima detection was first applied in order to detect the R-waves and extract the RR intervals. Then, the RR-intervals were uniformly re-sampled at 4 Hz to compute the true instantaneous HR. This instantaneous HR was then averaged on 4-second windows (3-second overlap) to compute the reference HR. In order to quantify the fluctuations of HR induced by the respiration and handgrip tasks, the HR range was computed for each record as $(HR_{max} - HR_{min})$. Among all records, the average range was 25.4 ± 8.1 beats per minute (bpm).

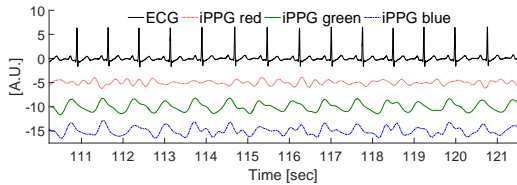


Figure 1. An example of the R/G/B iPPG waveforms. The ECG is displayed on top as a visual reference.

2.2. Heart rate estimation

SWASVD algorithm: The adaptive sliding-window singular value decomposition (SWASVD) algorithm described in [5] was used to process the iPPG signals. This algorithm, derived from the sequential bi-iteration SVD approach, yields a recursive computation of the subspace components using a sliding window. In this study, a 50-sample causal sliding window (step size of one sample) and six subspaces were used. At each time step, the frequency of each subspace component $(f_1(t), \dots, f_6(t))$ was

computed from the DFT of the matrix with columns converging to the right dominant singular vectors. In order to enhance the precision of the frequency estimate, a two-step procedure was used. First, the frequency bin corresponding to the largest DFT sample was extracted. Then, an iterative frequency estimation was performed, as proposed in [6]. Our observations showed that, most of the time, the oscillation of interest (i.e. reflecting the blood volume changes) was contained in the first component of the decomposition. However, in some situations (artifacts, bad-quality waveforms), it corresponded to the second component. For this reason a supplementary mechanism was developed to automatically switch to the second component f_2 in such cases. More specifically, at each time step the median of f_1 for the previous 250 samples was computed. Then, if the value $f_2(t)$ was closer to this median value compared to $f_1(t)$ the algorithm automatically switched to $f_2(t)$.

OSC-ANF-W algorithm: The OSC-ANF algorithm, described in [7], is based on a time-varying band-pass filter, constantly updated to track the instantaneous frequency of the input signal. The adaptive mechanism, derived from the oscillator equation, is designed to maximize the oscillatory behavior of the signal. The resulting cost function is used to update the central frequency of the filter at each time step, such that it tracks the main frequency of the signal. A multi-signal version of this algorithm (OSC-ANF-W) has been proposed in [8]. In this case, the same band-pass filter is used for all signals and its central frequency is updated to follow the common instantaneous frequency of the signals. Individual frequency estimates are computed and finally combined into a global frequency estimate. More specifically, the contribution of each input signal is weighted according to a mechanism based on the minimization of the variance of the linear combination of the individual instantaneous frequency estimates. Three parameters can be selected in the OSC-ANF-W algorithm: β ($0 < \beta < 1$), which is related to filter bandwidth, δ ($0 \ll \delta < 1$), a forgetting factor controlling the convergence rate and μ ($0 \ll \mu < 1$), related to the weights. In this study, we set $\beta = \delta = \mu = 0.97$. Our previous observations showed that the robustness of the frequency estimation usually increases when multiple inputs are used [9]. For this reason, the green/NIR channel and smoothed versions of this signal (causal window, filter lengths of 3 and 5 samples) were provided as inputs for the light/dark conditions, respectively. It is important to mention that the tracking is not affected by phase differences between the inputs.

NFB algorithm: This algorithm uses a bank of notch filters to measure the common instantaneous frequency in

several input signals [10]. More specifically, the filter bank is composed of length-3 FIR notch filters, equally spread over a range of discrete frequencies. In this setting, the main frequency of the signal is closer to the notch frequency of the filter resulting in the smallest output. The output-to-input power is computed and then used to calculate a set of weights. The frequency estimate is finally derived from the weighted sum of the notch frequencies. The multiple inputs are combined in a similar scheme to the one used in the OSC-ANF-W. In this study, the green/NIR channel and smoothed versions (causal window, filter lengths of 3 and 5 samples) were provided as inputs for the light/dark conditions, respectively. The number of filters in the filter bank was chosen to be 50 in the frequency band [0.5-5] Hz.

2.3. Performance measurement

In order to assess the performance of the different algorithms, the average absolute error (AAE) was computed between the true and the estimated HR values. For this purpose, HR estimates were averaged on 4-second windows (3-second overlap). A Bland-Altman analysis was also performed to define another measure of the similarity between the true and the estimated HR values. The limits of agreement (LOA) $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ to contain 95% of the differences were calculated. In addition to the aforementioned measures, the estimation delay was computed using the cross-covariance between the true and the estimated HR. The average delay of each algorithm was then used to re-compute the previously described performance metrics for the aligned HR sequences. Occasionally, artifacts on iPPG waveforms were induced by small movements of the subjects. These artifacts were characterized by sudden amplitude increases in the iPPG waveforms. In order to avoid large estimation errors, a method was developed to automatically detect and remove these segments, based on the amplitude ratio of consecutive signal windows.

3. Results

Figure 2 shows an example of the estimated HR for a subject performing the handgrip isometric exercise. An averaged estimation delay of 4 seconds was found for the SWASVD and OSC-ANF-W algorithms and 3 seconds for the NFB algorithm. The AAE values of each sequence have been plotted in Figure 3, for both visible light and dark conditions (delay compensated). The results for the overall AAE and the LOA are shown in Table 1. These performance metrics are reported for both real-time setting and for the aligned HR series. Figure 4 illustrates the automatic detection of bad-quality segments. In total, 1.7% of HR estimates were removed based on this detection.

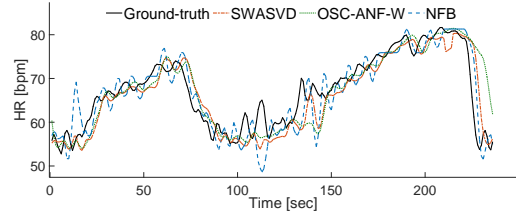


Figure 2. Estimated HR values and ground-truth for a sequence of a subject performing handgrip exercise.

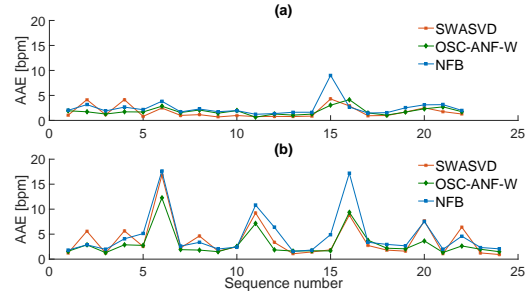


Figure 3. AAE results for all sequences recorded in visible light (a), and in the dark (b). The estimation delay was compensated.

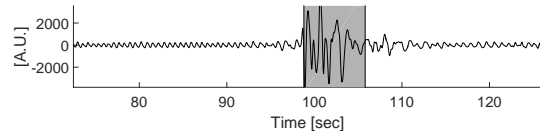


Figure 4. Example of an iPPG segment identified as bad-quality (grey zone)

4. Discussion

Overall results are encouraging and confirm that the investigated algorithms were suitable to estimate HR from the video-sequences of the subjects. In order to better emphasize the accuracy, and given that large HR variations were present in the database, it was important to report the performance also when the estimation delay was compensated. Although better and more constant results were obtained for the video-sequences recorded in visible light, HR monitoring in the dark using an NIR camera and infrared illumination seems conceivable with an overall AAE under 5 bpm (see Table 1 and Figure 3). Regarding the differences between the three evaluated algorithms, it can be noted that the results are in the same range. In terms of AAE, the best overall performance was achieved by the OSC-ANF-W algorithm. More precisely, this algorithm resulted in average AAE of 3.14/4.21 bpm for visible/dark sequences with the real-time setup and 1.85/3.07 bpm when the delay was compensated. It can be noted in Figure 3 that the individual AAE values were very similar

Table 1. Overall results for the three algorithms.

Algorithm		Visible light			Dark		
		SWASVD	OSC-ANF-W	NFB	SWASVD	OSC-ANF-W	NFB
Real-time	mean AAE [bpm]	3.42 ± 1.17	3.14 ± 0.84	3.98 ± 1.49	5.25 ± 3.43	4.21 ± 2.51	6.02 ± 4.00
	LOA 95% [bpm]	[-8.73 10.60]	[-7.79 9.65]	[-11.03 12.60]	[-12.67 18.46]	[-10.66 13.77]	[-17.05 20.97]
Aligned	mean AAE [bpm]	1.69 ± 1.19	1.85 ± 0.78	2.49 ± 1.61	3.88 ± 3.76	3.07 ± 2.71	4.75 ± 4.46
	LOA 95% [bpm]	[-6.99 8.90]	[-6.59 8.50]	[-9.26 10.87]	[-11.76 17.59]	[-9.51 12.66]	[-15.98 19.94]

between the three methods (expected for some recordings that were more challenging for the NFB or SWASVD algorithms), indicating that the three algorithms were sensitive to the same kind of perturbations. The smallest estimation delay was achieved by the NFB algorithm. Concerning the choice of the input channel, the initial idea was to combine the R/G/B channels for the multiple-input algorithms (OSC-ANF-W and NFB). However, it was noticed during the development of the HR processing schemes that better results were achieved with the signals derived from the green channel only, compared to the RGB combination. This is consistent with previous findings, where the correlation between estimated and ground-truth HR was larger for the green channel only than for the RGB [3]. Variable results have been reported in prior works regarding the accuracy of HR estimation in iPPG applications. For instance, an AAE of 6.1 bpm was found using the RADICAL ICA method, for video-sequences recorded with a webcam in stable conditions [11]. In another study, a correlation of 1 was reported between HR derived from video-sequences and from a contact PPG sensor [3]. The pole cancellation method described in [4] resulted in AAE of approximately 3 bpm. Although it is debatable to compare the performance of the different studies (different cameras, different subjects, variable experimental settings and diverse methods to compute error), it allows us to draw the following conclusions: the investigated algorithms meet the current performance requirements and they have the shortest estimation delays [1–4, 11].

5. Conclusion

Three HR estimation methods were evaluated on our database composed of video-sequences of subjects performing various tasks supposed to induce large variations of HR. We have demonstrated that it is possible to achieve good performance while having a short estimation delay, acceptable for real-time monitoring applications. The presence of motion artifacts is a remaining challenge and a face tracking algorithm should be included in further work.

Acknowledgements

This work was funded thanks to the Swiss NanoTera initiative of the Swiss National Science Foundation, RTD

project NewbornCare.

References

- [1] Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. *Opt Express* 2008; 16:21434–21445.
- [2] Poh MZ, McDuff D, Picard R. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans Biomed Eng* 2011;58:7–11.
- [3] McDuff D, Gontarek S, Picard R. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Trans Biomed Eng* 2014;61:2593–2601.
- [4] Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton DA, Pugh C. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiol Meas* 2014;35:807–831.
- [5] Badaeu R, Richard G, David B. Sliding window adaptive SVD algorithms. *IEEE Trans Signal Process* 2004;52:1–10.
- [6] Aboutanios E, Mulgrew B. Iterative frequency estimation by interpolation on Fourier coefficients. *IEEE Trans Signal Process* 2005;53:1237–1242.
- [7] Liao HE. Two discrete oscillator based adaptive notch filters (OSC ANFs) for noisy sinusoids. *IEEE Trans Signal Process* 2005;53(2):528–538.
- [8] Prudat Y, Vesin JM. Multi-signal extension of adaptive frequency tracking algorithms. *Signal Process* 2009;89:963–973.
- [9] Fallet S, Vesin JM. Adaptive frequency tracking for robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise. In *Computing in Cardiology Conference (CinC)*. 2015; 925–928.
- [10] Mirmohamadsadeghi L, Vesin JM. Real-time multi-signal frequency tracking with a bank of notch filters to estimate the respiratory rate from the ECG. *Physiol Meas* 2016; 37(9):1573–1587.
- [11] Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC. Signal recovery in imaging photoplethysmography. *Physiol Meas* 2013;34:1499.

Address for correspondence:

Sibylle Fallet
 EPFL SCI STI JMV - ELD 224 - Station 11
 1015 Lausanne, Switzerland
 sibylle.fallet@epfl.ch