

Imaging Photoplethysmography: A Real-time Signal Quality Index

Sibylle Fallet¹, Yann Schoenenberger¹, Lionel Martin¹, Fabian Braun^{1,2}, Virginie Moser² and Jean-Marc Vesin¹

¹Swiss Federal Institute of Technology, Lausanne, Switzerland

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

Abstract

Imaging photoplethysmography (iPPG) is a promising technology for contactless heart rate (HR) monitoring. However iPPG signals are easily deteriorated by subject movements and illumination changes. The purpose of this study was to develop a signal quality index (SQI) for real-time HR monitoring applications and to assess its performance on a challenging dataset composed of videos of moving subjects. HR was estimated using a multi-input adaptive frequency tracking scheme, in which the iPPG signals derived with different methods and their corresponding SQIs were provided as inputs. Using the proposed SQI, the average absolute error was reduced by 42%/45% when the forehead/entire face region was used to derive iPPG signals, respectively.

1. Introduction

Imaging photoplethysmography (iPPG), also known as remote photoplethysmography (rPPG), was introduced for the first time in 2008 by Verkruyse et al. [1]. It consists of the remote measurement of the cardiac pulse from facial images of the subjects captured with a simple camera, using visible light as illumination source. The most simple technique to compute the iPPG signal from the images is the spatial averaging of the pixels encompassed in a region of interest (ROI). More robust techniques have been proposed to recover the cardiac pulses. Blind-source separation-based methods were used in [2–4]. In [5], a chrominance-based method was proposed. Feng et al. [6] proposed an optical model was built to understand better the origins of the iPPG signal and the impact of subject movements. Based on their findings, the authors developed an adaptive green-red difference method (GRD) to attenuate the effect of motion artifacts. In another study, a promising technique called Spatial Subspace Rotation (SSR) was introduced [7]. In this data-driven approach, the temporal rotation of the RGB subspace of skin pixels is estimated to derive the pulse. Later on, the same authors proposed another approach called Plane-Orthogonal-

to-Skin (POS) [8], in which a plane orthogonal to the skin tone is used for projection, leading to SNR comparable to the one obtained with the SSR method. The use of other color spaces is another aspect that can be studied in the context of the iPPG technology. For example, Tsouri et al. [9] showed that the hue channel leads to lower error when estimating HR, compared to the green channel.

However, despite the considerable progresses in the field, the contactless nature of the iPPG technique makes it very sensitive to disturbances induced by subject movements and illumination changes. Moreover, the iPPG-signal derivation technique leading to the highest SNR depends on experimental conditions [8], which may additionally vary over time. Therefore, a signal quality index (SQI) can be valuable to 1) develop processing schemes involving a dynamic combination of the various reconstructed signals and 2) detect bad-quality epochs during which the estimated HR is not reliable.

2. Methods

Database: The database used in this study includes two subsets: the “static subset”, composed of 22 4-minute video-sequences of immobile subjects performing respiration and handgrip exercises and the “motion subset”, composed of nine 2-minute video-sequences of subjects rotating their head to the left and to the right according to an audio stimulus. Subjects were lying down during the recordings. For each sequence, one-lead ECG and video-sequence of the upper body region, in artificial lighting conditions, were recorded simultaneously. The video sequences, acquired with a commercial RGB camera, were sampled at 20 frames per second with a resolution of 1.3 megapixels. 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. All the subjects gave informed consent.

Calculation of iPPG signals: Our previous investigations showed that the forehead is the best region on the face, in terms of the fraction of power at HR [10]. Therefore, for each sequence, the ROI was selected on the subject’s forehead. For the motion subset, the tracker described in [11] based on the algorithm described in [12] was used. In this algorithm, the adaptive tracking-by-detection is based on structured output prediction achieved using an SVM learning framework. A budgeting mechanism is used to limit the number of support vectors and allows the tracker to run at high frame rates. In addition, a skin-segmentation step was performed by selecting pixels whose hue and saturation match known skin color. It should be mentioned that, for some sequences, the fast rotations of the head were challenging for the tracker, which had difficulties to follow. As a consequence, some frames without skin pixels were dropped for the computation of iPPG signals. For this reason, all the analyses of the motion subset were also performed for an ROI encompassing the whole face.



Figure 1. Example to illustrate a typical tracker output for the forehead ROI (on the left) and the corresponding skin segmentation mask (on the right). White pixels of the segmentation mask correspond to pixels detected as skin.

Different techniques were used to derive the iPPG time series from the skin pixels of the ROI and the resulting signals were further combined in order to increase the robustness of HR estimation. The following iPPG signals were computed:

- iPPG_{green}: Spatial averaging of pixels for green channel, and band-pass filtering between 0.6 and 4 Hz.
- iPPG_{hue}: Transformation of RGB space to HSV, spatial averaging of pixels for hue channel and band-pass filtering between 0.6 and 4 Hz.
- iPPG_{GRD}: Adaptive green-red difference [6].
- iPPG_{SSR}: Spatial subspace rotation, using a sliding window of length $l = 20$ samples [7].
- iPPG_{POS}: Plane-orthogonal-to-skin, using a sliding window of length $l = 20$ samples [8].

SQI computation: The frame-to-frame average absolute difference between pairs of corresponding pixels in the ROI is first computed as follows:

$$DI[n] = \frac{1}{N_{pix}} \sum_{k=1}^{N_{pix}} |pix_k[n] - pix_k[n-1]| \quad (1)$$

with $pix_k[n]$, the green channel value of the k^{th} pixel of the ROI, at frame n . Sudden surges of this quantity are correlated with various disturbances such as motion, illumination changes and camera occlusions, as shown in Figure 2. When a sample is associated with a significant DI increase, empirically defined as a current DI value at least five times larger than the median DI over the past 50 samples, the sample is labeled as a candidate for the beginning of a bad-quality region (i.e. SQI = 0). Besides, it was observed that the iPPG signals were not all affected in the same way by the disturbances. Therefore, the final step of the SQI calculation was designed to be signal-dependent. Two configurations were considered:

- iPPG_{green}, iPPG_{hue} and iPPG_{GRD}: For each labeled candidate, the SQI is set to zero for the next five seconds.
- iPPG_{POS}, iPPG_{SSR}: As shown in [8], these signals are relatively resistant against disturbances. Our observations showed that bad-quality regions were associated with local amplitude increases of these iPPG signals. For every labeled candidate, the presence of such an increase is checked using the following rule. The amplitude before the labeled candidate is first computed as: $amp_{before} = \max\{x[n-win], \dots, x[n]\} - \min\{x[n-win], \dots, x[n]\}$ with $win = 20$ samples and x , the iPPG signal of interest. After 10 samples, this amplitude is compared with the current amplitude, computed in the same way (using a causal window of 20 samples). If the ratio $amp_{current} / amp_{before}$ is larger than 1.6 (empirically selected threshold), the SQI is set to zero for the current sample.

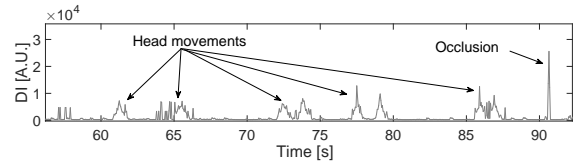


Figure 2. Example illustrating the time evolution of the DI quantity for a segment of a sequence taken in the motion subset.

Heart rate estimation: Instantaneous frequency estimation was performed using a modified version of the multi-signal adaptive frequency tracking algorithm described in [13], which is an extension of the algorithm introduced in [14]. This algorithm 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. In the multi-signal case, the same band-pass filter is used for the M input signals and its central frequency is updated to follow their common instantane-

neous frequency. Individual frequency estimates are computed and finally combined into a global frequency estimate. As described in [13], three parameters can be selected in the multi-signal extended OSC-MSE algorithm: β ($0 < \beta < 1$), which is related to filter bandwidth, δ ($0 \ll \delta \leq 1$), a forgetting factor controlling the convergence rate and λ ($0 \ll \lambda \leq 1$), related to the weights. In the present study, a modification was introduced to make the forgetting factors (δ and λ) time-varying and signal-dependent. Specifically:

$$\delta_m[n], \lambda_m[n] = \begin{cases} 1, & \text{if } SQI_m[n] = 0 \\ 0.95, & \text{otherwise} \end{cases} \quad (2)$$

with m ($0 < m \leq M$). This additional mechanism makes it possible to minimize the contribution of a bad-quality signal in the filter central frequency update mechanism. If all the $SQI_m[n]$ are equal to zero at a given time, the corresponding frequency estimate is not taken into account and labeled as not reliable. All the computed iPPG signals, as well as their smoothed versions (moving average using $L=3$ and 5 samples), were provided as inputs. It is important to mention that the frequency tracking scheme used is not affected by phase differences between the inputs.

Performance measurement: In order to investigate the potential benefits of the proposed SQI, HR was estimated with and without SQI. As for the ground-truth, HR estimates were averaged on 4-second windows (3-second overlap). The accuracy of the estimated HR was quantified with the average absolute error (AAE), computed between the true and the estimated HR values. A Bland-Altman analysis was also performed to compute the limits of agreement (LOA) [$\mu - 1.96\sigma, \mu + 1.96\sigma$] to contain 95% of the differences. The percentage of removed data points, considered as not trustworthy (i.e. when all the $SQI_m[n]$ were equal to zero) is also reported.

3. Results

The overall results for the two data subsets are shown in Table 1. The performance metrics are reported for both real-time setting and for the aligned HR series, for which the averaged estimation delay of 4 seconds was compensated. The individual AAE values are plotted in Figure 4. For the motion subset, using the forehead ROI, the error was reduced by using the SQI for all the sequences and an average error reduction of 42% was achieved. A fraction of 24% of data was considered as not reliable. When the entire face was used as ROI, the error was reduced for 8 sequences by using the SQI and an average error reduction of 45% was achieved, with 13% of data considered as not reliable. Figure 3 is an example of the time evolution of the proposed SQI for a data segment taken in the motion

subset, for the iPPG_{SSR} signal.

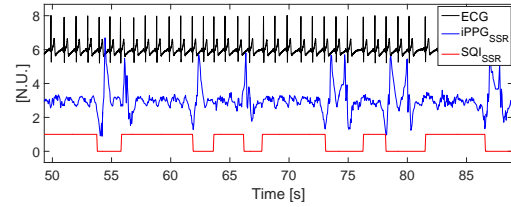


Figure 3. Example of the computed SQI for a segment of a sequence taken in the motion subset (forehead ROI). The ECG is also displayed as visual reference.

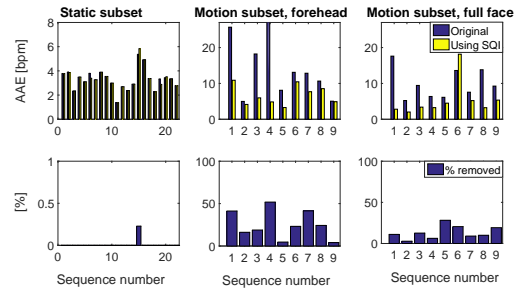


Figure 4. Individual AAE values and percentages of data considered as not reliable.

4. Discussion

The accuracy of HR estimation was considerably increased by the addition of the SQI for the motion subset, while the results for the static subset remained almost unchanged. This can be explained by the small amount of subject movements in these videos. The overall error of about five bpm for the motion subset is encouraging given the challenging nature of this dataset.

Regarding the computation of the SQI, it should be mentioned that SQI metrics have been proposed already in the context of traditional PPG signals. Li et al. [15] proposed an SQI based on dynamic time warping for pulsatile signals (PPG and ABP). Unfortunately, it seems difficult to adapt this approach to iPPG signals, because morphology of the latter is very variable. We believe that the proposed method based on the DI value has some strengths. First, it only requires the pixel values at the current frame and at the frame $[n-1]$ and is very simple to compute. Second, this quantity raises with all kinds of disturbances that are likely to deteriorate the quality of the iPPG signals. For example, the time evolution of the position of the tracked ROI could have been used instead. However, sudden illumination changes would not be detected. Finally, the fact that it led to similar results for the two types of ROI is also a good point.

Table 1. Overall results.

		Static subset (forehead ROI)		Motion subset (forehead ROI)		Motion subset (full face)	
		Original scheme	SQI-based	Original scheme	SQI-based	Original scheme	SQI-based
Real-time {	mean AAE [bpm]	3.28 ± 0.86	3.27 ± 0.92	16.46 ± 14.04	6.74 ± 2.77	9.89 ± 4.23	5.32 ± 4.94
	LOA 95% [bpm]	[-8.62 10.05]	[-8.64 10.00]	[-27.79 32.14]	[-16.06 19.26]	[-20.16 29.35]	[-14.31 19.51]
Aligned {	mean AAE [bpm]	1.92 ± 0.89	1.89 ± 0.93	11.57 ± 6.21	6.21 ± 2.55	9.85 ± 4.32	4.72 ± 5.27
	LOA 95% [bpm]	[-5.75 7.08]	[-5.68 6.95]	[-24.75 28.36]	[-14.93 18.41]	[-20.41 29.70]	[-13.93 19.24]
Percentage removed		0	0.01	0	23.65	0	13.30

In addition to the proposed SQI, we introduced in this study a complete HR estimation scheme taking into account signal quality and integrating different state-of-the-art methods. This approach improves both the robustness of HR estimation and the fraction of usable data. The lack of labels for waveform quality was a limitation of this study. For future work, it would be interesting to test this SQI on videos with subjects performing smoother movements, in time-varying lighting conditions.

5. Conclusion

An iPPG-SQI was presented in this study, as a tool to improve the reliability of iPPG-based HR monitoring applications. The computation of this SQI is causal, and thus integrable in real-time HR monitoring application based on iPPG. We believe that the importance of quality measures for biomedical signals becomes even more important nowadays with the introduction of new cardiovascular monitoring tools, that are usually minimally invasive or contactless, and thus resulting in signals that are more prone to artifacts.

Acknowledgements

This work was funded by 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(26):21434–21445.
- [2] Poh MZ, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express* 2010;18(10):10762–10774.
- [3] Poh MZ, McDuff D, Picard R. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans Biomed Eng* 2011;58:7–11.
- [4] 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.

- [5] de Haan G, Jeanne V. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Trans Biomed Eng* 2013; 60(10):2878–2886.
- [6] Feng L, Po LM, Xu X, Li Y, Ma R. Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin. *IEEE Trans Circuits Syst Video Technol* 2015;25(5):879–891.
- [7] Wang W, Stuijk S, de Haan G. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Trans Biomed Eng* 2016;63(9):1974–1984.
- [8] Wang W, den Brinker AC, Stuijk S, de Haan G. Algorithmic Principles of Remote PPG. *IEEE Trans Biomed Eng* 2017; 64(7):1479–1491.
- [9] Tsouri GR, Li Z. On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras. *J Biomed Opt* 2015;20(4):048002–048002.
- [10] Fallet S, Moser V, Braun F, Vesin JM. Imaging photoplethysmography: What are the best locations on the face to estimate heart rate? In *2016 Computing in Cardiology Conference (CinC)*. September 2016; 341–344.
- [11] Willemin F, Vanderghyest P, Paratte J. Robust visual tracking using feature selection. *EPFL Master thesis* 2014; <https://infoscience.epfl.ch/record/201836?ln=en>.
- [12] Hare S, Golodetz S, Saffari A, Vineet V, Cheng MM, Hicks SL, Torr PHS. Struck: Structured Output Tracking with Kernels. *IEEE Trans Pattern Anal Mach Intell* 2016; 38(10):2096–2109.
- [13] Prudat Y, Vesin JM. Multi-signal extension of adaptive frequency tracking algorithms. *Signal Process* 2009;89:963–973.
- [14] Liao HE. Two discrete oscillator based adaptive notch filters (OSC ANFs) for noisy sinusoids. *IEEE Trans Signal Process* 2005;53(2):528–538.
- [15] Li Q, Clifford GD. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiol Meas* 2012;33(9):1491–1501.

Address for correspondence:

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