More Reliable Remote Heart Rate Measurement by Signal Quality Indexes

Hannes Ernst, Hagen Malberg, Martin Schmidt

Institute of Biomedical Engineering, TU Dresden, Germany

Abstract

Accuracy of camera-based heart rate (HR_{cb}) measurement is often impaired by artifacts, which leads to erroneous HR_{cb} and reduced confidence in the measurement.

To avoid erroneous HR_{cb}, we investigated six signal quality indexes (SQIs) from the literature in terms of their effect size and combined them to a novel SQI-filter. All analyses were performed on the "Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database" (BP4D+) in three important color channels.

Signal-to-noise ratio, average maximum cross correlation of consecutive segments, and relative difference of spectral peaks were the most powerful SQIs. The SQI-filter increased accuracies of all color channels. Largest improvements (up to 60%) were achieved in the green channel resulting in 80% accuracy. The overall highest accuracy of 84% was reached in the hue channel. Motion-rich videos benefited most from the developed SQI-filter.

The presented methodology helps to discard distorted signals. This enables more reliable HR_{cb} data in further applications and increases confidence in the measurement.

1. Introduction

Camera-based photoplethysmography (cbPPG) is a non-contact optical technique for haemodynamic monitoring [1]. The characteristic of contactlessness offers many advantages, for example in terms of disinfection and patient comfort. On the other hand, cbPPG is particularly susceptible to unsteady conditions. It is often reported that signals are heavily distorted by patient motion or transient illumination [1]. In these cases, vital parameter extraction involves the risk of not analyzing physiological information at all. For this reason the idea of examining the quality of the analyzed signal was developed and several signal quality indexes (SQIs) have been introduced. The application of SQIs is not limited to cbPPG. For example, Li et al. [2] used SQIs on electrocardiograms (ECGs) and arterial blood pressure signals for robust estimation of the heart rate (HR) from multiple noisy channels. In [3], the quality of ECG was rated into five groups with a support vector machine and 13 SQIs. Quality of a HR estima-

tion algorithm is often assessed as accuracy in comparison with a gold standard (e.g. HR from ECG). However, if no gold standard is available, quality must be evaluated directly from the signal. This work outlines six SQIs for cbPPG signals derived from videos for camera-based heart rate (HR_{cb}) measurement. The ability of these SQIs to predict the correctness of HR_{cb} from the signal itself is then tested both statistically and as a classification problem. Since cbPPG is used with different color spaces, all analyses are repeated for three important color channels.

2. Notation

A video V of N frames with P pixels per frame consists of P signals $x_p(n)$ with n as discrete time and step size $\Delta n = 1/f_s$, where f_s denotes the sampling frequency (or frames per second). Let $x(n)$ denote a signal deduced from V. Figure 1 shows a cbPPG signal (average green intensity of middle forehead) before and after filtering as well as the amplitude spectra $|X(k)|$ derived by fast Fourier transform (FFT) $\mathcal{F}$: $X(k) = \mathcal{F}\{x(n)\}$ with frequency step size Δf.
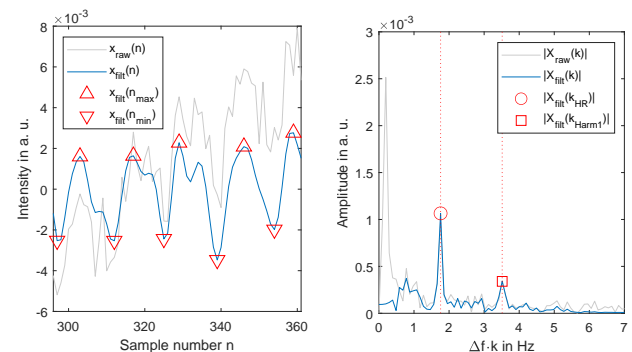
(a) Section of a $10\,\text{s}$ green channel cbPPG signal ($f_s = 25\,\text{Hz}$) in time domain. The raw signal was detrended to oscillate around zero intensity. Filtering was performed with a 10th order Butterworth band-pass ($[0.5, 5]$ Hz). Distolic minima and systolic maxima are marked in red.

(b) Detrended raw and filtered cbPPG signal ($N = 250$) in frequency domain (zero-padding active, $\Delta f \approx 0.10$ Hz). Detrending suppressed low frequency signal components. The highest peak in the amplitude spectrum, which is associated with HR, and its first harmonic are highlighted in red.

Figure 1: Exemplary cbPPG signal with average heart rate of $1.75\,\text{Hz}$ ($105\,\text{bpm}$) in time and frequency domain.

3. Signal Quality Indexes

In general, cbPPG signal quality can be analyzed in time or frequency domain at video or signal level. This work focuses on signal level SQIs, which are listed in Table 1. However, it should be noted that signal level SQIs can be applied to each pixel trace $x_p(n)$ of V as for example Zaunseder et al. [4] did with signal-to-noise ratio (SNR).

Table 1: Overview of considered signal quality indexes.

Domain	Signal Quality Indexes
Time	$sSQI$ [5], $psSQI$ [6], $CCmSQI$ [7] (by $aCCmSQI$ and $sCCmSQI$)
Frequency	$snrSQI$ [8,9], $rdspSQI$ [10]

3.1. Time Domain

Motion or transient illumination cause strong signal deflections. This leads to a high standard deviation of $x(n)$, which is therefore a suitable SQI [5]. With $\bar{x}$ as the mean of $x(n)$, the so-called $sSQI$ is defined as:

$$sSQI = \sqrt{\frac{1}{N-1} \cdot \sum_{n=1}^{N} |x(n) - \bar{x}|^2}. \qquad (1)$$

Trumpp et al. [6] used pulse strength $psSQI$ (systolic-diastolic amplitude difference) as a quality measure. Pulse strength can be estimated beat-wise or from an average beat signal $\bar{b}(n)$ as in eq. 2. Similar to $sSQI$, beat-wise $psSQI$ can be used to identify distorted signal segments as they show high amplitude alterations. Poor blood volume pulsation in the region of interest of V leads to low average $psSQI$ because the optical characteristics vary less intense and the signal is therefore less deflected. [6]

$$psSQI = \bar{b}(n_{max}) - \bar{b}(n_{min}) \qquad (2)$$

If signal quality is high, cbPPG is quasi-periodic. Feng et al. [7] used the maximum of a normalized cross correlation function of consecutive signal segments (length $2\,\mathrm{s}$, shift limited to $\pm 0.5\,\mathrm{s}$) to assess signal quality. The cross correlation function $CC_{x1x2}(m)$ for two real signals $x_1(n)$ and $x_2(n)$ of equal length N is defined as:

$$CC_{x1x2}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_1(n+m) \cdot x_2(n) & m \geq 0, \\ CC_{x1x2}(-m) & m < 0. \end{cases} \qquad (3)$$

The SQI $CCmSQI$ is then derived after normalization as:

$$CCmSQI = \max \frac{CC_{x1x2}(m)}{\sqrt{CC_{x1x1}(0) \cdot CC_{x2x2}(0)}}. \qquad (4)$$

$CCmSQI$ contains a value for each $2\,\mathrm{s}$ segment and can be summarized to $aCCmSQI$ and $sCCmSQI$ by calculating the mean and standard deviation respectively.

3.2. Frequency Domain

Frequency domain SQIs are calculated from $|X(k)|$. They are affected by N and f_s as these parameters determine spectral properties. The most popular SQI is SNR. It was first applied to cbPPG by de Haan and Jeanne [8]. With the mask function $\Pi(k)$, signal power is taken from a band around k_{HR} and its first harmonic, which must either be known from a reference or estimated from cbPPG. Here, the implementation of Rasche et al. [9] is used:

$$snrSQI = 10 \cdot \lg \frac{\sum_{\Delta f \cdot k=30\,\mathrm{bpm}}^{200\,\mathrm{bpm}} \Pi(k) \cdot |X(k)|^2}{\sum_{\Delta f \cdot k=30\,\mathrm{bpm}}^{200\,\mathrm{bpm}} (1 - \Pi(k)) \cdot |X(k)|^2}$$

$$\text{with} \quad \Pi(k) = \begin{cases} 1 & \Delta f \cdot |k_{HR} - k| \leq 5\,\mathrm{bpm}, \\ 1 & \Delta f \cdot |2 \cdot k_{HR} - k| \leq 5\,\mathrm{bpm}, \\ 0 & \text{otherwise.} \end{cases}$$

$$(5)$$

Finally, $rdspSQI$ is the relative difference of the highest ($|X(k_1)|$) and second highest ($|X(k_2)|$) peaks in the amplitude spectrum (eq. 6) [10]. In a high quality signal $k_1 = k_{HR}$ and $k_2 = k_{Harm1}$, and their spectral amplitudes differ clearly leading to high $rdspSQI$. In case of artifacts, reduced $rdspSQI$ can be observed [10].

$$rdspSQI = (|X(k_1)| - |X(k_2)|)/|X(k_1)| \qquad (6)$$

4. Influence of Signal Quality Indexes on Accuracy of Heart Rate Measurement

4.1. Data Material

The performance evaluation of the SQIs was based on the investigation of camera-based HR accuracy ACC presented in [11], which utilizes the "Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database" (BP4D+) [12]. In contrast to [11], the full data base (140 subjects of all skin types, 10 videos under different emotional stimuli per subject, 1392×1040 pixels, $3 \times 8\,\mathrm{bit}$ RGB, $f_s = 25\,\mathrm{fps}$) was analyzed in this study.

4.2. Methods

Red, green, and blue channel cbPPG signals were extracted from the average intensity in the forehead region of interest (in [11] called FHL) and transformed to HSV and NTSC (also known as YIQ) color spaces. Only the green channel G from RGB, the hue channel H from HSV, and the chrominance channel Q from NTSC were processed further, as they achieve the highest accuracy in their color space [11]. The signals were split into $10\,\mathrm{s}$ segments ($5\,\mathrm{s}$ overlap) leading to a total of 8207 segments. Each segment was detrended and filtered with a 10th order Butterworth band-pass filter (range $[0.5, 5]\,\mathrm{Hz}$ as in Figure 1a). The filtered segments were zero-padded before applying a

FFT. k_{HR} was determined from the peak in the amplitude spectrum (see Figure 1b). If HR_{cb} of a segment j was correct (absolute difference to HR from reference (HR_{ref}) smaller than the greater of 5 bpm or 10 % of HR_{ref} [11]), it was labeled $w_j = \circ$, otherwise $w_j = \bullet$. The proportion of $w = \circ$ yields ACC.

All SQIs were computed for each segment from the detrended and filtered signal as described. To generate $snrSQI$ independent of any reference, k_{HR} from cbPPG was used.

The labels w_j were used to group the signal segments in two classes ($w = \circ$, $w = \bullet$). The SQIs of the two groups were compared statistically with the software JASP 0.11.1 by Mann-Whitney U tests that return effect size as rank-biserial correlation r, which may be interpreted as Pearson's r. Furthermore, a classification problem was opened. A decision tree was set up with 5-fold stratified cross-validation using SQIs as predictors, w as response variable, and error rate as objective function for optimization. For evaluation, several parameters were calculated from the confusion matrix of Table 3 as defined in Table 4. It should be emphasized that the predictors do not analyze HR, but shall indicate if a HR_{cb} can be trusted. The goal is to avoid erroneous HR_{cb} by sorting out segments with prediction $w_j = \bullet$, which is called SQI-filtering. Statistical analyses and w classification were performed for each color channel.

4.3. Results

Results of statistical analyses are shown in Table 2. Comparison of HR_{cb} to HR_{ref} led to ACC of 0.50 (G), 0.73 (H), and 0.70 (Q). All SQI effect sizes were small to medium. The overall most powerful SQI was $snrSQI$ (r: G: -0.52, H: -0.55, Q: -0.54), even though $rdspSQI$ provided a slightly stronger effect size (-0.54) in G. $aCCmSQI$ (r: G: -0.51, H: -0.50, Q: -0.50) performed second and $rdspSQI$ (r: G: -0.54, H: -0.43, Q: -0.46) third best. The effect sizes of the SQIs in G were increased or similar to those in H and Q. Compared to H, a decrease in $sSQI$ and $psSQI$ effect sizes was observed in Q.

Results of w classification are shown in Table 3 (confusion matrix) and Table 4 (evaluation parameters). Up to 81 % (G) of erroneous HR_{cb} were detected (Sensitivity). The amount of actually correct HR_{cb} that were classified as erroneous (FP rate) varied with 21 % (G), 9 % (Q), and 5 % (H). Comparable across the color channels were the values of w classification accuracy (80–83 %), error rate (17–20 %), PPV (72–79 %), and NPV (80–84 %). NPV can be interpreted as the new accuracy of HR_{cb} (with SQI-filtering) ACC'. Only segments with prediction $w_j = \circ$ (TN, FN in Table 3) from the cross-validation test sets are taken into account. ACC' showed considerable relative improvement to ACC (G: $+60\,\%$, H: $+15\,\%$, Q: $+19\,\%$).

Table 2: Statistical evaluation of SQIs in the color channels G, H, and Q with group size $N_s(w)$, HR_{cb} accuracy $ACC = N_s(\circ)/(N_s(\circ) + N_s(\bullet))$, and effect size r (rank-biserial correlation).

	RGB-G	HSV-H	NTSC-Q
$N_s(w = \circ)$	4095	6023	5752
$N_s(w = \bullet)$	4112	2184	2455
ACC	0.50	**0.73**	0.70
$r(sSQI)$	$+0.44$	$+0.34$	$+0.11$
$r(psSQI)$	$+0.48$	$+0.35$	$+0.12$
$r(rdspSQI)$	$-\mathbf{0.54}$	-0.43	-0.46
$r(snrSQI)$	-0.52	$-\mathbf{0.55}$	$-\mathbf{0.54}$
$r(aCCmSQI)$	-0.51	-0.50	-0.50
$r(sCCmSQI)$	$+0.26$	$+0.21$	$+0.23$

Table 3: Confusion matrix definition and results for the color channels G (RGB-G), H (HSV-H), and Q (NTSC-Q).

		predicted class	
w		$\bullet$	$\circ$
true class $\bullet$		true positive (TP) G: 3325, H: 1133, Q: 1393	false negative (FN) G: 787, H: 1051, Q: 1062
true class $\circ$		false positive (FP) G: 868, H: 331, Q: 537	true negative (TN) G: 3227, H: 5692, Q: 5215

Table 4: Evaluation of decision tree w classification by color channel. The pairs accuracy* and error rate, sensitivity and FN rate, specificity and FP rate complement each other to one. N_s=TP+TN+FP+FN (see Table 3). PPV: Positive Predictive Value, NPV: Negative Predictive Value.
*Accuracy of w classification is not to be confused with ACC.

Parameter	Definition	RGB-G	HSV-H	NTSC-Q
Accuracy*	(TP+TN)/N_s	0.80	**0.83**	0.81
Error rate	(FP+FN)/N_s	0.20	0.17	0.19
Sensitivity	TP/(TP+FN)	**0.81**	0.52	0.57
FN rate	FN/(TP+FN)	0.19	0.48	0.43
Specificity	TN/(TN+FP)	0.79	**0.95**	0.91
FP rate	FP/(TN+FP)	0.21	0.05	0.09
PPV	TP/(TP+FP)	**0.79**	0.77	0.72
NPV	TN/(TN+FN)	0.80	**0.84**	0.83

The achieved improvement is broken down for the 10 emotional stimuli of BP4D+ for G in Figure 2. The strongest relative increase from ACC to ACC' was reached in stimulus 4 (startle induced by sudden sound burst [12]) in all color channels (G: $+191\,\%$, H: $+53\,\%$, Q: $+65\,\%$). ACC and ACC' peaked in stimulus 3 (sadness induced by video clip of 911 emergency phone call [12]) in all color channels (G: 0.77 and 0.91, H: 0.87 and 0.92, Q: 0.87 and 0.91).
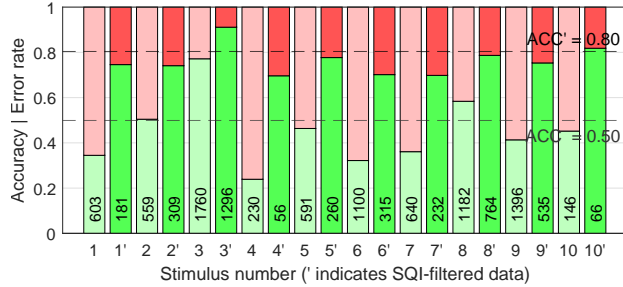
Figure 2: Accuracy (green) stacked with error rate (red) of all data (ACC) and SQI-filtered data (ACC', stimulus numbers with ') from BP4D+ by stimulus for the green channel. Each stimulus targets to an emotion: 1 Happiness, 2 Surprise, 3 Sadness, 4 Startle, 5 Scepticism, 6 Embarrassment, 7 Fear, 8 Physical pain, 9 Anger, 10 Disgust [12]. Rotated numbers show the number of used segments. SQI-filtering improved accuracy across all stimuli.

5. Discussion

Our results show, there is no single SQI that can clearly separate the groups $w = \circ$ and $w = \bullet$. Interestingly, the effect sizes of almost all SQIs are strongest in G, the channel with the lowest ACC. SQIs are not only of greater importance here because more signals have to be rejected, they also perform better in this channel. The combination of SQIs for SQI-based filtering with a decision tree markedly improves ACC across all color channels. Even though sensitivity of H and Q are lower compared to G, the generally higher signal quality in these channels leads to fewer erroneous HR_{cb}. The large increase from ACC to ACC' in G attributes HR_{cb} from G similar confidence as HR_{cb} from H or Q. Even though HR_{cb} from all color channels are equally reliable now, H and Q retain more data than G.

Stimulus-wise analyses confirm the negative impact of motion artifacts on ACC, e.g. by head movement or facial expression. Whilst subjects remain still during the sad phone call of stimulus 3, they startle on the sudden sound burst of stimulus 4. The proposed SQI-filtering increased ACC across all stimuli. The improvement is particularly great for motion-rich videos, whose ACC' is now approaching the level of low-motion videos. This holds across all color channels. It should be noted that the improvement by SQI-filtering does not require any external reference, but can be obtained directly from the signal.

Future investigations should include video level SQIs such as the average frame-wise intensity change or phase maps [4] to exploit spatio-temporal information. Also other signal level SQIs such as distance and relative difference of spectral peaks which depend on Singular Spectrum Analysis [5] could be integrated.

References

[1] Zaunseder S, Trumpp A, Wedekind D, Malberg H. Cardiovascular assessment by imaging photoplethysmography - a review. Biomed Eng Biomed Te Jun. 2018;63(5):617–634.

[2] Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiol Meas Jan. 2008; 29(1):15–32.

[3] Li Q, Rajagopalan C, Clifford GD. A machine learning approach to multi-level ECG signal quality classification. Comput Meth Prog Biomed Dec. 2014;117(3):435–447.

[4] Zaunseder S, et al. Spatio-temporal analysis of blood perfusion by imaging photoplethysmography. In Coté GL (ed.), Opt Diag Sens XVIII: Toward Point-of-Care Diagnostics. SPIE, Feb. 2018; Art. no. 105010X.

[5] Wang W, Balmaekers B, de Haan G. Quality metric for camera-based pulse rate monitoring in fitness exercise. In 2016 IEEE Int Conf Image Process (ICIP). IEEE, Sep. 2016; 2430–2434.

[6] Trumpp A, et al. The value of polarization in camera-based photoplethysmography. Biomed Opt Express Jun. 2017; 8(6):2822–2833.

[7] Feng L, et al. Dynamic ROI based on K-means for remote photoplethysmography. In 2015 IEEE Int Conf on Acoust, Speech and Signal Process (ICASSP). IEEE, Apr. 2015; 1310–1314.

[8] de Haan G, Jeanne V. Robust Pulse Rate From Chrominance-Based rPPG. IEEE Trans Biomed Eng Oct. 2013; 60(10):2878–2886.

[9] Rasche S, et al. Camera-based photoplethysmography in critical care patients. Clin Hemorheol and Micro Jan. 2016; 64(1):77–90.

[10] Gupta P, Bhowmick B, Pal A. Accurate heart-rate estimation from face videos using quality-based fusion. In 2017 IEEE Int Conf Image Process (ICIP). IEEE, Sep. 2017; 4132–4136.

[11] Ernst H, Scherpf M, Malberg H, Schmidt M. Color Spaces and Regions of Interest in Camera-Based Heart Rate Estimation. In 2020 Conf European Study Group on Cardiovascular Oscillations (ESGCO). Jul. 2020; 1–2.

[12] Zhang Z, et al. Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis. In 2016 IEEE Conf Comput Vis Pattern Recogn (CVPR). IEEE, Jun. 2016; 3438–3446.

Address for correspondence:

Dipl.-Ing. Hannes Ernst
Institute of Biomedical Engineering, TU Dresden
Fetscherstr. 29, 01307 Dresden, Germany
hannes.ernst@tu-dresden.de