Improved heart rate detection for camera-based photoplethysmography by means of Kalman filtering

Fernando Andreotti, Alexander Trumpp, Hagen Malberg, Sebastian Zaunseder

Improved Heart Rate Detection for Camera-Based Photoplethysmography by means of Kalman Filtering

Fernando Andreotti, Alexander Trumpp, Hagen Malberg, Sebastian Zaunseder

Institute of Biomedical Engineering
Faculty of Electrical and Computer Engineering
TU Dresden, Dresden, Germany
fernando.andreotti@mailbox.tu-dresden.de

Abstract – **Camera-based photoplethysmography (cbPPG) is a recent development in biomedical engineering, which is primarily used to the contactless assessment of subject's heart rate. cbPPG's easy acquisition scheme comes at cost of an usually low signal-to-noise ratio. Artefacts such as subject's movements and light changes impose severe difficulties and have hindered a wider use of the technique up-to-date. This work aims at providing a more reliable heart rate estimation under real application conditions. For this purpose, a Kalman filter complemented by signal quality indexes is proposed. The method is able to reduce the root mean-squared error of estimated heart rates by approximately 37%, while slightly increasing the detection accuracy.**

Keywords — camera-based photoplethysgraphy, heart rate detection, Kalman filter, signal quality index

I.   INTRODUCTION

Camera-based photoplethysmography (cbPPG) is a contactless measuring technique that allows the acquisition of cardio-respiratory signals using cameras [1]. Similar to the clinical used photoplethysmography (PPG), blood volume changes in the microvascular bed of tissue modulate light which exogenously penetrates the skin [2]. These intensity changes are remotely detectable by cameras. Concomitantly, mechanical changes on the skin surface contribute to the measured cbPPG signal [3]. In contrast to a PPG sensor, that uses an adjusted light source, cbPPG operates with both natural and artificial ambient lighting [4]. Furthermore, cbPPG overcomes the drawback of punctual measurements and offers the possibility to capture spatio-temporal information related to hemodynamics, which renders the technique interesting for many applications.

Over the last years, cbPPG has been used to determine the heart rate [5], [6], respiratory rate [7], heart rate variability [8], changes in oxygen saturation [9], [6] and analyze peripheral blood perfusion [4], [10]. However, under real conditions and non-laboratory settings, studies focus mostly on heart rate (HR) detection. Subject movements and light changes lead to a generally low signal-to-noise ratio (SNR) of the extracted cbPPG signals. A common approach is the use of the frequency domain for HR detection. In particular, the Fast Fourier Transform (FFT), or the Short Time Fast Fourier Transform (STFT), is often applied by sliding a temporal window through the raw or preprocessed cPPG signal [8], [11], [12]. This technique has been successfully applied by different researchers. However, such approaches

independently determine heart rates for every segment, therefore disregarding contextual information.

In this contribution, a real-time procedure is presented for accurate HR detection from cbPPG signals. The detection is based on two input information: i) rough HR estimate (based on standard HR values) and ii) a signal quality index (SQI), which is based on frequency spectrum or image velocity. Such information is assimilated using a Kalman Filter algorithm, SQIs are applied in a univariate and simple multivariate form. Differing from former approaches, the proposed method incorporates previous knowledge concerning physiological and technical aspects, whilst avoiding any complex calculations.

II.   MATERIALS AND METHODS

A. Dataset and Technical Set-up

Camera data was obtained within a psychomotor vigilance study, conducted at the Institute of Biomedical Engineering at the TU Dresden. The selected subset featured 22 healthy subjects of both genders (9 female and 13 males) at ages of 24.6 ± 3.8 years. All participants had light skin colors, four had glasses and six males were bearded. This contribution considers an excerpt of the recordings, namely a 4 min vigilance task, during which subjects were lying on a tilt table and occasionally moved slightly in response to the task. All participants gave their written informed consent. The study was approved by the Ethics Committee of the TU Dresden (EK:281092012).

For cbPPG measurement, we used an industrial RGB camera (IDS UI-5240CP-C-HQ) at a resolution of 320x240 pixels, a color depth of 3x10 Bit and a frame rate of 100 fps. The camera, set up with a lense of Schneider-Kreuznach (Cinegon 1.4/8), was positioned above the lying subject, so that recorded videos obtained subjects' frontal face (see Fig. 1). All measurements were executed indoors, where a fluorescent light source was used for illumination. Reference measurements were performed synchronously to cbPPG using an ADInstruments' Powerlab 16/35 (ADInstruments, Dunedin, NZ). Besides respiration and PPG signals, an ECG (modified Einthoven III lead - ADInstruments' Bio Amp FE 132) was recorded.

B. Image Processing

For each subject, 47 subsequent video sequences of 10 s (50 % overlap) were considered. At the beginning of every sequence, the face was detected by an algorithm based on the work of Viola and Jones [13]. The algorithm returns a box
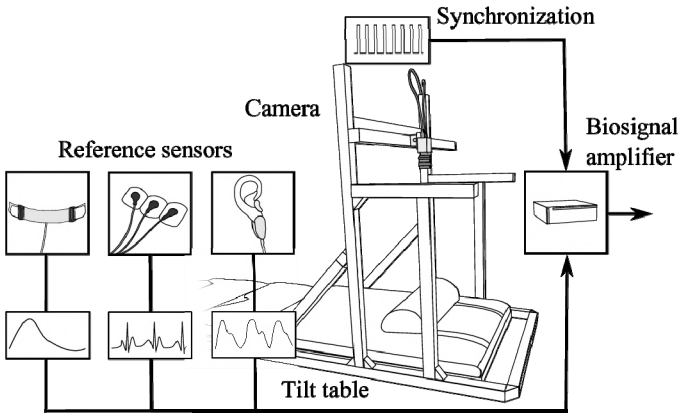
Fig. 1. Measurement setup for the study, consisting of tilt table, RGB camera, reference sensors (left to right: respiration, ECG, PPG) and a biosignal ADC. Modified from [5].
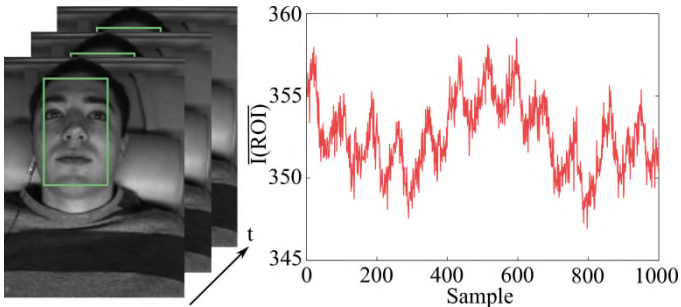

Fig. 2. Example for a cbPPG signal extraction. Left: estimated ROI (green) for a number of frames. Right: out of ROI extracted cbPPG signal ($\overline{I(ROI)}$: mean ROI intensity) of videos green channel (the dominant wave represents the pulse rate).

the subject's face, also containing exterior face sides that are not parallel to the image plane. To select more suitable areas (i.e. cheeks, forehead) as region of interest (ROI), 20 % of boxes' left and right borders were discarded as in [8] (see Fig. 2). To avoid any unwanted signal components in later cbPPG analysis, the ROI was hold statically for the whole sequence. Since it has been previously shown that video's green channel provides the highest cbPPG signal amplitudes [4], the other channels were disregarded. The signal segments were then extracted by averaging the ROI pixel values of every frame in each sequence. An example of such a cbPPG signal segment is shown in Fig. 2. Therefore, $4 \times 60/5 - 1 = 47$ cbPPG signal segments were obtained for each subject. These segments were further used in this work for estimating subjects' HR.

C. Heart Rate Detection

In order to obtain a rough HR estimate, FFT was applied to the extracted cbPPG 10 s signal segments (thus resulting in a STFT). Every windowed segment $x(n)$ was first detrended (best straight-line fit is removed) and multiplied by the Hanning window, to avoid frequency leakage. To obtain a higher frequency resolution Δf, zero-padding was applied. The requirement of $\Delta f \leq 1$ bpm led to 2^{13} signal points. Finally, the Fast Fourier Transform (FFT) provided the spectrum $X(k)$ with $\Delta f = 0.73$ bpm.

HR was defined as the maximal peak in $|X(k)|$, considering only the spectral band between $\Delta f \lambda_1 = 30$ bpm and $\Delta f \lambda_2 = 200$ bpm. Such limits were derived according to a standardized (ANSI/AAMI) permitted heart rate meter range [14].

D. Signal Quality

To assess the reliability of the calculated HR, two measures were applied, namely a frequency (SNR) and image velocity-based (v) measure. The first was suggested by De Haan and Jeanne [12]. Therefore, the power of the heart frequency $f_{hr1} \pm 5$ bpm and its first harmonic $f_{hr1} \pm 5$ bpm were considered as wanted signal. The noise component was determined by the residual power of the spectrum within $[\lambda_1, \lambda_2]$ (see Fig. 3). The measure reads to (in dB):

$$ SNR = 10\log_{10} \frac{\sum_{k=\lambda_1}^{\lambda_2} \Pi(k) |X(k)|^2}{\sum_{k=\lambda_1}^{\lambda_2} \left((1-\Pi(k)) |X(k)|^2 \right)^2} \quad (1) $$

where Π is defined as:

$$ \Pi(k) = \begin{cases} 1, & if \; \left| f_{hr} - \Delta f \cdot k \right| \leq 5 \; bpm \\ 1, & if \; \left| f_{hr1} - \Delta f \cdot k \right| \leq 5 \; bpm \\ 0, & otherwise \end{cases} \quad (2) $$

The latter measure, i.e. v is applied to detect motion which disturbs a reliable cbPPG measurement. An optical flow technique by Lucas and Kanade [15] was used to obtain a velocity vector field for every fifth video image. In each image, the dominant motion field with a certain direction was segmented and a mean field velocity was calculated. Thus, for every 10 s segment, 200 average velocity values were produced. Since only one movement (characterized by high velocity) is often sufficient to make the rough HR estimate untrustworthy, v was defined as the maximum velocity value for each segment.

E. Kalman Filtering

So far, both information required for further analysis were described, i.e. a rough HR calculations and reliability measures for those HR estimates (i.e. SNR and v). In order to improve the obtained HR estimates, an autoregressive (AR) model was applied within a Kalman Filter (KF) framework. This choice
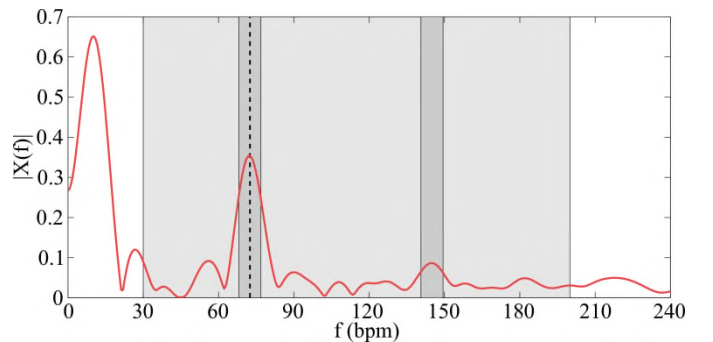

Fig. 3. Amplitude spectrum of a 10 s cbPPG signal segment. The light grey area represents the considered spectrum for heart rate detection and SNR calculation. The dashed line represents the detected heart rate, dark grey areas the spectral band (± 5 bpm) around the heart rate and its first harmonic (SNR calculation).

was motivated by KF's ability to track short-time stationary signals [16], [17] and real-time capabilities. Moreover, KF enables an elegant inclusion of SQI for more reliable HR estimation. This latter advantage was exploited by Li *et al* [18], who made use of KF and different SQIs for robust estimation of heart rates in multimodal data. AR models have been often referred to in biomedical engineering, e.g. in estimating subject's heart rate of cbPPG data [6] (similar to the aim of STFT in obtaining rough HR measures in this contribution), assessing the breathing rate from ECGs [17], [19] and analysis of multichannel electroencephalograms [20].

A univariate p^{th} order AR(p) process can be described as [16] [20]:

$$s(n+1) = \sum_{i=1}^{p} \left[\varphi_i(n) \cdot s(n-i+1) \right] + \xi(n) = \\ = \varphi^T(n) \cdot s(n) + \xi(n) \tag{3}$$

where the index $i \in [1, p]$, $s(n)$ denotes the pth order AR process, $\varphi_i(n)$ the time-dependent AR model parameter vector and $\xi(n)$ is an additive zero-mean white Gaussian noise process. The order p of an AR model implies how many previous time values are used to predict future values [17]. By applying the KF framework, we aim at obtaining a weighted moving average filter (with weights $\omega_1...\omega_p$) using the p previous rough HR values (Eq. 4).

$$\widehat{HR} = [HR(n-1) \cdot \omega_1 + HR(n-2) \cdot \omega_2 + \\ + ... + HR(n-p) \cdot \omega_p] \tag{4}$$

Eq. 4 can be modelled into KF's space-state representation as the following process and measurement equations:

$$x(n) = \mathbf{A} \cdot x(n-1) + \omega(n-1) \\ y(n) = \mathbf{H} \cdot x(n) + \nu(n) \tag{5}$$

being

$$x(n) = \begin{bmatrix} HR(n) \\ HR(n-1) \\ \vdots \\ HR(n-p+1) \end{bmatrix}$$

and

$$\mathbf{A} = \begin{bmatrix} \omega_1 & \omega_2 & \dots & \omega_{p-1} & \omega_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

using boldface capitalized notation for matrices and boldface lowercase for vectors, $n \in N$ are the discrete time samples. $\mathbf{x}(n)$ represents the $p \times 1$ state vector containing the p previously estimated HR values, $y(n)$ the observations (rough HR values), $\omega(n)$ is the process noise and $\nu(n)$ the observational noise. The noise covariance matrices are defined by $\omega \sim \mathcal{N}(0, \mathbf{Q}(n))$ and $\nu \sim \mathcal{N}(0, \mathbf{R}(n))$. The state transition matrix $\mathbf{A}$ is the $p \times p$ matrix, describing the AR model, while the observational matrix $\mathbf{H}$ is a $p \times p$ identity matrix. The weights $\omega_1...\omega_p$ were chosen to be linearly decreasing so that the most recent HR estimate has highest impact for the next estimate. The first p samples of the signal were mirrored for initializing the filter. KF weights are mostly defined by its modelled process and observational noise covariance matrices ($\mathbf{Q}$ and $\mathbf{R}$). In order to integrate the SQI information previously described, we are particularly interested in the $\mathbf{R}$ matrix, which is designed to be time-varying and dependent on SQI, as follows.

According to Equation 1, SNR is defined in absolute value (dB) and v in pixels/s, both with unrestricted range. However, normalized SQI measures are required. In order to normalize these SQIs, a regression was performed using a probability density function (PDF) onto the histogram of experimental SNR and v values. This PDF's cumulative distribution function (CDF) was used to map the calculated SNR (or v) values to the interval [0,1]. This mapping assures that small differences in the SQI resulted in considerable variations, as most of the experimental SNRs (or v) are centered around the slope of the CDF. The underlying distribution was assumed to be a generalized extreme value (GEV) distribution, which is described by its parameters location μ, scale σ and shape ξ. GEV's PDF is defined by:

$$GEV(\mu, \sigma, \xi) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)} \tag{6}$$

where

$$t(x) = \begin{cases} \left(1 + \left(\dfrac{x-\mu}{\sigma} \right) \xi \right)^{-\frac{1}{\xi}}, & if \ \xi \neq 0 \\ e^{-\frac{x-\mu}{\sigma}}, & if \ \xi = 0 \end{cases} \tag{7}$$

For clarity, the normalized SNR is henceforth noted as κSQI the SQI from v as $vSQI$. Fig. 4 demonstrates the original κSQI distribution, the experimentally determined parameters of the fitted GEV and its respective CDF. The same procedure was performed for $vSQI$, obtaining as regression parameters $\mu=0.024$, $\sigma=0.006$ and $\xi=0.039$. Differing from κSQI, the smaller the values for $vSQI$, the higher is the signal quality. Therefore, $vSQI$ was defined as one minus its normalized value.

Besides the proposed SQI measures, the novelty of this work is the use of these measures in conjunction with KF for weighting HR estimates from cbPPG recordings. This
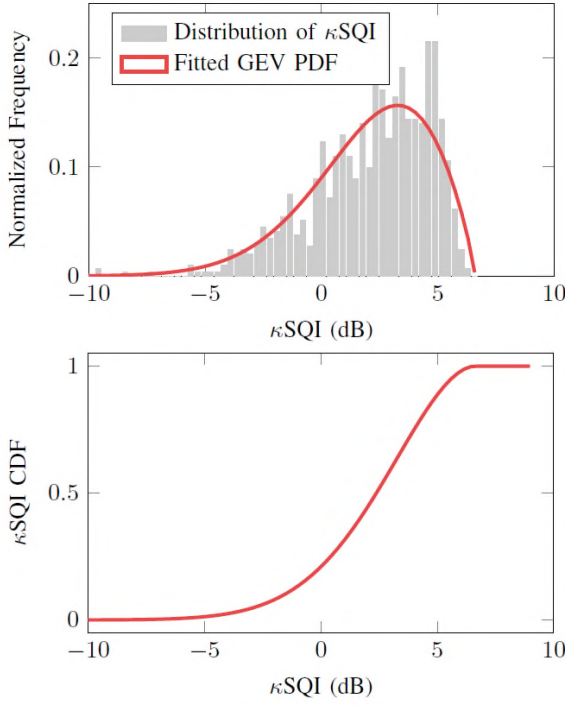
Fig. 4. Distribution of experimental SNRs with fitted GEV PDF (upper figure), with parameters μ=1.437, σ=2.844 and ξ=–0.551. The lower figure shows the resulting CDF.

procedure is done similarly to [18], [17]. KF's time-varying observational noise covariance matrix is modelled as:

$$\mathbf{R}(n) = \mathbf{R}_0 \cdot (1 - SQI(n)) \qquad (8)$$

It is important to notice that differently from [18], a linear relationship between **R** and SQI is used. This because during the mapping procedure, the small differences in SQI have already been taken into consideration.

III. RESULTS

The accuracy of our estimated HR was evaluated based on two measures, namely the root mean squared error (RMSE) and the HR detection rate (HDR). In order to generate the HR reference, QRS complexes were manually annotated in the ECG to assure that wrong or missing detections were excluded. The reference HR was then obtained by calculating the mean of the annotated RR intervals in each window of 10 seconds. RMSE was calculated from the difference between reference HR and obtained HR. The second measure, HDR, considers all detections as correct within a ±5 bpm interval from the reference. The interval is based on the allowed accuracy of heart rate meters in ANSI/AAMI EC13:2002 [14]. HDR results are given in percent by dividing the number of true positives by the total number of measured HR.

κ*SQI* and *vSQI* measures were applied individually, as well as combined by simply taking the minimum values between both. The order of the AR model was empirically evaluated and set to p=6, which corresponds to a 30 s segment. Aside from the model order, KF $\mathbf{Q}_0$ and $\mathbf{R}_0$ covariance matrices were calibrated. Grid search was used in assessing
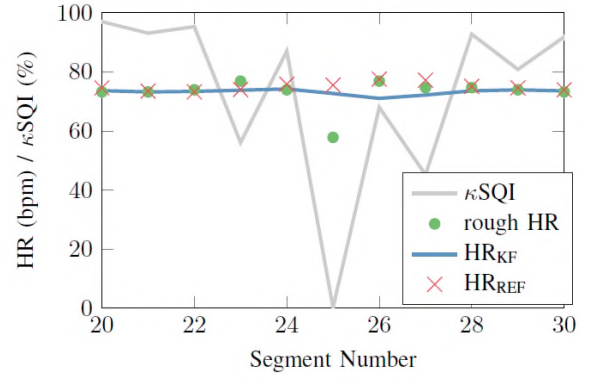

Fig. 5. Example of filtered data by the proposed KF methodology, subject number 22 using κ*SQI*.
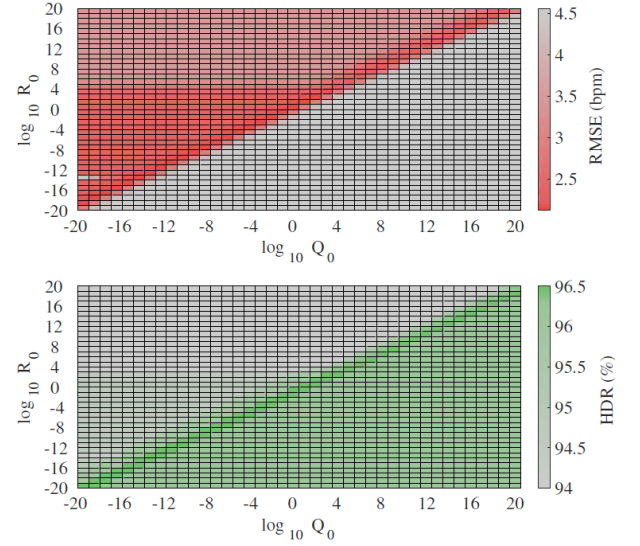

Fig. 6. Grid search for different $\mathbf{Q}_0$ and $\mathbf{R}_0$ values. Plots made use of all subjects and κ*SQI*. Similar results were obtained by *vSQI*.

how the HR estimation was affected by such parameters. Fig. 6 demonstrates these results in terms of RMSE and HDR, using κ*SQI*, similar results were obtained for *vSQI* and *min*(κ*SQI*,*vSQI*). For the further tests, $\mathbf{Q}_0 = \mathbf{R}_0 = 0.1$ were used. Fig. 5 shows an exemplary filtered signal using the proposed KF methodology. Table I summarizes the results for RMSE and HDR for each subject.

IV. DISCUSSION

From Fig. 5 it is remarkable that, despite the inaccurate initial HR estimate, KF is capable of weighting down estimates due to their low κ*SQI*. A 6th order model allows a plausible degree of dependency on former HR rough estimates, since it covers the previous 30 s segments. The proposed methodology produces a general improvement for HR estimation in terms of both HDR and RMSE (see Table I). This result indicates that our methodology is sound, despite some incorrect information observed in both SQI indexes.

TABLE I. Statistics for RMSE and HDR considering the approach with and without the proposed KF filtering. Different SQIs were applied in a univariate and simple multivariate manner

Subject	RMSE (bpm)	$RMSE_{KF}$			HDR (%)	HDR_{KF}		
		κSQI	$vSQI$	$min(\kappa SQI, vSQI)$		κSQI	$vSQI$	$min(\kappa SQI, vSQI)$
1	2.12	1.45	1.63	1.46	95.7	97.9	97.9	97.9
2	1.02	0.83	0.96	0.83	100	100	100	100
3	1.27	0.90	0.92	0.90	100	100	100	100
4	4.12	2.27	3.89	2.27	97.9	97.9	97.9	97.9
5	8.22	4.63	4.58	4.58	95.7	95.7	95.7	95.7
6	7.77	4.57	6.42	4.54	93.6	93.6	91.5	93.6
7	1.82	1.38	1.22	1.22	100	100	100	100
8	11.88	7.35	7.22	7.25	91.5	89.4	89.4	89.4
9	3.17	1.84	3.03	1.84	87.2	95.7	89.4	95.7
10	1.76	1.19	1.41	1.40	95.7	100	100	100
11	1.74	1.29	1.54	1.29	97.9	100	100	100
12	1.11	1.03	1.02	1.04	100	100	100	100
13	4.05	2.56	2.52	2.58	97.9	95.7	97.9	95.7
14	17.9	11.3	12.8	11.3	91.5	80.9	87.2	80.9
15	7.39	4.02	4.75	4.06	97.9	97.9	95.7	95.7
16	2.48	1.67	1.69	1.60	97.9	97.9	97.9	97.9
17	1.55	1.12	1.25	1.17	97.9	100	100	100
18	2.11	1.41	1.53	1.48	97.9	100	97.9	100
19	7.51	4.62	4.25	4.26	95.7	95.7	95.7	95.7
20	6.61	4.73	4.81	4.78	87.2	89.4	89.4	89.4
21	1.93	1.58	1.56	1.57	97.9	97.9	97.9	97.9
22	2.79	1.71	1.75	1.74	97.9	97.9	97.9	97.9
Average	4.56	2.89	3.22	2.87	96.1	96.5	96.3	96.4

This study counted with a limited database, meaning short segments, a limited number of subjects and restricted HR variability. Further studies should apply the proposed method to larger amount of data, separated into training and test sets, in order to confirm its real world performance.

Further validation is also required for the available SQI measures. This should be possible using a carefully annotated training set, containing e.g. markers for movement and/or lightning changes. Moreover, the result indicate, KF could benefit greatly from multivariate SQI approaches and further combinations should be evaluated. A deeper insight into the method in its current state can be gained by an additionally evaluating a virtual SQI incorporating the reference HR. Using a binary SQI based on the HDR measure that indicates true positives (with ones) and false negatives (zeros), results for $HDR > 99\%$ and $RMSE < 1.6$ were obtained. Such findings indicate that the proposed KF is conceptually well suited for the purpose, however the described SQIs have to be further improved.

Lastly, the STFT step could be eliminated and substituted by an AR model, similar to [6], and incorporated into KF's model. A deeper analysis comprising a larger dataset would enable a fair comparison between STFT and AR approaches for the cbPPG application.

V. Conclusion

This work proposes the use of KF for filtering and improving cbPPG HR estimates. The method is capable of improving the HR estimation based on different signal quality measures. KF is an interesting solution to the problem, since it enables various extensions of this work within a well-designed framework.

References

[1] M. Huelsbusch and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI," in Medical Imaging 2002, A. V. Clough and C.-T. Chen, Eds. San Diego, USA: International Society for Optics and Photonics, Apr. 2002, pp. 110–117.

[2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement." Physiol. Meas., vol. 28, no. 3, pp. R1–39, Mar. 2007.

[3] F. P. Wieringa, F. Mastik, and A. F. W. van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology." Ann. Biomed. Eng., vol. 33, no. 8, pp. 1034–41, Aug. 2005.

[4] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Opt. Express, vol. 16, no. 26, p. 21434, Dec. 2008.

[5] S. Zaunseder, A. Heinke, A. Trumpp, and H. Malberg, in 2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO), Kiev, Ukraine, Apr., pp. 286–290.

[6] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models." Physiol. Meas., vol. 35, no. 5, pp. 807–31, May 2014.

[7] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image." Med. Eng. Phys., vol. 29, no. 8, pp. 853–7, Oct. 2007.

[8] M. Poh, D. McDuff, and R. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express, vol. 18, no. 10, pp. 795–805, 2010.

[9] L. Kong, Y. Zhao, L. Dong, Y. Jian, X. Jin, B. Li, Y. Feng, M. Liu, X. Liu, and H. Wu, "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light." Opt. Express, vol. 21, no. 15, pp. 17 464–71, Jul. 2013.

[10] N. Zaproudina, V. Teplov, E. Nippolainen, J. A. Lipponen, A.A. Kamshilin, M. Na̎rhi, P. A. Karjalainen, and R. Giniatullin, "Asynchronicity of facial blood perfusion in migraine." PloS one, vol. 8, no. 12, p. e80189, Jan. 2013.

[11] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone." in EMBC, vol. 2012. IEEE EMBS, Jan. 2012, pp. 2174–7.

[12] G. de Haan and V. Jeanne, "Robust pulse-rate from chrominance-based rPPG." IEEE Trans. Biomed. Eng., no. c, pp. 1–9, Jun. 2013.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cas- cade of simple features," in Computer Vision and Pattern Recognition (CVPR), vol. 1. IEEE Comput. Soc, 2001, pp. I–511–I–518.

[14] Assoc. for the Advancement of Medical Instrumentation, "Cardiac monitors, heart rate meters, and alarms," pp. 13, 35, 2002.

[15] B. D. Lucas, T. Kanade et al., "An iterative image registration technique with an application to stereo vision." in IJCAI, vol. 81, 1981, pp. 674– 679.

[16] G. Doblinger, "An adaptive kalman filter for the enhancement of noisy ar signals," in IEEE International Symposium on Circuits and Systems (ISCAS), vol. 5, 1998, pp. 305–308.

[17] A. Johnson, S. Cholleti, T. Buchman, and G. Clifford, "Improved respiration rate estimation using a Kalman filter and wavelet cross-coherence," in Annals of Computing in Cardiology Conference (CinC), 2013, 2013, pp. 791–794.

[18] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter." Physiol. Meas., vol. 29, no. 1, pp. 15–32, Jan. 2008.

[19] L. Tarassenko, L. Mason, and N. Townsend, "Multi-sensor fusion for robust computation of breathing rate," Electron. Lett., vol. 38, no. 22, p. 1314, 2002.

[20] M. Arnold, W. H. Miltner, H. Witte, R. Bauer, and C. Braun, "Adap- tive AR modeling of nonstationary time series by means of Kalman filtering." IEEE Trans. Biomed. Eng., vol. 45, no. 5, pp. 553-62, May 1998.