DE GRUYTER Current Directions in Biomedical Engineering 2017; 3(2): 483–487

Christian S. Pilz*, Sebastian Zaunseder, Ulrich Canzler and Jarek Krajewski

Heart rate from face videos under realistic conditions for advanced driver monitoring

Abstract: The role of physiological signals has a large impact on driver monitoring systems, since it tells something about the human state. This work addresses the recursive probabilistic inference problem in time-varying linear dynamic systems to incorporate invariance into the task of heart rate estimation from face videos under realistic conditions. The invariance encapsulates motion as well as varying illumination conditions in order to accurately estimate vitality parameters from human faces using conventional camera technology. The solution is based on the canonical state space representation of an Itô process and a Wiener velocity model. Empirical results yield to excellent real-time and estimation performance of heart rates in presence of disturbing factors, like rigid head motion, talking, facial expressions and natural illumination conditions making the process of human state estimation from face videos applicable in a much broader sense, pushing the technology towards advanced driver monitoring systems.

Keywords: Photoplethysmography Imaging, Diffusion Process, SDE, Invariance, Psychophysiology, Human State Computing.

https://doi.org/10.1515/cdbme-2017-0102

1 Introduction

During the last years, the task of measuring skin blood perfusion and heart rate measurements from facial images became inherent part of several top conferences. Interestingly, most contributions focus on how to cope with motion like head pose variations and facial expressions since any kind of motion on a specific skin region of interest will destroy the underlying blood perfusion signal in a way that no reliable information can be extracted anymore. **Figure 1** illustrates the disturbing influence of head motions on the raw pulse signal.

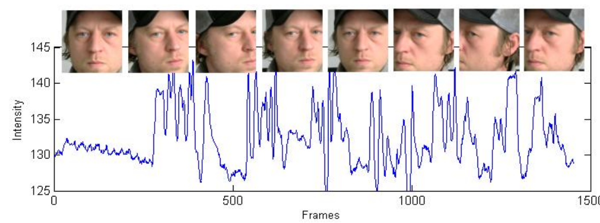

Figure 1: A Typical scenario where heart rate estimation becomes challenging, rigid head motions. In the first 250 frames the user is in a resting state and the fine pulsation of blood flow is visible on the averaged green channel of skin pixels. After 300 frames the user starts to move his head and the pulse signal gets lost.

1.1 Related work

The term Photoplethysmography, short PPG, dates back to the late first half of the 20th century, when Molitor and Kniazak [1] recorded peripheral circulatory changes in animals. One year later, Hertzmann [2] introduced the term Photoelectric Plethysmograph as he observed "the amplitude of volume pulse as a measure of the blood supply of the skin". With the ongoing fast development of semiconductor technology, the last three decades has seen large progress in the PPG instrumentation. PPG sensors have been explored extensively, including the ring finger, wrist, brachia, earlobe, and external ear cartilage. Advancement to the classical PPG is the camera based Photoplethysmography Imaging (PPGI) method introduced by the pioneering work of Blazek [3]. Since his first published visualisation of pulsatile skin perfusion patterns in the time and frequency domain, classical signal processing methods are applied commonly to extract reasonable information out of the perfusion signals [4][5][6]. Hülsbusch [4] realized that motion of the skin area of interest inherently induces artifacts into the extracted signal. Therefore, canceling motion artifacts during signal processing became an important aspect for reliable skin

*** Corresponding author: Christian S. Pilz:** CanControls GmbH, Markt 45, 52064 Aachen, Germany, e-mail: pilz@cancontrols.com
Sebastian Zaunseder: TU Dresden, Institut für Biomedizinische Technik, Fetscherstr. 29, 01307 Dresden, Germany
Ulrich Canzler: CanControls GmbH, Markt 45, 52064 Aachen, Germany, e-mail: canzler@cancontrols.com
Jarek Krajewski: Bergische Universität Wuppertal, Institut für Sicherheitstechnik HF - Arbeitsbereich Human Factors & Diagnostik, Gaußstraße 20 42119 Wuppertal, Germany krajewsk@uni-wuppertal.de

blood perfusion measurements. From the basic early idea of compensating the motion of the skin area of interest by optical flow methods directly in the image plane [4], Poh et al. [6] regarded the problem solution for facial videos as a blind source separation task using Independent Component Analysis (ICA) over the different color channels. De Haan and Jeanne [10] proposed to map the PPGI-signals by linear combination of RGB data to a direction that is orthogonal to motion induced artifacts. A recent alternative, which does not require skin-tone or pulse-related priors in contrast to the channel mapping algorithms, determines the spatial subspace of skin-pixels and measures its temporal rotation for signal extraction [11]. We go beyond the state of the art and propose a holistic classical interpretation of the blood perfusion phenomena.

2 Methodology

The underlying system of measuring heart rates from face regions using conventional camera technology is modelled upon a diffusion process. The entire process itself is divided into independent single processes; the heart frequency, the

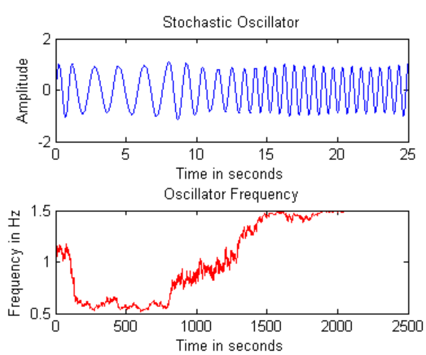

Figure 2: A simulated trajectory of a stochastic oscillator with frequency trace in a range typical for a human in resting state

illumination and the users head movement and facial motion. The periodic event of heart frequency is expressed in form of a stochastic resonator

$$\frac{d^2 c_n(t)}{dt^2} = -\left(2\pi f(t)\right)^2 c_n(t) + e_n(t) \qquad \textbf{(1)}$$

with

$$c_n(t) = a_n \cos(2\pi nft) + b_n \sin(2\pi nft) \qquad \textbf{(2)}$$

representing the solution of a second order differential equation with respect to the classical mechanics of circular motion [13]. The white noise component $e_n(t)$ reflects small changes in amplitude and phase. The major advantage of such a stochastic representation of a resonator is, even when the frequency has discontinuous the signal is always continuous. **Figure 2** shows a single stochastic oscillator with time-varying frequency and amplitude. The illumination as well as the head movement and facial motion are expressed as a Wiener process

$$\frac{d^2 x(t)}{dt^2} = w(t) \qquad \textbf{(3)}$$

whereby a violation of the smoothness criterion yields to a generalized Poisson (e.q. Cox) process

$$x(t) = x_0 + \sum\nolimits_{i:s_i \le t} \varepsilon_i \qquad \textbf{(4)}$$

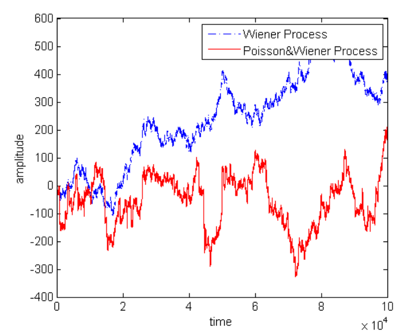

Figure 1: A simulated trajectory of a Wiener process and its realization modulated by a Poisson process with jump frequency 1.95 and magnitude 35.

describing the time varying jump frequency and magnitude of pixel intensities. **Figure 3** shows a simulated trajectory of a Wiener process and its realization modulated by a Poisson process.
The general solution of the corresponding stochastic differential equations

$$\frac{d\vec{x}(t)}{dt} = F\vec{x}(t) + L\vec{w}(t) \qquad \textbf{(5)}$$

is given by Itô's lemma [12].

The discrete-time approximation yields to [14]

$$\vec{x}(t_i + 1) = \exp(\Delta t F)\vec{x}(t_i) + \int_{t_i}^{t_i+1} \exp((t - s)F)L\vec{w}(s)\ ds \tag{6}$$

with $\Delta t = t_{i+1} - t_i$, the Wiener process $w(t)$ with spectral density W and the covariance of the stochastic integral

$$Q_i = \int_0^{\Delta t} A_i L W L^T A_i^T d\tau \tag{7}$$

with $A_i = \exp(\Delta t F)$, which results to the discrete-time model

$$\vec{x}(t_{i+1}) = A_i\vec{x}(t_i) + \vec{q}(t_i) \tag{8}$$

$$\vec{y}(t_i) = H_i\vec{x}(t_i) + \vec{e}(t_i) \tag{9}$$

with process noise

$$\vec{q}(t_i) = N(0, Q_i) \tag{10}$$

and measurement noise

$$\vec{e}(t_i) = N(0, \Sigma_i) \tag{11}$$

If the resonator's fundamental frequency is known, the solution yields to a general time-discrete linear dynamic system [9]. However, since the resonator's fundamental frequency is unknown, the problem is given as latent state of the frequency. This results in a Markov process, whereby the latent states are time-discrete linear dynamic systems. The closed form solution to this problem is described by Bloom and Bar-Shalom [8]. The advantage of this kind of formulation is that the case of non-uniform sampling as well as missing observations is naturally included in the model. The basic idea of methodology is inspired by the work of Särrkä [7].

3 Experiments

To evaluate the proposed model, empirical data is collected under natural environmental conditions with a typical low-cost opto-electronical sensor device, a Logitech HD C270 webcam, as well as reference ground truth measurements using a common finger pulseoximeter, a CMS50E PPG device. 25 users were asked to perform video recordings in two sessions resulting in a total amount of 50 videos. The first session is selected to be even-tempered without any kind of larger head or body movements and facial expressions. During the second session, participants were free to move their head naturally while remaining seated. Typical movements included tilting the head sideways, nodding the head, looking up/down and leaning forward/backward. Some participants also made facial expressions, or started to talk. There reflecting typcial driver behaviour. The recording illumination environment was chosen as daylight scenario without any additional lighting. The duration of each session is approximately one minute. The frame rate was fixed to 15 fps in average and the corresponding time stamps for each frame were captured too. The finger pulseoximeter data for each session and participant was stored for later comparison. For every video recording a standard face finder was used to determine the analysis region of interest. The extracted averaged gray intensity feature was feed into the vector valued representation of the diffusion process on a frame by frame basis. On every estimated pulse trace a spectral peak is determined by the Lomb periodogram. The frame duration was set to 10 seconds with 90 percent overlap. The correlation and Bland-Altman plots for the resting and head motion condition are reported in the following **figure 5** and **figure 6** respectively. To obtain further insides about the potential strength of the diffusion process model, the approach is compared against the recently published Spatial Subspace Rotation (SSR) [11] and the baseline ICA approach

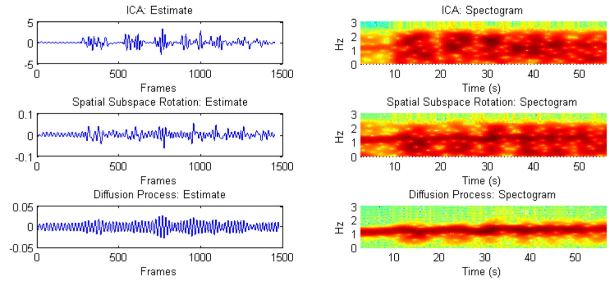

Figure 2: Comparison of an estimated pulse signal under rigid head motions and the corresponding spectrogram for the ICA [6], the SSR [11] and the diffusion process method. These estimates are based upon the video illustrated in **figure 1**.

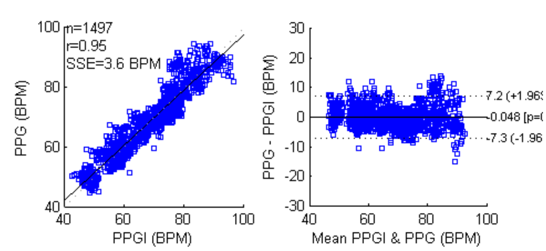

Figure 5: Correlation and Bland-Altman plots of PPGI diffusion process estimated heart rate against CMS50E PPG reference of 25 users in resting state.

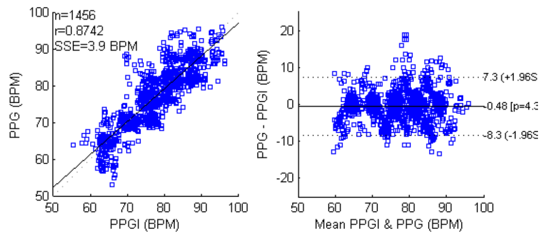

Figure 6: Correlation and Bland-Altman plots of PPGI diffusion process estimated heart rate against CMS50E PPG reference of 25 users performing head rotations.

[6]. **Figure 4** compares an users estimated pulse signal under rigid head motions and the corresponding spectrogram for the three methods. The heart rate for the ICA methods nearly gets lost completely. For the SSR method the frequency trace is better visible but cannot compete against the diffusion process model where the heart rate is very clear over the entire sequence of head movements. The detailed correlation coefficients and squared errors of prediction for all approaches are provided for the two data sessions in Table 1. ICA performs worst and is not able to provide reliable heart rate information during head motion. Although SSR performs better it cannot compete against the robustness of the diffusion process.

Table 1: Pearson's correlation coefficient and squared errors of prediction of ICA, the SSR and the diffusion process (DP) method under different scenarios.

Type	Resting	Head Rotation
ICA	0.61/7.8	0.21/14.6
SSR	0.78/4.8	0.47/7.6
DP	0.95/3.6	0.87/3.9

4 Conclusion

In this work, we have presented a holistic signal interpretation of heart rate estimation from face videos under realistic simulated driving conditions. The closed form solution of the corresponding stochastic differential equations yields to a diffusion process where the exact estimate of the source separated heart rate signal is obtained via the posterior distribution of the process. We compared the model against two common approaches on face videos under resting as well as head and facial motion scenarios under natural illumination conditions. Measurements on a 25 user experiment showed clearly superior robustness of the diffusion process modelling, although the uncertainty of prediction still gets slightly increased during natural head motion. We conclude that an entirely invariant process model still depends on a more robust feature representation.

Author's Statement

Research funding: The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) under grant agreement 01|S15024 (VIVID - IKT2020/2015-2018). The opinions expressed here are those of the authors and may or may not reflect those of the sponsoring parties. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] H. Molitor and M. Knaizuk, A new bloodless method for continuous recording of peripheral change. Jour. Phar. Expr. Ther., 27: 5-16.1936.
[2] A.B. Hertzman. Photoelectric Plethysmography of the Fingers and Toes in Man. Exp. Biol. Med.,37:,529-534.1937.
[3] V. Blazek. Optoelektronische Erfassung und rechnerunterstützte Analyse der Mikrozirkulations-Rhythmik. Biomed. Techn. 30 (1):121-122.1985.
[4] M. Hülsbusch. A functional imaging technique for opto-electronic assessment of skin perfusion. PhD thesis, RWTH Aachen University.2008
[5] W. Verkruysse, L.O. Svaasand and J.S. Nelson. Remote plethysmographic imaging using ambient light. Optics Express, 16 (16):21434-21445. 2008.
[6] M.Z. Poh, J.D. McDuff and R.W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express, 18 (10): 10762-10774. 2010.
[7] S. Särkkä. Recursive Bayesian Inference on Stochastic Differential Equations. PhD thesis, Helsinki University of Technology. 2006
[8] H.A.P Bloom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with Markovian switching coefficients. IEEE Transactions on Automatic Control, 33 (8):780-783.1988.
[9] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. Transactions of the ASME-Journal of Basic Engineering, 83:95-108.1961.
[10] G. de Haan and V. Jeanne. Robust pulse-rate from chrominance-based rPPG. IEEE Transactions on Biomedical Engineering, 60 (10): 2878-2886.2014.

[11] W. Wang, S. Stuijk and G. de Haan. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. IEEE Transactions on Biomedical Engineering, 63 (9):1974-1984.2016
[12] K. Itô. On Stochastic Differential Equations. Memoris Of The American Mathematical Society, 4.1951.
[13] R. Feynman, R. Leighton and M. Sands. The Feynman Lectures on Physics Vol. 1. Chapter 21. Addison-Wesley. 1963.
[14] B. Øksendal. Stochastic Differential Equations. Springer, 2003.