Local Group Invariance for Heart Rate Estimation from Face Videos in the Wild

Christian S. Pilz
CanControls, Aachen, Germany

Sebastian Zaunseder
TU Dresden, Germany

Jarek Krajewski
University of Wuppertal, Germany

Vladimir Blazek
RWTH Aachen University, Germany
CTU Prague, Czech Republic

Abstract

We study the impact of prior knowledge about invariance for the task of heart rate estimation from face videos in the wild (e.g. in presence of disturbing factors like rigid head motion, talking, facial expressions and natural illumination conditions under different scenarios). We introduce features invariant with respect to the action of a differentiable local group of local transformations. As result, the energy of the blood volume signal is re-arranged in vector space with a more concentrated distribution. The uncertainty in the feature distribution is incorporated with a model that leverages the local invariance of the heart rate. During experiments the method achieved strong estimation performance of heart rate from face videos in the wild. To demonstrate the potential of the approach it is compared against recent algorithms on data collected to study the impact of the mentioned nuisance attributes. To facilitate future comparisons, we made the code and data for reproducing the results publicly available.

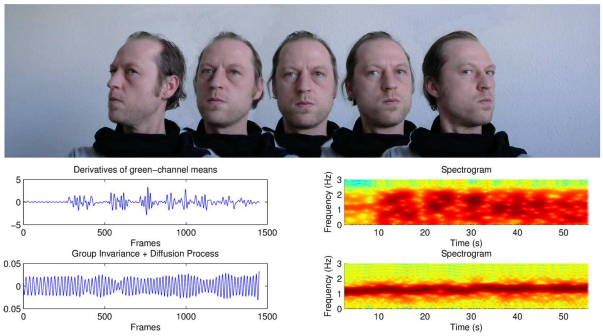

Figure 1. Rigid and non-rigid facial motions act as nuisance factors on the tiny blood volume changes inherently destroying the target information of heart rate under the conventional formulation of the problem. Utilizing classical group theory we are able to define features invariant with respect to the action of the group of nuisance transformations making it possible to estimate heart rate information under everyday facial motions. As illustrated for the above head motions, the green channel information doesn't yield to reasonable heart rate information in the frequency domain. However, the proposed invariant feature shows a clear signal in both time and frequency domain.

1. Introduction

In general, it can't be expected to obtain clear signals from sensors. Signals are often affected by nuisance factors hiding the target function. This is a major drawback and makes the analysis of processes difficult, enforcing several constraints for real applications. One of these applications is the task of vital parameter estimation from face videos under arbitrary natural conditions. The signal is heavily influenced by endogenous as well as exogenous factors, like face and head motion, illumination changes as well as specific sensor and image properties [21] and, last but not least, human conditions itself like age and health [37].

The role of physiological states has a large impact on human state computing in computer vision, since it holds informations about the affective nature of the human interacting with the machine. During the last years, measuring blood volume changes and heart rate measurements from facial images became a part of top computer vision conferences [18, 19, 26, 38]. All these 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 (ROI) will destroy the raw signal in a way that no reliable information can be extracted anymore. Besides being able to estimate vitality parameters like heart rate and respiration, the functional survey of wounds as well as quantification of allergic skin reaction [3] are further applications of camera-based blood perfu-

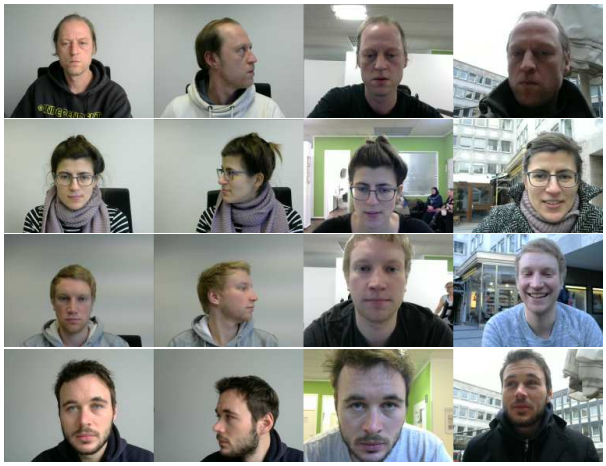

Figure 2. Example images of the 25 user data collection. From left to right: Face recordings during head resting conditions, during head rotations, during an exercise on a bicycle ergometer in a gym and during an urban conversation.

sion analysis. Recently, prediction of emotional states, stress [22, 30, 5], fatigue [35] and sickness [14] became interesting new achievements in this area, pushing the focus of this technology further towards human-machine interaction.
In contrast to the genuine medical use-case of the technology, in computer vision and human-machine interaction we can't expect any cooperative behavior of the user without introducing lack of convenience and a reduction of the general acceptance. Thus, there's a demand to produce features and models robust to nuisance factors, still preserving the desired target information. In a more philosophical sense, where everything seems to change just search for what is not changing.

The main contributions of this work are

- a feature representation for the problem of estimating heart rate using low-cost camera sensor technology invariant under rigid and non-rigid facial motions and varying illumination,
- a direct connection of the proposed feature presentation to a functional mathematical formulation for the quasi-periodic and non-stationary nature of heart rate and
- a set of uncompressed face video recordings with reference measurements collected under resting, head motion, a bicycle ergometer scenario and during an urban conversation.

The outline of this work is as follows. From the historical genuine up to the development of the state of the art in computer vision, the methodology of heart rate estimation from face videos will be reviewed. Followed by theoretical aspects, the feature and model space will be described. Based upon an evaluation on collected data the results will be presented and finally discussed.

2. Related Work

Photoplethysmography, short PPG, dates back to the late first half of the 20th century, when Molitor and Kniazak [25] recorded peripheral circulatory changes in animals. A year later, Hertzman [15] introduced the term Photoelectric Plethysmograph as "the amplitude of volume pulse as a measure of the blood supply of the skin". Hertzman's instrumentation comprised mainly of a tungsten arc lamp and a photomultiplier tube. An advancement to the classical PPG is the camera based PPGI (with I for Imaging) introduced by the work of Blazek [4]. The basic principle behind the measurement of blood volume changes in the skin by means of PPG is the fact that hemoglobin absorbs light much stronger at specific frequency bands than the remaining skin tissues. The first published visualization of pulsatile skin perfusion patterns in the time and frequency domain is given by Blazek [4]. Since classical signal processing is applied mostly to extract information out of the perfusion signals [17, 28, 40]. However later it is realized that motion of the skin ROI [17] and micro motion of the head due to cardiac activity [2, 23] inherently induces artifacts into the extracted signal, especially when lighting is neither uniform nor orthogonal, canceling motion artifacts during signal processing became an important aspect for reliable skin blood perfusion measurements [24]. A basic early idea of compensating the motion of the skin ROI by optical flow methods directly in the image plane [17] is followed by Poh *et al*. [28], who proposed to solve the problem by blind source separation using Independent Component Analysis (ICA) over the different color channels. However, Wedekind *et al*. [43] compared ICA in multiple setting and principal component analysis and showed limitations of either transform. In fact, in case the underlying signal basis is majoritarian Gaussian, ICA will not be able to determine a proper de-mixing matrix and the independent components cannot be obtained in a deterministic order [8]. A solution to this problem excluding scenarios that have periodic movements is discussed by Macwan *et al*. [20]. Tarassenko *et al*. [36] attempted to cope with light flicker by using an auto-regressive modeling and pole cancellation. Haan and Jeanne [9] and De Haan and Van Leest [10] proposed to map the PPGI-signals by linear combination of RGB data to a direction that is orthogonal to motion induced artifacts. An 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 measure its temporal rotation for signal extraction [42]. Tulyakov *et al*. [38] proposed matrix completion to jointly estimate reli-

able regions and heart rate estimates whereby Li *et al*. [19] applied an adaptive least square approach to extract robust pulse frequencies. Both reported performance gains similar to De Haan and Jeanne [9], however they used compressed video data during their experiments, which raises some doubts on the validity of results. Wang *et al*. [41] reported an orthogonal behavior of skin color and motion artifacts but introduced a static operator for feature transformation representing results on private data. A new stochastic model formulation was introduced by Pilz *et al*. [27] outperforming Wang *et al*. [42], however the proposed features based upon vector quantization seem to be a rather time consuming heuristic approach. All these important contributions share the problem that there exists no consensus about an unique benchmarking criterion as well as an agreement on suitable open data sets for fair comparison of algorithms.

3. Methodology

In pattern recognition the invariance problem is anchored as traditional paradigm where the classification aims to be invariant with respect to action of a group that acts on set of features [32]. More generally, an invariant with respect to an equivalence relation is a property that is constant on each equivalence class [44]. Therefore, the invariance criterion is a necessary condition for the generalization ability of learning algorithms. Invariance can be regarded in the feature and model space. For heart rate from face videos the basic feature is usually computed over a set of pixel intensities out of aligned face regions and the model space over a set of suitable frequencies. In the following we describe features invariant with respect to the action of the Lie group [12] of rigid transformations, the Special Euclidean group $SE(3)$ and a stochastic frequency representation invariant with respect to the quasi-periodic nature and non-stationarity of heart rate. The model space is based upon the previous works of Särrkä [31] and Pilz *et al*. [27].

The Feature Space

Regarding a common optical sensor signal

$$\vec{p} \in \mathbb{R}^n = \{R, G, B\}, n = 3 \tag{1}$$

as spatial expectation over a skin operator s and function of time t

$$\vec{x}(t) = \int_0^\infty \mathbb{E}[\{\vec{p} \mid s(\vec{p})\}]\mathrm{d}t \tag{2}$$

we assume this multivariate observation is drawn by a normal distribution

$$\vec{x}(t) \sim \mathcal{N}(\vec{\mu}, \vec{\sigma}^2). \tag{3}$$

Local invariance of blood volume changes as function of time for each input feature $\vec{x}(t)$ under transformations of a differentiable local group of local transformations $\mathcal{L}_T$ [33]

$$\frac{\partial}{\partial T}\Big|_{T=0} = f(\mathcal{L}_T, \vec{x}(t)) = 0 \tag{4}$$

can be approximately enforced by minimizing the regularizer

$$\frac{1}{l}\sum_{i=1}^{l}(\frac{\partial}{\partial T}\Big|_{T=0} f(\mathcal{L}_T, \vec{x}_i))^2. \tag{5}$$

For the covariance matrix of the observation

$$\{\vec{x}_i : i = 1, ..., l\} \tag{6}$$

with respect to the transformations $\mathcal{L}_T$

$$C := \frac{1}{l}\sum_{i=1}^{l}(\frac{\partial}{\partial T}\Big|_{T=0}\mathcal{L}_T, x_i)(\frac{\partial}{\partial T}\Big|_{T=0}\mathcal{L}_T, x_i)^\top \tag{7}$$

and the corresponding symmetric eigenvalue problem

$$CV = V\Lambda \tag{8}$$

we find an operator P with corank $k = 1$ [34] for

$$\lim_{l\to\infty} P = I - VV^\top \tag{9}$$

and the corresponding feature vector

$$\vec{\tilde{x}} = P \cdot \vec{x}. \tag{10}$$

The observation $\{\vec{\tilde{x}}_i : i = 1, ..., l\}$ is located on the null space defined by the projection operator P

$$\mathcal{H} = \mathcal{N}(P) \tag{11}$$

The hyperplane $\mathcal{H}$ is a linear subspace of $\mathbb{R}^n$.
Fig. 3 illustrates the feature transformation step.

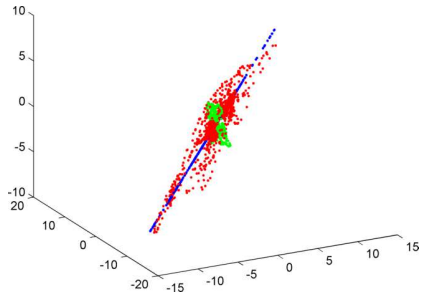

Figure 3. The raw features $\vec{x}$ (red dots, determined over the head motion image sequence illustrated in Fig.1 and centered for visualization purpose) are first transformed by projecting these onto the largest eigenvector of the covariance matrix C (blue dots) and then rescaled (green dots). In other words, the directions of main variance of the random vector emphasis is put on features which are less variant under L_T. The projection matrix P carries out an orthogonal projection to the blood volume change complementary space, which is assumed to accommodate the major load of nuisance factors not related to the target function.

For each resulting input feature $\vec{\tilde{x}}$ under transformations of a differentiable local group of local transformations $\mathcal{L}_R$ minimizing the corresponding regularizer, according to the principles in (4) and (5), yields to [29]

$$\lim_{l\to\infty} C_{\vec{\tilde{x}}_i} - A \cdot \lim_{l\to 0} C_{\vec{\tilde{x}}_i} \cdot A^\top = 0 \tag{12}$$

with

$$A = \lim_{l\to\infty} C_{\vec{\tilde{x}}_i}^{\frac{1}{2}} \cdot \lim_{l\to 0} C_{\vec{\tilde{x}}_i}^{-\frac{1}{2}} \tag{13}$$

and

$$b = \mathbb{E}[\{\vec{\tilde{x}}_i\}] - \mathbb{E}[\{A \cdot \vec{\tilde{x}}_i\}] \tag{14}$$

resulting to feature representation obtained by the linear transform

$$\vec{x}_{SE(3)} = A \cdot \vec{\tilde{x}} + b \tag{15}$$

Other common convenient solutions for the estimation of rigid transforms can be found following the works of Arun *et al*. [1] and Umeyama [39].

The choice of the interval $\lim_{l\to 0}$ plays a crucial role. We did experiments by computing the argument of the derivatives of the largest eigenvalue over the integral

$$\arg(t) = \int_0^{2\pi} \frac{d\lambda_1(t)}{dt} \mathrm{d}t \tag{16}$$

with respect to the symmetric eigenvalue problem

$$\lim_{l\to\pi ft} C_{\vec{\tilde{x}}_i} V = V\Lambda. \tag{17}$$

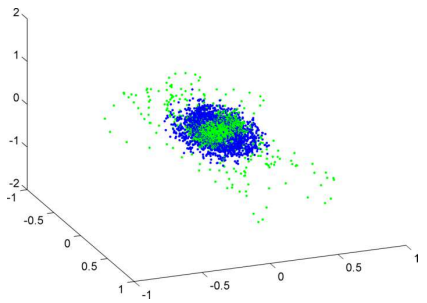

Figure 4. The rescaled features $\vec{\tilde{x}}$ (green dots) are then first rotated to the direction of the largest variance of the new covariance matrix $C_{\vec{\tilde{x}}}$, scaled and then translated to center of mass $\mathbb{E}[\{\vec{\tilde{x}}_i\}]$ (blue dots). These steps essentially represent linear transformations of the coordinate system in order to maintain a homogeneous view onto the feature space of blood volume changes.

The observation $\{\vec{x}_{SE(3)_i} : i = 1, ..., l\}$ is be distributed on the topological manifold

$$S^1 = \{\vec{x}_{SE(3)} \in \mathbb{C} \mid \mathbb{C} \in \mathcal{H} : \left\|\vec{x}_{SE(3)}\right\| = 1\}, \tag{18}$$

for any real number $0 \leq c < 1$ with

$$\vec{x}_{SE(3)_c} \in \mathbb{C} \tag{19}$$

$$\vec{x}_{SE(3)_c} = e^{2\pi jc} \tag{20}$$

$$\nu_{\vec{x}_{SE(3)_c}} : t \mapsto e^{2\pi jt} \tag{21}$$

Fig. 4 illustrates the feature rotation.

Selecting

$$\lim_{l\to\infty} C_{\vec{\tilde{x}}_i} = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{Bmatrix} \tag{22}$$

enforces $\vec{x}_{SE(3)_i}$ to be distributed on the unit circle.

The Model Space

Deterministic systems are processes producing the exact output from an initial state. This behavior can be observed in mathematics and physics. Particular in the future states of such systems there's no randomness involved. Utilizing physical laws these can be described by differential equations. In nature, especially regarding biological systems like humans, this is rarely the case. Furthermore, this can be expressed better as random variables which may be completely different to various times. Their different quantities take values in the same space solely.
By recalling the classical mechanics of circular motion, the deterministic system of a single harmonic oscillator yields to a 2*nd* order differential equation [11]

$$\frac{d^2c_n(t)}{dt^2} = -(2\pi nf)^2 c_n(t) \tag{23}$$

with the solution

$$c_n(t) = a_n cos(2\pi nft) + b_n sin(2\pi nft) \tag{24}$$

where the constants a_n and b_n are set by the initial conditions of the differential equation. Accounting for non-stationary frequency as a function of time and changes in amplitude and phase, like it is expected for PPGI signals, essentially leads to a stochastic differential equation

$$\frac{d^2c_n(t)}{dt^2} = -(2\pi nf(t))^2 c_n(t) + e_n(t) \tag{25}$$

for each harmonic component. Even when the frequency is discontinuous the signal remains continuous. Fig. 5 shows such a stochastic oscillator with time-varying frequency and amplitude. The stochastic state space for the resonator signal yields to

$$\frac{dx(t)}{dt} = F_0(f(t))x(t) + Le(t), \tag{26}$$

$$c(t) = Hx(t). \tag{27}$$

Interestingly, this corresponds to the kernel formalism of a Gaussian process, where the Gaussian process is constructed as a solution to a *m*th order linear stochastic differential equation [13]. Here, the computational complexity

is linear, whereby the genuine kernel formalism is of cubic nature.

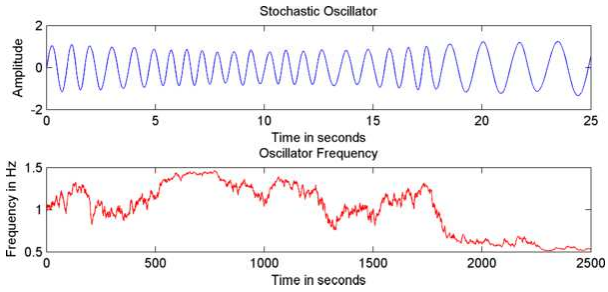

Figure 5. A simulated trajectory of a stochastic oscillator with frequency trace in a possible range typical for a human in resting state.

Since the frequency is unknown, the state space depends on an additional latent variable

$$\frac{dx(t)}{dt} = F_0(\theta)x(t) + Le(t), \tag{28}$$

$$c(t) = H(\theta)x(t) \tag{29}$$

such that

$$\theta \in \Omega = \{\theta^1, \ldots, \theta^S\} \tag{30}$$

forming a Markov chain with transition matrix Π with transition probabilities

$$P(\theta_t^i | \theta_{t-1}^j) = \Pi_{ij}. \tag{31}$$

The solution is given by computing the Gaussian mixture approximation to the joint posterior distribution of the latent variables and states [6]. Pilz *et al*. [27] stated that slow varying drifts can be modeled by a Wiener process, whereby any kind of violation of the smoothness criterion yields to a Poisson process. We didn't consider the drift and jump case since the feature representation based upon local group invariance already accounts for such behavior.

4. Experiments

We divided the experimental procedure into two phases, a data collection and an algorithmic benchmarking.

Data

To justify the data collection effort, we state that to the best of our knowledge other data is either private therefore not accessible, recorded or distributed using image compression techniques or doesn't reflect multiple recordings in order to study the specific questions concerning the mentioned problems that comes with the task of heart rate measurement from face videos. Although Heusch *et al*. [16] and Bobbia *et al*. [7] introduced public data for evaluating remote heart rate measurements along with reference implementation of recent algorithms, we received the video recordings done using image compression techniques or didn't receive access privileges yet. Therefore we designed scenarios ranging from controlled and easy to uncontrolled and more difficult. We built up on the concept and user pool of Pilz *et al*. [27] but re-recorded the users to archive a better time synchronization between the camera and reference device. As result we recorded four different sessions. The first session consists of a resting scenario where no head or facial motion is performed and the illumination is more or less static. In the second session the users are asked to perform head as well as facial motions but the illumination remains static. The third session is performed during an exercise on a bicycle ergometer in a gym where no further instructions are given to the user. The fourth sessions is recorded during an urban conversation including head and facial motions as well as natural varying illumination conditions. Fig. 2 shows some example images taken from the recordings during the different sessions. Every session is recorded over an one minute time span, except the ergometer session which is recorded over a 5 minute period. In total 25 users were ask to participate resulting in an amount of 100 video recordings with approximately 200 min total duration; more than three hours. The data collection consists of 20 male and 5 female in the range of 25-42 years. The majority ethnicity is Caucasian. The camera device is selected as Logitech HD C270 webcam and as reference ground truth measurements we synchronized a common finger pulseoximeter, a CMS50E PPG device, over its serial port communication protocol. The average frame rate of the camera is set to 25 FPS and for the pulseoximeter 60 FPS. The camera video stream is captured uncompressed with auto-exposure and stored into an AVI container. For every captured image the time stamp is stored too. The pulseoximeter signal is stored together with the device's pre-computed heart rate information.

Evaluation

The benchmarking of the described feature and model space is conducted against established methods. To this we count the ICA [28] approach as source separation method, the Spatial Subspace Rotation (SSR) [42] and the Projection Orthogonal to Skin (POS) [41] as feature transform methods [1]. We performed tests for each session respectively as well as a separate run for the POS and the Local Group Invariance (LGI) on the entire data set. The signal processing procedure is selected to be equal for every approach. For each video frame a common face finder is used to pre-select the region of interest. A skin operator is applied onto the region by thresholding the blue- and red-difference chroma components. For the set of obtained RGB-pixels the expectation is computed and stored as three-dimensional time

[1] We also re-implemented other methods [9, 19, 38, 20], since their code is not available. However, we obtained worse results.

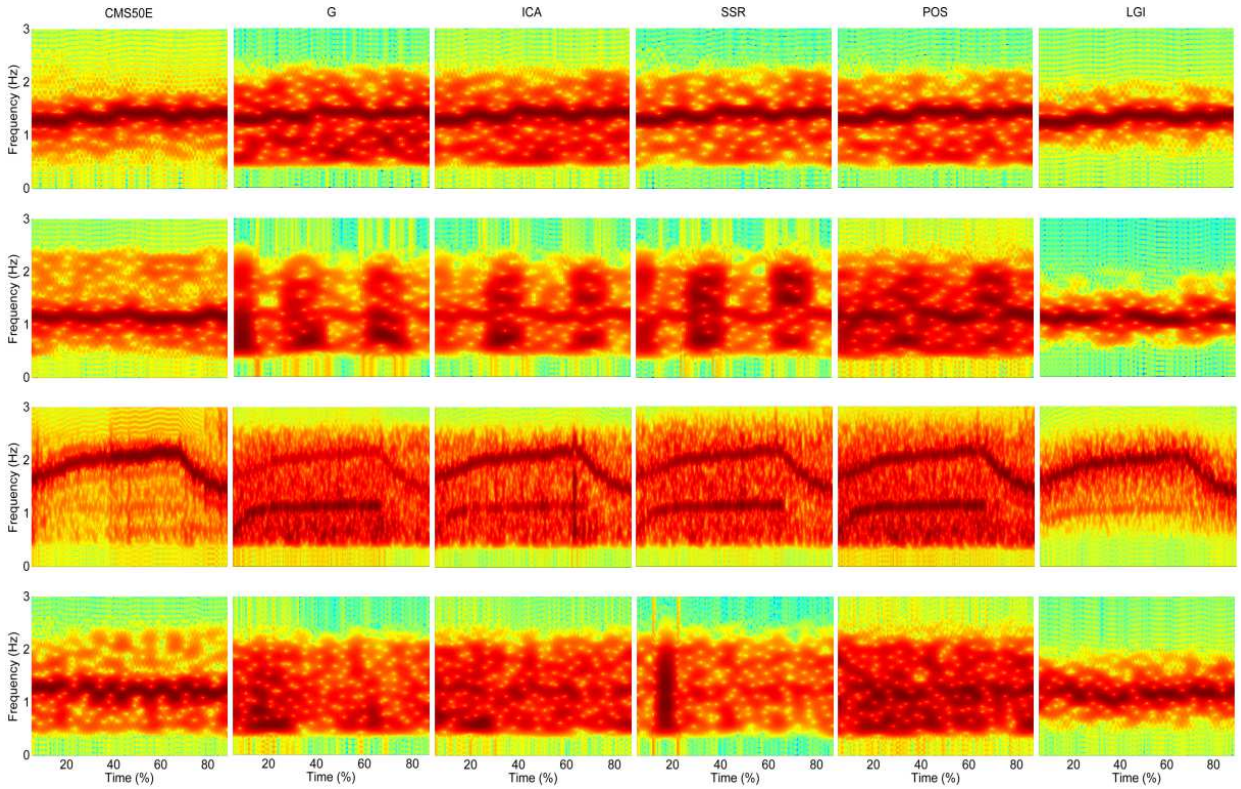

Figure 6. Comparison of a users spectrograms for the different data sessions computed with different algorithms. The rows (from top to down) reflect a head resting session, a head rotation session, a bicycle ergometer exercise and an urban conversation. The columns (from left to right) correspond to the reference PPG device, the green channel, the ICA source separation, the SSR, the POS and the Local Group Invariance (LGI) method. During the resting scenario each algorithm is able to extract reasonable heart rate information. When moving the head nearly every algorithm starts to fail. During the ergometer exercise the interference of the pedal frequency is visible over all algorithms. Here the LGI method benefits from its model space. The urban conversation as fully uncontrolled scenario with rigid and non-rigid facial motion along with illumination changes makes it very difficult to extract robust signals

series for further processing. Each signal obtained by the different algorithms is band-filtered in the range between 0.5 and 2.0 Hz. We increased the upper ranger to 2.5 Hz for the processing of the ergometer sessions. All filtered signals are then analyzed by standard Fourier based spectral method with windows size of 256 samples and overlap of 90 percent. A maximum peak energy criterion is applied over the spectral traces to determine the heart rate candidates. The PPG signals are analyzed in the same way but initially resampled to the camera frame rate. Correlation coefficients are computed against the PPG reference heart rate together with the root-mean-square error (RMSE) for each session and algorithm respectively. For the two full database runs correlation plots and Bland-Altman plots are computed additionally.

Fig. 6 compares the spectrograms for a single users over all sessions against the different algorithms given the reference measurements. Under controlled motion free conditions stable performance is obtained by all algorithms. Under motion scenarios it becomes more and more challenging to robustly extract the heart rate whereby under the fully uncontrolled urban conversation most of the algorithms to fail completely. We observed this behavior during the entire evaluation. The results for each session are presented in Table 1. During the resting scenario the LGI method performs slightly worse. For all other sessions the LGI method archives quite robust results where the others mostly start to fail. Fig. 7 compares the estimation performance between the POS and the LGI approach. The correlation for the POS method is heavily affected by outliers. Although the LGI approach results in a better statistical performance, it shows an estimation bias of approximately 4 BPM. This also explains why the LGI method performs slightly worse during the resting scenario.

5. Conclusions

In this work we have presented a functional approach for the task of heart rate estimation from face videos under the load of nuisance factors. We performed evaluation on data collected under everyday facial motions and environmen-

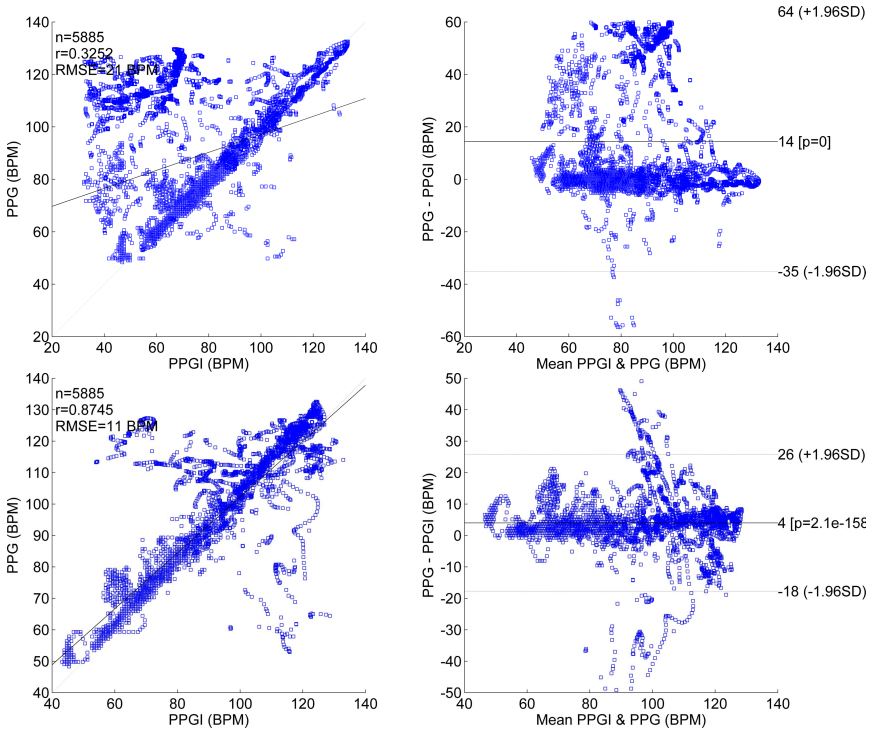

Figure 7. Correlation and Bland-Altman plots for the POS [41] (upper plots) and LGI (lower plots) method over the entire data collection. POS archives a correlation of 0.35 with a RMSE of 21 BPM and LGI a correlation of 0.87 with a RMSE of 11 BPM. Many outliers can be attributed to false predictions during the gym session where the heart rate is confused with the pedal frequency.

Session	ICA	SSR	POS	LGI
Resting	0.97/1.4	0.97/2.0	0.96/2.1	0.96/3.3
Rotation	0.16/10.8	0.51/7.6	0.56/5.3	0.97/2.9
Gym	0.41/16.6	0.08/18.6	0.09/23.1	0.63/13.1
Talk	0.13/23.1	0.14/15.4	0.3/12.5	0.72/4.3

Table 1. Pearson's correlation coefficient and RMSE of prediction for the ICA [28], the SSR [42], the POS [41] and the LGI method.

tal conditions. In contrast to classical optical and spectral interpretation we emphasized the point of view on unsupervised learning of invariant features. A stochastic representation of the heart rate's quasi-periodical process dynamics is obtained by recursive inference. The proposed methodology achieves dramatic improvements in some situations. We experienced cases were the camera based measurements are highly in phase with the reference system but often just not. Therefore, the general error of prediction is still relatively large. With regard to future comparisons we suggest an agreement on a broader necessity for public challenges on heart rate estimation from face videos.

6. Acknowledgments

We would like to thank the MedAix fitness center in Aachen for providing the bicycle ergometers. This work was funded, in part, by the German Federal Ministry of Education and Research (BMBF) under grant agreement VIVID $01|S15024$ and by CanControls GmbH Aachen.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[2] N. Blanik, C. Blazek, C. Pereira, V. Blazek, and S. Leonhardt. Wearable photoplethysmographic sensors: Past and present. *Proc. SPIE 9034, Medical Imaging: Image Processing*, 2014.

[3] C. Blazek and M. Hülsbusch. Assessment of allergic skin reactions and their hemodynamical quantification using photoplethysmography imaging. *Computer-aided Noninvasive Vascular Diagnostics. Vol. 3: Proc. of 11th Int. Symposium CNVD*, 3:85–90, 2005.

[4] V. Blazek. Optoelektronische Erfassung und rechnerunterstützte Analyse der Mikrozirkulations-Rhythmik. *Biomed. Techn.*, 30(1):121–122, 1985.

[5] V. Blazek, N. Blanik, C. Blazek, M. Paul, C. Pereira, M. Koeny, B. Venema, and S. Leonhardt. Active and passive optical imaging modality for unobtrusive cardiorespiratory monitoring and facial expressions assessment. *Assessment. Anesth Analg.*, 124:104–119, 2017.

[6] H. 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.

[7] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 2017.

[8] J. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[9] G. de Haan and V. Jeanne. Robust pulse-rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2014.

[10] G. de Haan and A. van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiol. Meas*, 3(9):1913–1926, 2014.

[11] R. Feynman, R. Leighton, and M. Sands. The feynman lectures on physics vol. 1. chapter 21. *Addison-Wesley*, 1963.

[12] B. C. Hall. Lie groups, lie algebras, and representations. *Springer, New York*, 2003.

[13] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal gaussian process regression models. *IEEE International Workshop on Machine Learning for Signal Processing*, pages 379–384, 2010.

[14] A. Henderson, J. Lasselin, M. Lekander, M. Olsson, S. Powis, J. Axelsson, and D. Perrett. Skin colour changes during experimentally-induced sickness. *Brain, Behavior, and Immunity*, 60:312–318, 2017.

[15] A. Hertzman. Photoelectric plethysmography of the fingers and toes in man. *Exp. Biol. Med.*, 37(3):529–534, 1937.

[16] G. Heusch, A. Anjos, and S. Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017.

[17] M. Hülsbusch. A functional imaging technique for optoelectronic assessment of skin perfusion. *PhD thesis, RWTH Aachen University*, 2008.

[18] A. Lam and Y. Kuno. Robust heart rate measurement from video using select random patches. *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 3640–3648, 2015.

[19] X. Li, J. Chen, G. Zhao, and M. Pietikinen. Remote heart rate measurement from face videos under realistic situations. *IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH*, 2014.

[20] R. Macwan, Y. Benezeth, and A. Mansouri. Remote photoplethysmography with constrained ica using periodicity and chrominance constraints. *BioMedical Engineering OnLine*, 17(1), 2018.

[21] D. McDuff, E. B. Blackford, and J. R. Estepp. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. *12th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 63–70, 2017.

[22] D. McDuff, S. Gontarek, and R. Picard. Remote measurement of cognitive stress via heart rate variability. *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2957–2960, 2014.

[23] A. Moço, S. Stuijk, and G. de Haan. Ballistocardiographic artifacts in ppg imaging. *IEEE Transactions on Biomedical Engineering*, 63(9):1804–1811, 2015.

[24] A. Moço, S. Stuijk, and G. de Haan. Motion robust ppg-imaging through color channel mapping. *Biomed. Opt. Express*, 7:1737–1754, 2016.

[25] H. Molitor and M. Knaizuk. A new bloodless method for continuous recording of peripheral change. *Jour. Phar. Expr. Ther.*, 27:5–16, 1936.

[26] A. Osman, J. Turcot, and R. E. Kaliouby. Supervised learning approach to remote heart rate estimation from facial videos. *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–6, 2015.

[27] C. S. Pilz, J. Krajewski, and V. Blazek. On the diffusion process for heart rate estimation from face videos under realistic conditions. *Pattern Recognition. GCPR 2017. Lecture Notes in Computer Science, vol 10496. Springer*, 10496:361–373, 2017.

[28] M. Poh, J. McDuff, and R. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10 62–10774, 2010.

[29] L. Qi, A. Parthasarathy, and A. E. Rosenberg. A fast algorithm for stochastic matching with application to robust speaker verification. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2:1543–1546, 1997.

[30] G. Ramirez, O. Fuentes, S.Crites, M. Jimenez, and J. Ordonez. Color analysis of facial skin: Detection of emotional state. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 474–479, 2014.

[31] S. Särkkä. Recursive bayesian inference on stochastic differential equations. *PhD thesis, Helsinki University of Technology*, 2006.

[32] B. Schölkopf and A. Smola. Learning with kernels: Support vector machines, regularization, optimization, and beyond. *MIT Press, Cambridge, MA, USA*, 2001.

[33] B. Schölkopf, A. Smola, and V. Vapnik. Prior knowledge in support vector kernels. *Advances in neural information processing systems 10, Cambridge, MA: MIT Press.*, pages 640–646, 1998.

[34] A. Solomonoff, W. Campbell, and I. Boardman. Advances in channel compensation for svm speaker recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 629–632, 2005.

[35] T. Sundelin, M. Lekander, G. Kecklund, E. V. Someren, A. Olsson, and J. Axelsson. Cues of fatigue: Effects of sleep deprivation on facial appearance. *Sleep*, 36(9):1355–1360, 2013.

[36] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807–831, 2014.

[37] A. Trumpp, S. Rasche, D. Wedeking, M. Rudolf, H. Malberg, K. Matschke, and S. Zaunseder. Relation between pulse pressure and the pulsation strength in camera-based photoplethysmograms. *Current Directions in Biomedical Engineering*, 3(2):489–492, 2017.

[38] S. Tulyakov, X. A. Pineda, E. Ricci, L. Yin, J. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2016.

[39] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.

[40] W. Verkruysse, L. Svaasand, and J. Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.

[41] W. Wang, A. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[42] W. Wang, S. Stuijk, and G. de Haan. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Transactions on Biomedical Engineering*, 63(9), 2015.

[43] D. Wedekind, A. Trumpp, F. Gaetjen, S. Rasche, K. Matschke, H. Malberg, and S. Zaunseder. Assessment of blind source separation techniques for video-based cardiac pulse extraction. *J. Biomed. Opt.*, 22(3):35002, 2017.

[44] H. Weyl. The classical groups. their invariants and representations. *Princeton University Press*, 1939.