Contact-free Monitoring of Physiological Parameters in People with Profound Intellectual and Multiple Disabilities

Gašper Slapničar*
Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
gasper.slapnicar@ijs.si

Erik Dovgan*
Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
erik.dovgan@ijs.si

Pia Čuk
Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
pia.cuk@gmx.de

Mitja Luštrek
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
mitja.lustrek@ijs.si

Abstract

This paper presents a contact-free method for physiological parameter estimation in people with profound intellectual and multiple disabilities (PIMD). We used an existing state-of-the-art algorithm Plane-Orthogonal-to-Skin (POS) in order to obtain an initial remote photoplethysmogram (rPPG) reconstruction from facial videos. We enhanced this signal by applying a long-short-term-memory (LSTM) neural network to the initial PPG reconstruction. Evaluation of our method on a public database DEAP has shown heart rate (HR) error of 8.09 beats-per-minute, suprpassing the state-of-the-art POS algorithm implementation, which had error of 13.36 BPM. More importantly, a good correlation between our predictions and ground-truth HRs has been observed. The method is currently being implemented as part of a system which aims to monitor people with PIMD in real time in order to obtain information about their physiological and psychological state and in turn increase their quality of life.

1. Introduction

Physiological signals offer valuable insights into a person's physical and psychological state and well-being. Traditionally, these signals are obtained using wearable devices with embedded sensors, however, there are numerous downsides to such devices. Many people experience discomfort while wearing any device, especially during physical activity, which leads to decreased adherence towards monitoring of physiological parameters. Another issue are potential cables, which severely limit movement. In case of wireless devices, battery life becomes a problem, as usability time is limited and recharging requires removal of the device.

The downsides of wearable sensor devices are even more prominent in specific groups of people with disabilities, such as people with profound intellectual and multiple disabilities (PIMD). These people often reject or show extreme displeasure towards any form of additional device on their bodies. Additionally, due to their disability, they are often nearly immobile and not capable of symbolic communication with other people. Consequently, these people are not capable of understanding and properly using sensor devices on their own. In this case, contact-free solutions are preferable in order to avoid causing discomfort, while still obtaining important information about physiological state.

Physiological state is typically determined by considering a set of common physiological parameters, such as heart rate (HR), heart rate variability (HRV) and respiratory rate (RR). One way to estimate these parameters, is to use a photoplethysmogram (PPG). This physiological signal describes periodic changes of blood volume in the tissue in accordance with heart beats. It is traditionally obtained with a fingertip sensor that returns a high quality signal, or with a wristband which gives a lower quality signal. Both devices leverage a light source and a photodiode to measure the changes in light absorption and reflection of the skin containing different amounts of blood. In addition, PPG can also be obtained from facial camera recordings, as it is reflected in subtle changes in skin color that can be captured on video. This camera-based PPG is commonly re-

*The first two authors should be considered as joint first author.

ferred to in related work as remote PPG (rPPG) or image PPG (iPPG), and has become a prominent research field in the past decade.

In this paper we present a prototype version of a system for contact-free monitoring of HR using rPPG obtained via the Plane-Orthogonal-to-Skin algorithm (POS) and enhanced using signal pre-processing, signal post-processing and a long-short-term-memory neural network (LSTM). It was developed and evaluated using the large open-source DEAP database, however, the main aim is to use it as part of a larger system aimed at helping people with PIMD. Knowing their physiological parameters can potentially give an important insight into their mental state, as the connection between physiological and psychological state is known to exist in people without PIMD [20] and is subsequently postulated to also exist in people with PIMD.

The rest of the paper is organized as follows. Section 2 reviews the related work on rPPG reconstruction. The used datasets and the developed approach for reconstructing the rPPG signal and estimating the heart rate are described in Section 3. Section 4 presents the experiments and the obtained results. Finally, Section 5 concludes the paper with discussion and ideas for future work.

2. Related Work

There are two main approaches for obtaining rPPG, each relying on different underlying physiological phenomena.

The first approach for PPG reconstruction from video analyzes small head movements that are induced by the blood being pumped into the head, as proposed by Balakrishnan et al. [1]. They extracted HR by tracking movement of the head. Afterweard, principal component analysis (PCA) was applied to obtain the component that best corresponds to heartbeats in the frequency domain. It should be noted that such head movements are very subtle and might not be detectable with a low-quality camera, imposing additional hardware requirements on this approach. Additionally, it is quite common for people with PIMD, for example, to have continuous head movements due to their condition, which typically obscure the slight movements due to the influx of blood at each heartbeat.

The second, more common approach focuses on variations in blood volume, which is reflected in the changes of the skin color, as described earlier. To detect the variations of blood volume, tiny changes in color of skin pixels between two sequential video frames are analyzed. For example, Poh et al. [15, 16] applied independent component analysis (ICA) on RGB color signals, which were computed as the average of the red, green and blue intensity of all the skin pixels over time. They then chose the most PPG-like resulting signal returned by ICA.

Lewandowska et al. [11] used PCA instead of ICA to obtain the independent components and subsequently the rPPG signal.

Haan et al. [4] proposed a chrominance-based method, which helps with the motion problems that hinder the separation techniques such as PCA and ICA. They reconstructed the PPG signal by calculating a specific linear combination of normalized RGB traces. The proposed algorithm was shown to work regardless of the color of the illuminant as well as being very robust against motion.

Other approaches do not calculate the average of all skin pixels, but treat each skin pixel independently. For example, Wang et al. [26] tracked the variation of color in each skin pixel independently and individual traces were then overlap-added to obtain rPPG.

Petil et al. [13] reported using basic RGB signals as inputs to ICA to obtain independent components. The average of pixels in the red plane were taken and a set of features was computed from the resulting waveform. These features were then fed to a very simple feed-forward neural network with a single hidden layer to estimate blood pressure, which is a very challenging task even with traditional contact sensors.

Wang et al. [25] introduced a new mathematical model that incorporates pertinent optical and physiological properties of skin reflections. They used the model to design a rPPG method, where a projection plane orthogonal to skin (POS) tone is used for rPPG extraction. This algorithm is explained in more detail in Section 3.1.2.

Chen et al. [2] presented a novel end-to-end deep learning approach, which takes raw video frames as input and estimates rPPG. They proposed a convolutional attention neural network, which features a new motion represention and attention mechanism. It is reported to be robust under heterogeneous lighting and major motions, and to significantly outperform all current state-of-the-art methods. In an effort to replicate this approach as a foundation for us to build upon, we contacted the authors regarding the availability of either their model, or the code to train such a model, in 2018. The authors said that both the code and the data will be released in the future, however, it was not made available to the time of writing of this paper and the communication has since stopped from the authors' side. We have thus tried to replicate this work by following the paper in detail, however, we were unable to reproduce their results on MANHOB-HCI dataset.

Although the presented methods seem promising, an independent evaluation conducted by Heusch et al. [7] on a publicly available dataset showed that they are not accurate enough to be used in real-world scenarios. More precisely, their evaluation re-implemented three state-of-the-art methods for reconstructing PPG from RGB cameras, and the results showed that there is a very low correlation between the reconstructed and ground-truth PPG. Nonetheless, we chose to use the POS algorithm as a starting point of our work, as

it is state-of-the-art and was reported to outperform all other traditional methods, such as ICA, PCA, CHROM, etc. [25]. In order to additionally improve the accuracy of the rPPG reconstruction, we then developed a deep-learning-based approach described in the following sections.

3. Materials and Methods

In this section we present the data used as input to our pipeline and the developed deep-learning-based approach for reconstructing rPPG and calculating physiological parameters. Schematic representation of our pipeline is given in Figure 1 and is discussed in more detail in the subsequent sections.

3.1. Materials

Here, we describe the dataset on which the developed approach was evaluated, and the POS algorithm, an existing state-of-the-art algorithm for rPPG reconstruction, which is incorporated in our approach as one step of the pipeline.

3.1.1 The INSENSION and DEAP datasets

The presented approach for reconstructing rPPG is part of the INSENSION project[1], which aims at detecting non-symbolic behaviour signals, including physiological parameters, of people with PIMD, in order to determine their mental state and communication attempts. The project is currently collecting data of people with PIMD who will be used for evaluating and tuning the developed algorithms. An example of already collected video of a person with PIMD is shown in Figure 2. Since the database of videos of target users and their corresponding PPGs is currently under collection, we had to use an already available database.

We evaluated our approach on a public multimodal dataset for analysis of human affective states called "Database for Emotion Analysis using Physiological signals" (DEAP) [10]. This dataset is commonly used in related work on physiological signal analysis and emotion detection (see, for example, [6, 12, 27]). For the evaluation of our approach, facial videos of recorded subjects together with their ground-truth PPG were used. These videos were recorded with a SONY DCR-HC27E camcorder on a tripod placed behind a computer screen. In addition, the PPG signal was recorded with a BioSemi fingertip device. The recorded subjects were watching excerpts of music videos that aimed at eliciting specific emotions. In total, videos of 22 persons watching 40 one-minute music videos were used in the presented experiment. Example frames from the DEAP dataset are shown in Figure 3.

[1] http://www.insension.eu/

3.1.2 The POS algorithm

The presented approach applies a neural network to enhance the reconstructed rPPG obtained with the chosen state-of-the-art algorithm, i.e., the Plane-Orthogonal-to-Skin algorithm (POS) [25]. This algorithm computes rPPG in two steps. In the first step, (X_S, Y_S) is calculated as:

$$X_S = G_N - B_N$$

$$Y_S = -2R_N + G_N + B_N,$$

where $[R_N, G_N, B_N]$ are zero-mean-scaled, detrended and filtered color signals R, G and B. In the second step, rPPG is obtained as:

$$rPPG = X_S + \alpha Y_S$$

$$\alpha = \frac{\sigma(X_S)}{\sigma(Y_S)},$$

where σ is the L-point standard deviation with L corresponding to the number of samples contained in 1.6 seconds of video. Such a number of samples was empirically selected in order to contain at least one heart beat.

Figure 4 shows an example of the ground-truth PPG from the DEAP database, together with rPPG, reconstructed by POS. This example shows that POS does not accurately reconstruct rPPG during the entire signal. This is also confirmed by the evaluation results presented in Section 4. To improve the rPPG reconstruction, we enhanced the methodology with pre-processing of the RGB signals, post-processing of the POS output and a neural network described in Section 3.2.

3.2. Methods

Our enhancement approach consists of the following steps:

1. Face detection and skin segmentation
2. Signal pre-processing
3. The POS algorithm for obtaining rPPG (described in Section 3.1.2)
4. Neural network rPPG enhancement
5. Physiological parameter calculation

These steps are described in the following sections.

3.2.1 Face detection and skin segmentation

The face, more precisely, its bounding box was detected with a pre-trained cascade object detector based on the Viola-Jones algorithm [24]. However, due to time constrains, it was infeasible to apply this algorithm on each

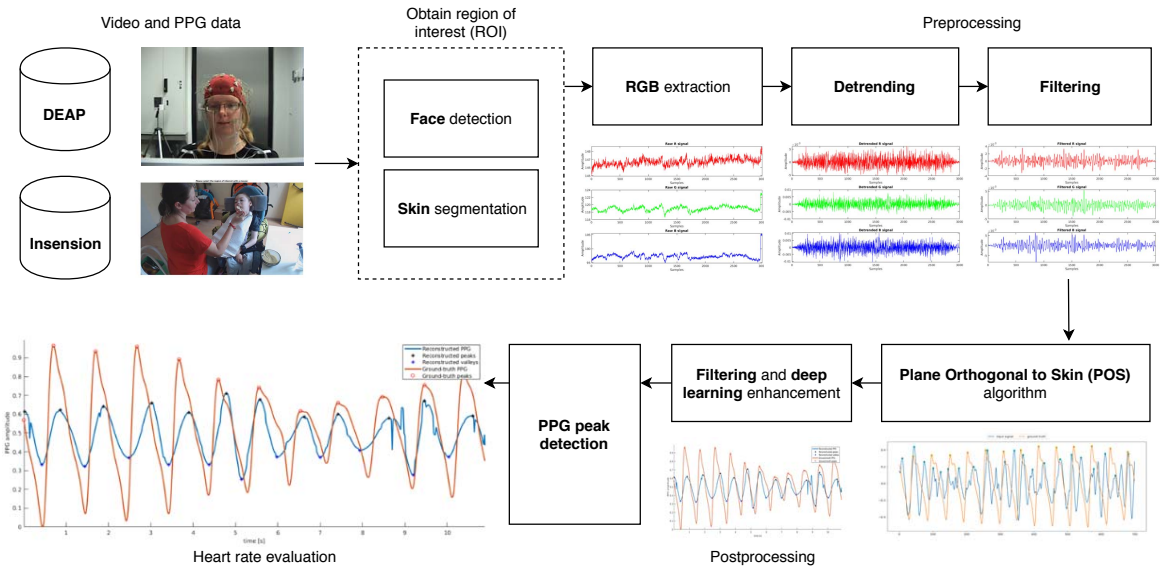

Figure 1. Schematic representation of our pipeline. The x axes on all graphs denote time, while y axes denote RGB value on the top three graphs and rPPG amplitude on the bottom three.

Figure 2. An example of a frame in a video of the target users, i.e., people with PIMD. Informed consent was obtained from legal guardians of all the participants in the project for this data to be used in both research and publications related to this project.

frame. Therefore, only the initial three frames were processed by this algorithm. If the algorithm did not detect the face in one of these initial frames, the given trial video was discarded. On the other hand, when the face was detected, its bounding box was slightly extended and used for the entire video. Note that the recorded subjects did not move significantly, but only sightly, therefore the extended bounding box was able to handle such small movements and contain the entire face in every frame. More precisely, the bounding box was increased by 10 % on its top, left and right sides, and by 20 % at the bottom. A larger extension at the bottom also enabled to capture the neck skin, which was useful for the POS algorithm.

The next step consisted of segmenting the skin. This step

Figure 3. Example frames from the DEAP dataset.

filtered out common things, such as glasses or facial hair, as well as the recording devices and their cables (see Figure 3). The segmentation was done with a HSV (hue, sat-

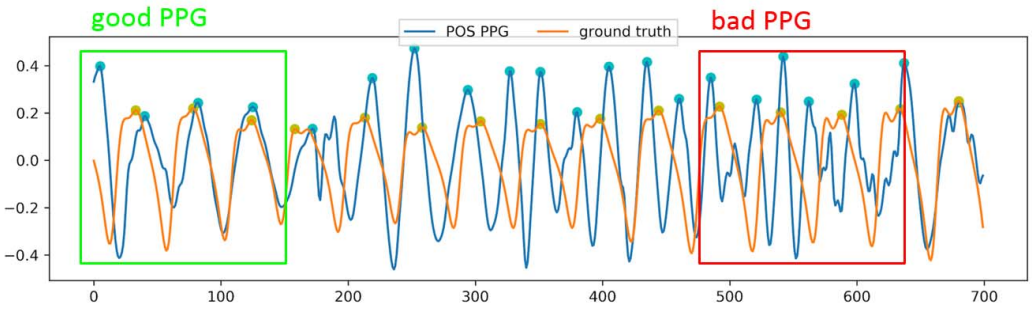

Figure 4. Result of the POS algorithm compared to the ground truth.

uration, value) masking approach, as is common in related work [23]. Acceptable ranges were set to [0, 46] for hue, [23, 123] for saturation, and [88, 255] for value. All pixel values outside these ranges was discarded. An example of face detection and skin segmentation result can be seen in Figure 5.

These approaches work well for the DEAP dataset where the subjects are alone, well-exposed and stationary, however, there are some issues when working with our real-life data from people with PIMD. First is the fact that Viola-Jones algorithm detects the face quite well, but we often have scenarios on our recordings where there is heavy movement or there are more people in the frame, typically caregivers alongside the PIMD person. Thus, we must handle detection of several faces and, more importantly, identification of people, so more sophisticated approaches are already being considered – Histogram of Oriented Gradients proposed by Dalal et al. [3] for face detection and FaceNet neural network [17] for encoding and recognizing faces. Additionally, the skin detection thresholds did not work that well on our recordings, especially because the color of the wall behind the person with PIMD is very similar to their skin, as seen in Figure 2. We will thus also investigate additional skin segmentation methods as part of our future work.

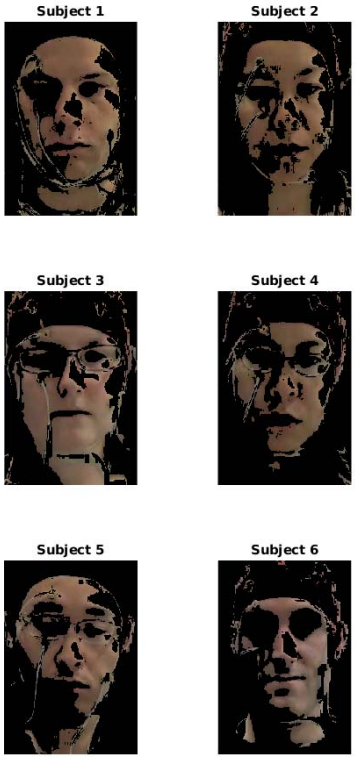

Figure 5. An example of the result of the face detection and skin segmentation steps.

3.2.2 Signal pre-processing

The initial signals were computed as spatial average of the skin pixels at every frame. This was done for the red, green and blue color channels independently. These signals were noisy thus the following pre-processing steps were applied.

Firstly, each RGB signal was processed by a zero-mean-and-scaling technique [4], which is a common step in signal pre-processing since it helps alleviating some edge effects with certain filters. These occur due to the filter window passing the edge of the signal and padding with zeros (by default). If the signal is made to be zero-mean, this padding usually represents less of an issue since it does not deviate very much from the signal values, and the shape of the waveform is better preserved. Another way to resolve this is to simply repeat the final value of the signal for some samples or extrapolate the signal, then do the filtering, and finally cut away the added extra samples.

Secondly, the scaled RGB signals were detrended using the Smoothness Prior Approach (SPA) [19]. More precisely, the signals were detrended piece-wise using overlapping windows with 50% overlap and glued using Hamming windows. This step is useful when short segments are sometimes above the mean and sometimes bellow, as in the

dataset that we used.

Finally, the scaled and detrended RGB signals were filtered using a moving average filter with window length of five samples and a fourth-order Butterworth band-pass filter with cutoff frequencies at 0.5 Hz and four Hz.

3.2.3 Calculation of rPPG with POS

The pre-processed RGB signals were processed by the POS algorithm as described in Section 3.1.2 in order to obtain the POS rPPG signal.

3.2.4 Signal post-processing

Finally, since the POS algorithm reconstruction also contains some artefacts and frequency noise, we have applied some post-processing. As no baseline drifting or larger trends were observed, we have applied normalization to [0, 1] range and then filtered the resulting signal with an adaptive band-pass filter. More precisely, a two-step wavelet filter was applied. First we performed continuous wavelet transform of rPPG and filtered wavelet coefficients with a wide Gaussian window centred at scale corresponding to the maximum of squared wavelet coefficients averaged over a running window of 15 seconds. Then we applied a usual Gaussian filter. The filtered signal was reconstructed by performing the inverse continuous wavelet transform. Details are given by Unakafov [22].

3.2.5 Enhancing POS rPPG with LSTM and CNN neural networks

Two neural network architectures have been tested for rPPG enhancement. First, a Long-Short-Term-Memory neural network (LSTM) [14] was tested, as it is known to be good for capturing temporal dependencies in the data. Additionally, a convolutional autoencoder [14] was also tested, as it is developed specifically to learn shapes and encode them into an embedding.

The LSTM-POS method that includes LSTM layers for improving the rPPG signal obtained with the POS algorithm was our initial attempt. This network consists of two LSTM layers combined with dropout layers [18]. On top of them, a fully-connected layer combines the data from the lower layers into the enhanced rPPG. The input to the network is a window of 50 samples of the POS rPPG signal, which corresponds to one second. The output is a single enhanced rPPG sample, so the window advances by one sample. The network architecture is shown in Figure 6.

We also designed the CNN-POS method that consists convolutional layers for enhancing POS rPPG. This convolutional neural network (CNN) combines encoding layers with their mirrored version, i.e., decoding layers. More precisely, the encoding layers consist of 3 convolution layers, while the decoding layers contain 3 deconvolution layers. Similarly to the LSTM, this neural network takes as input a window of 50 samples, i.e. one second. The output of the network is the improved rPPG for the entire window of 50 samples, which is in contrast to LSTM, which outputs only a single enhanced rPPG sample. The encoding-decoding network architecture is shown in Figure 7.

To accelerate training, every other frame of a training instance was discarded, i.e. 25 frames were used per instance instead of 50. Our experiments have shown that this does not lead to a decrease in the network's accuracy, while reducing training time by more than 50%. For evaluation, all 50 frames of each instance were used. All networks were trained with MSE loss. The optimizers used were RMSprop [21] for LSTM and Adam [9] for CNN, with a learning rate of $\alpha = 0.001$. The dropout rate for the LSTM was $p = 0.2$. Batch normalization [8] with standard parameters was added to the convolutional layers of the CNN.

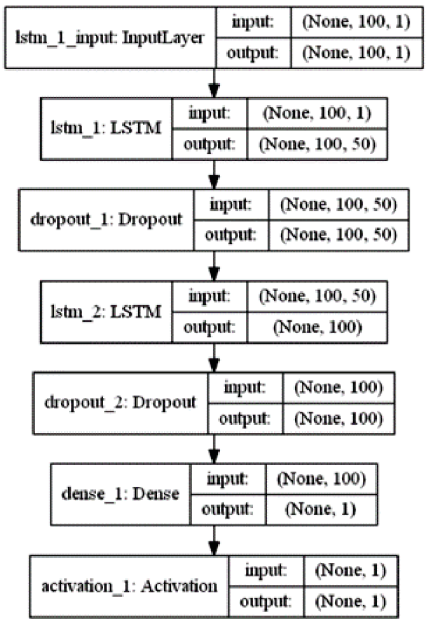

Figure 6. Architecture of the LSTM neural network as part of the LSTM-POS method.

3.2.6 Physiological parameter calculation

The last step of the presented approach estimates the heart rate from the reconstructed rPPG. Additional physiological parameter estimation will be attempted in the future, such as HRV and RR. Note that all these parameters (including the HR) require precise waveform reconstruction or heart beat detection.

The HR was estimated with a robust peak detection al-

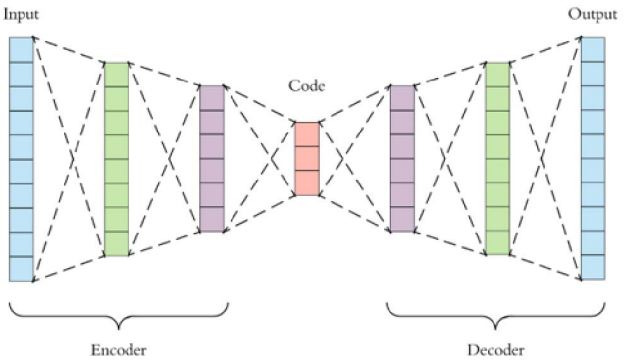

Figure 7. Conceptual architecture of the convolutional autoencoder as part of the CNN-POS method.

gorithm customized for the PPG signal obtained under challenging conditions [5]. This algorithm counts the number of dominant systolic peaks, corresponding to the systole in the cardiac cycle, within a window of PPG of such length that at least a few cycles are captured. In our case, it was applied to one minute windows, directly giving the HR in beats per minute (BPM).

4. Results

We evaluated our approach by comparing the enhanced signals obtained from the LSTM and CNN neural networks, and the signals reconstructed by applying the POS algorithm, against the ground-truth PPG measured with a professional fingertip PPG sensor. The same peak detection algorithm was applied to all four signals and the mean absolute error (MAE) between the number of detected peaks in the reconstructed signals and in the ground truth was calculated. In addition, we visually inspected the performance of the peak detector on the ground truth signals, where it worked almost perfectly, except for some edge cases in which it sometimes detected both systolic and diastolic peaks. These cases were excluded from further evaluation. Moreover, we also computed the correlation between the predictions and ground truth HRs, as MAE itself does not convey the full story. Good correlation is important, as it shows that LSTM-POS does not simply predict absolute HR but also predicts the changes correctly, demonstrating that the system did not only overfit to the dataset mean HR, but is capable of reconstructing the shifts away from the mean.

16 subjects of the 20 considered from the DEAP dataset were used for training, 2 were used for validation and the final 2 for testing. Due to aforementioned reasons (incorrect ground truth peak detections) the remaining 2 subjects in were not used in the evaluation. The split by subject ensures that instances corresponding to the same subject never appeared in more than one of the subsets in order to prevent overfitting. A baseline was established by using a dummy regressor that always predicted the mean HR of the training set. The errors and correlations are given in Table 1. These results show that using the POS signal as input into the LSTM network has the highest correlation with the ground truth, and close to the lowest error.

The LSTM-POS method as the best-performing method was further evaluated in terms of its predictions of HR for all instances of the DEAP dataset, barring those excluded due to incorrectly detected peaks in the ground truth. The obtained predictions and the corresponding HR from the ground truth were ordered from the lowest ground-truth HR to the highest. These data are shown Figure 8. Additionally, a linear regression line was computed from the predicted HR and included in this figure. These results visualize a significant correlation between the predictions and the ground truth. In addition, the linear approximation of the predicted HR matches the ground truth HR accurately, showing that the LSTM-POS method captured changes in HR adequately and is able to output sensible predictions in a large range, not just around the mean.

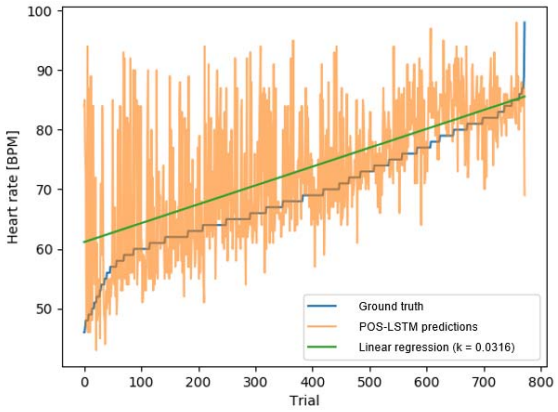

Figure 8. Ground-truth HR sorted from lowest to highest alongside corresponding HR predictions of our LSTM-POS method. A linear regression line computed from the predictions is also drawn, which shows a good correlation.

An example of the enhanced LSTM-POS rPPG alongside its pre-enhanced POS version and the ground truth is shown in Figure 9. One can observe some incorrectly reconstructed peaks from the POS algorithm getting suppressed by the LSTM network. Additionally, the temporal alignment of peaks from the enhanced signal with those from the ground truth has improved compared to the peaks from POS rPPG. Correct alignment is crucial for HRV, which will also be evaluated in future work.

Method	Mean of predictions [BPM]	MAE [BPM]	Correlation
Baseline	70.42	8.36	Inapplicable
POS	81.58	13.36	0.27
CNN-POS	71.72	**7.92**	0.24
LSTM-POS	73.50	8.09	**0.40**

Table 1. Results of evaluated methods for HR estimation from rPPG. The LSTM-POS method obtained the best results due to the highest correlation with the ground truth, which outweighs the slight disadvantage in MAE compared to CNN-POS.

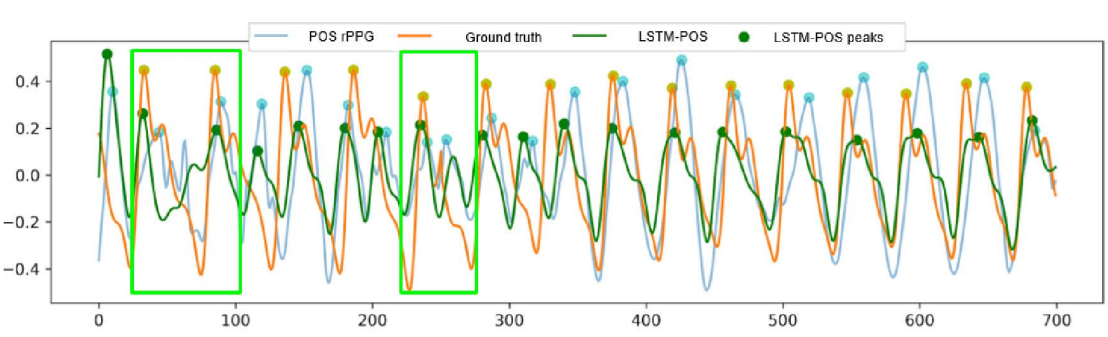

Figure 9. rPPG obtained with the LSTM-POS method and the POS algorithm, and ground-truth PPG. The green rectangular areas show a notable improvement as it suppresses some incorrect peaks in the rPPG obtained with only the POS algorithm. Better temporal alignment of LSTM-POS peaks with the ground truth can be seen in most cases.

5. Discussion and Conclusion

The results presented in Section 4 show that the LSTM-POS method outperforms the existing POS algorithm as well as the CNN-POS method. More precisely, Table 1 shows that HR obtained with the LSTM-POS method has the highest correlation with the ground truth HR, and close to the lowest error. It should be noted that a high correlation of HR is crucial for assessment of a person's physical and psychological state and well-being. Such an assessment will not require an exact prediction of HR, but rather the recognition of increase or decrease of HR. The ability of LSTM-POS to detect changes of HR is shown in Figure 8. More precisely, the linear regression line in this figure was computed from the LSTM-POS HR predictions and illustrates that the system is capable of reconstructing shifts in HR away from the dataset mean.

The ability of the LSTM-POS method to recognize changes in HR is very promising for the purposes of determining the psychological state of people with PIMD within the INSENSION project. Note that people with PIMD are often not capable of symbolic communication with their environment. As a consequence, detection of nonsymbolic communication such as changes of psychological state is of key importance for understanding the needs, wishes and preferences of people with PIMD. This recognition will enable us to provide better assistance to these people, which will include, for example, timely detection of uncomfortable situations and personalized actions to make the person with PIMD feel comfortable again. As a result, the quality of life of people with PIMD will increase. In addition, the INSENSION system will be also accessible to caregivers who will get better insights in peoples' nonsymbolic communication in order to provide better care and further increase patients' quality of life.

However, we have not yet been able to evaluate the LSTM-POS methods on people with PIMD due to the fact that their data are still being collected. Therefore, our future work will include the evaluation of the developed method on the INSENSION dataset and further method development in case of a decrease in performance with respect to the DEAP dataset. Similar to HR, we will attempt to estimate other physiological parameters from rPPG, such as HRV and RR. Based on these parameters, we will develop an approach for the determination of psychological state of people with PIMD, which will aim at understanding their needs, wishes and preferences.

6. Acknowledgments

This work is part of the *INSENSION* project that has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 780819. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research. Finally we would like to acknowledge the help of Andrejaana Andova, who previously also helped with this research and was the first to test the LSTM networks.

References

[1] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, June 2013. 2

[2] W. Chen and D. McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018. 2

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. 2005. 5

[4] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering*, 60:2878–2886, 2013. 2, 5

[5] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans. Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. *PLoS One*, 8(10):e76585, 2013. 7

[6] D. Girardi, F. Lanubile, and N. Novielli. Emotion detection using noninvasive low cost sensors. In *Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 125–130, 2017. 3

[7] G. Heusch, A. Anjos, and S. Marcel. A reproducible study on remote heart rate measurement. *CoRR*, abs/1709.00962, 2017. 2

[8] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 6

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[10] S. Koelstra, M. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: A database for emotion analysis using physiological signals. *IEEE Transaction on Affective Computing*, 3(1):18–31, 2012. 3

[11] M. Lewandowska, J. Rumiski, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, Sept 2011. 2

[12] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, and A. K. Nandi. Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction. *Concurrency and Computation: Practice and Experience*, 30(23):e4446, 2018. 3

[13] O. R. Patil, Y. Gao, B. Li, and Z. Jin. CamBP: A camera-based, non-contact blood pressure monitor. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 524–529, New York, NY, USA, 2017. ACM. 2

[14] J. Patterson and A. Gibson. *Deep Learning: A Practitioner's Approach*. O'Reilly, Sebastopol, 2017. 6

[15] M. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, Jan 2011. 2

[16] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010. 2

[17] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 5

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 6

[19] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, 49(2):172–175, 2002. 5

[20] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, 36(2):747–756, 2012. 2

[21] T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012. 6

[22] A. M. Unakafov. Pulse rate estimation using imaging photoplethysmography: Generic framework and comparison of methods on a publicly available dataset. *Biomedical Physics & Engineering Express*, 4(4):045001, 2018. 6

[23] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Proceedings of Graphicon*, pages 85–92, 2003. 5

[24] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages I511–I518, 2001. 3

[25] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017. 2, 3

[26] W. Wang, S. Stuijk, and G. de Haan. Exploiting spatial redundancy of image sensor for motion robust rPPG. *IEEE Transactions on Biomedical Engineering*, 62(2):415–425, Feb 2015. 2

[27] X. Zhuang, V. Rozgič, and M. Crystal. Compact unsupervised EEG response representation for emotion recognition. In *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 736–739, 2014. 3