Reconstructing PPG Signal from Video Recordings

Gašper SlapniÇcar, Erik Dovgan, Andrejaana Andova, Mitja Luštrek
Department of Intelligent Systems
Jozef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana, Slovenia
gasper.slapnicar@ijs.si
erik.dovgan@ijs.si
andrejaana.andova@ijs.si
mitja.lustrek@ijs.si

ABSTRACT: *Physiological signals give important insight regarding some one's health. It would be in the interest of people to monitor such signals without any wearable devices. We used RGB camera recordings of faces to reconstruct the PPG signal, which can be used to monitor many physiological signals such as heart rate, breathing rate, blood pressure, etc. A deep learning method was developed to enhance existing state-of-the-art methods. This method uses the output of an existing method as an input into a LSTM neural network, which substantially improves the reconstruction of PPG.*

Keywords: Remote PPG, Signal Processing, Deep Learning

Received: 6 April 2018, Revised 18 April 2018, Accepted 25 April 2018

1. Introduction

Physiological signals, such as photoplethymogram (PPG), are traditionally measured using wearable devices like cus and wristbands. While such devices are rather unobtrusive, it would be preferable to omit them completely. This can be achieved with the use of contact-free devices such as RGB cameras, which can blend into the environment, allowing for remote physiological signal reconstruction. An example group for whom such a system would be useful are people with profound cognitive impairment, who are the subjects participating in the INSENSION project1, for which our system is being developed.

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

This paper aims to compare and enhance existing approaches for reconstructing PPG from video data, i.e., remote PPG (rPPG). The PPG signal describes the changes of blood volume in the skin tissue, which corresponds to the heart periodically pushing the blood towards the periphery of the body with each beat. When skin tissue gets filled with blood, it becomes slightly darker and absorbs more light. The light source in contact sensors (e.g., wristbands) is concentrated and constant (a light emitting diode - LED) and enables high-quality PPG reconstruction. Reconstructing rPPG from a camera recording is more difficult as the source of light is most commonly the sun or the lighting of a room. This makes such an approach more sensitive to environmental conditions and less accurate compared to contact sensors. rPPG reconstruction would allow for estimation of several physiological parameters, such as heart rate (HR), breathing rate, heart-rate variability and blood pressure, without a wearable device.

The rest of the paper is organized as follows. Section 2 reviews the related work. The methods for reconstructing the PPG signal are described in Section 3, while the experiments and results are discussed in Section 4. Finally, Section 5 concludes the paper with ideas for future work.

2. Related Work

There are two main approaches for reconstructing rPPG, which are based on different underlying physiological phenomena.

The first approach focuses on variations in blood volume, which is reflected in the changes of the skin color. To detect the variations of blood volume using non-contact sensors (camera), tiny changes in RGB intensity of the skin pixels between two sequential video frames are analyzed. For example, Poh et al. [9, 10] applied independent component analysis (ICA) on the 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. Lewandowska et al. [5] used principal component analysis (PCA) instead of ICA to reconstruct the PPG signal. Haan et al. [2] reconstructed the PPG signal simply by calculating a specic linear combination of the obtained RGB traces. Other approaches do not calculate the average of all skin pixels, but treat each skin pixel independently. For example, Wang et al. [11] tracked the variation of color in each skin pixel independently and chose the most PPG-like signal afterwards. The changes of the skin pixel values were also tracked to reconstruct the PPG signal [12]. Petil et al. [7] used the basic RGB signals as inputs to a neural network to reconstruct various physiological signals. Another example by Wu et al. [13] amplified all the color changes of the facial pixels to follow the blood ow in these pixels. Although the presented methods seem promising, an independent evaluation conducted by Heusch et al. [3] on a publicly available dataset showed that they are not accurate enough to be used in real-world scenarios. More precisely, this 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.

The second approach for PPG reconstruction from video analyzes the small head movements that are induced by the blood being pumped into the head. Such a study was conducted by Balakrishnan et al. [1], however, it should be noted that such movements are very subtle and might not be detectable with a low quality camera, imposing an additional hardware requirement on this approach.

3. Reconstructing PPG with Vision Based Methods

This section presents the developed deep-learning-based method for reconstructing the PPG signal from video data. This method enhances the signal reconstruction of an existing state-of-the-art method, as none of them were satisfactory. We first present the state-of-the-art methods used in the evaluation. All of these methods have a similar preprocessing step, which is presented in Section 3.1. The steps specific for each of coh these methods are presented in Section 3.2. Finally, in Section 3.3 we present the developed method that takes as input the PPG reconstructed with an existing method and returns an enhanced reconstruction of PPG.

3.1 Preprocessing of Video Data

The first preprocessing step consists of the detection of the subject's face as the region of interest" (ROI). For detecting the face, we used the Haar cascades, implemented in the OpenCV library 2. More precisely, the video was segmented into individual frames and only the selected face ROI was cropped from each frame.

[2] https://opencv.org

The second step of video preprocessing aims at discriminating between skin and non-skin pixels. For this purpose, we implemented two classification methods. The first method transforms the RGB color space into the YCbCr color space, which contains less redundant information. Pixel values are then classified as either skin or non-skin using thresholds. This method is fast, simple and works well on the test dataset, however it probably does not generalize well to datasets where the degree of variation of skin colors and shades is higher. The second method applies one-class support vector machines (SVM) to classify the skin pixels. It learns a decision function for novelty detection from positive examples (corresponding to skin pixels), which are obtained from the forehead region of the first three frames of each video. New data is then classified as similar (skin) or different (not skin) to the training set. The forehead region is detected as the facial area of fixed dimensions above the eyes, which can be easily detected using OpenCV. The SVM method produces worse results than the threshold based method, but generalizes well for various skin colors and shades. Both skin classification methods incorrectly classify some of the non-skin pixels as skin. To avoid false positives, we selected only the pixels that are most likely to actually be skin. We did this by calculating the mean value of all the skin pixels returned by the classier and then removing the outlier pixels with respect to the mean in the YCbCr color space.

3.2 State-of-the-Art Methods

We have evaluated a set of state-of-the-art methods. These methods can be classified as color-based or movement-based as described in Section 2.

Poh-et-al Method: This is a color-based method that sequences the mean value of the red, green and blue intensity of all the skin pixels to create three different color traces. Since all the traces contain some information about the blood flow, it first normalizes them and then transforms them with ICA using the Fast ICA algorithm [4, 9, 10]. This method returns three signals, so we choose the one with most frequencies in the range [0.6 Hz, 4 Hz], i.e., the frequency range of PPG. This is done by analyzing the power spectrum of each output signal.

Haan-et-al Method: This is also a color-based method which, similarly to the previous method, uses the mean of the red, green and blue intensity of all the skin pixels [2]. It then creates a linear combination from the red, green and blue traces, resulting in two new traces X and Y, calculated as: $X = 3R - 2G$; $Y = 1{:}5R + G - 1{:}5B$. The X and Y traces are then filtered and combined to reconstruct the PPG signal. In our experiments, we used the method implementation from the BOB library[3].

Wang-et-al Method: This color-based method uses all the skin pixels from an individual frame to define the color space of frames [12]. By tracking the changes in this space, we reconstruct the PPG signal. To accomplish this, a covariance matrix is computed. This covariance matrix changes for each frame due to the blood owing into the skin. By calculating the eigenvectors of the original frame and the eigenvalues of the covariance matrix, we get a representation of the color space for the skin pixels. The rotation between two eigenvectors of sequential frames represents the changes of the color space. This rotation is also related to different relative PPG contributors. Therefore, by concatenating the rotation between the first opposing to the second and the third eigenvector, PPG-like traces are retrieved. The eigenvalues are also influenced by the pulsatile blood and are thus used to normalize the PPG-like signals. As for the previous method, we also used the method implementation from the BOB library[3].

Balakrishnan-et-al Method: In contrast to the previously presented methods, this is a motion-based method, since it focuses on the oscillations of the head [1]. This method does not need to detect skin pixels, therefore, the second step of data preprocessing is skipped. To reconstruct the oscillations, the Lucas-Kanade ow-tracking algorithm [6] is applied, which tracks the flow of the head movements in the vertical direction. The oscillation signals are then filtered using a band-pass filter with the frequency interval [0.6 Hz, 4 Hz], i.e., the frequency range of PPG. Afterwards, PCA is applied to select the most PPG-like signal.

3.3 Deep-Learning-Based Method

We developed a new method for reconstructing the PPG signal, which takes the PPG reconstructed by an existing method as the input, and outputs an improved reconstruction of the PPG signal. To achieve this, it applies deep learning, which has recently shown superior performance in machine learning on many domains compared to traditional approaches.

To build the deep learning model, we used a Long-Short Term Memory (LSTM) network [8]. Its architecture comprised two LSTM layers and one fully-connected layer. The window length was set to 100 samples, i.e., Five seconds. Each layer had 50 LSTM units, each taking input of length 100 and returning output of the same length, as shown in Figure 1. The output of the

[3]https://pypi.org/project/bob.rppg.base

Wang-et-al method [12] has been selected as the input to the LSTM network.

4. Experiments and Results

In order to evaluate the quality of methods described in Section 3.2 and our method, the reconstructed PPG was compared with the ground truth obtained with a fingertip PPG sensor.

4.1 Materials and Experimental Setup

The existing methods and the developed method have been evaluated on the COHFACE dataset4. This dataset consists of 160 videos from 40 different subjects with corresponding synchronized PPG collected with a fingertip device. The mean value of heart rate over the whole dataset is 70.25 beats per minute (BPM) with the corresponding standard deviation of 11.56.

A preliminary test has been done to select the best skin classification method. The evaluated skin classification methods were threshold-based method and SVM-based method as described in Section 3.1. Examples of the masks returned by both methods are shown in Figure 2. The results show that the threshold-based method is better than SVM-based. However, it should be noted that the selected thresholds were fitted to the selected dataset, therefore, the method might not generalize well to other data.

To evaluate the developed method, a leave-one-subject-out experiment was conducted with the aim of testing its predictive performance and generalization capability. To this end, mean absolute error (MAE) and mean squared error (MSE) were used as metrics. Additionally, to evaluate the quality of HR predictions, we computed the number of peaks in the reconstructed signals and compared it to the number of peaks in the ground truth PPG.

4.2 Experimental Results

Results of all the evaluated methods, as well as the developed method, are given in Table 1. All three previously mentioned

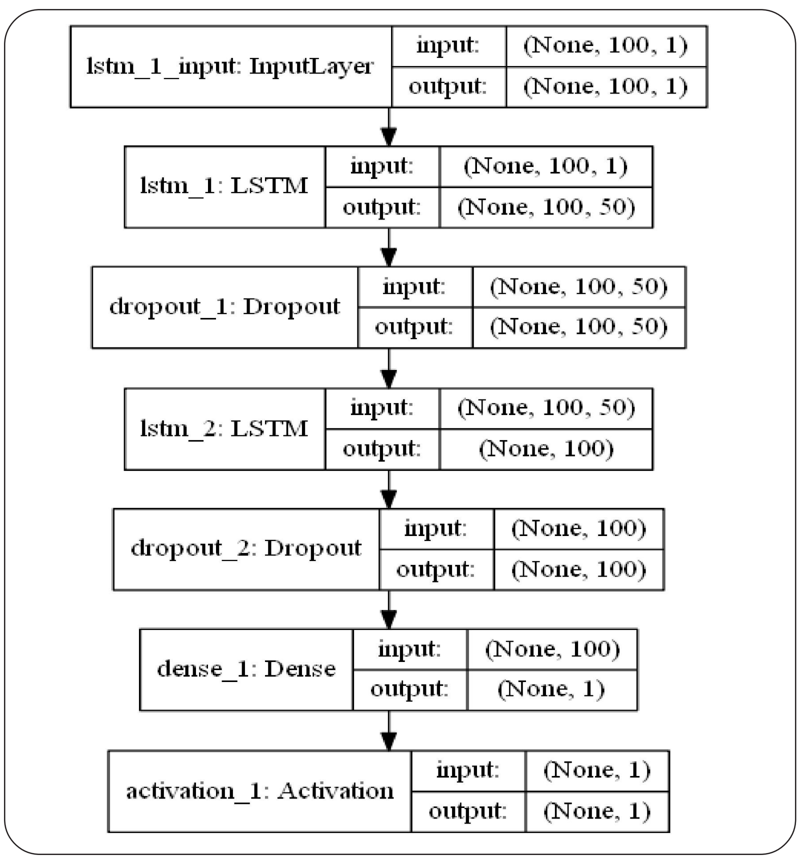

Figure 1. The architecture of the network used in the Deep-Learning-Based method

(a) (b)

Figure 2. Classified skin using the (a) threshold method, and (b) machine learning method

metrics are reported, i.e., MAE, MSE and HR MAE. Note that the heart rate MAE of the baseline is 9.67 BPM, while our method achieves 8.75 BPM. In addition, the first 10 seconds of the reconstructed PPG using each of the methods on a subset of videos are shown in Figures 3-5.

The results show that the Deep-Learning-Based method produces better reconstruction of the PPG signal, as the error between the estimated and actual HR is the lowest. Table 1 also shows that the developed method outperforms state of-the-art methods on the COHFACE dataset by a notable margin.

5. Conclusions

We presented a new approach for the reconstruction of the PPG signal from video data. This approach enhances state of-the-art methods with a deep learning model. It has been evaluated on the COHFACE dataset and the results show that it improves the PPG reconstruction with respect to the state-of-the-art methods.

However, the reconstructed PPG signal is still noisy and it would thus be difficult to estimate any physiological parameters from it, which will need to be improved in future work. Additionally, higher quality of the recordings, especially regarding the lighting conditions, will be evaluated with the aim of obtaining better results.

Method	**MAE (Signals)**	**MSE (Signals)**	**MAE [BPM] (Heart Rates)**
Poh-et-al	0.04	0.15	42.00
Haan-et-al	0.16	0.04	20.75
Wang-et-al	0.16	0.40	11.73
Balakrishnan-et-al	0.16	0.04	39.00
Deep-Learning	**0.04**	**0.01**	**8.75**

Table 1. Comparison between state-of-the-art methods and the developed method

[4] https://www.idiap.ch/dataset/cohface

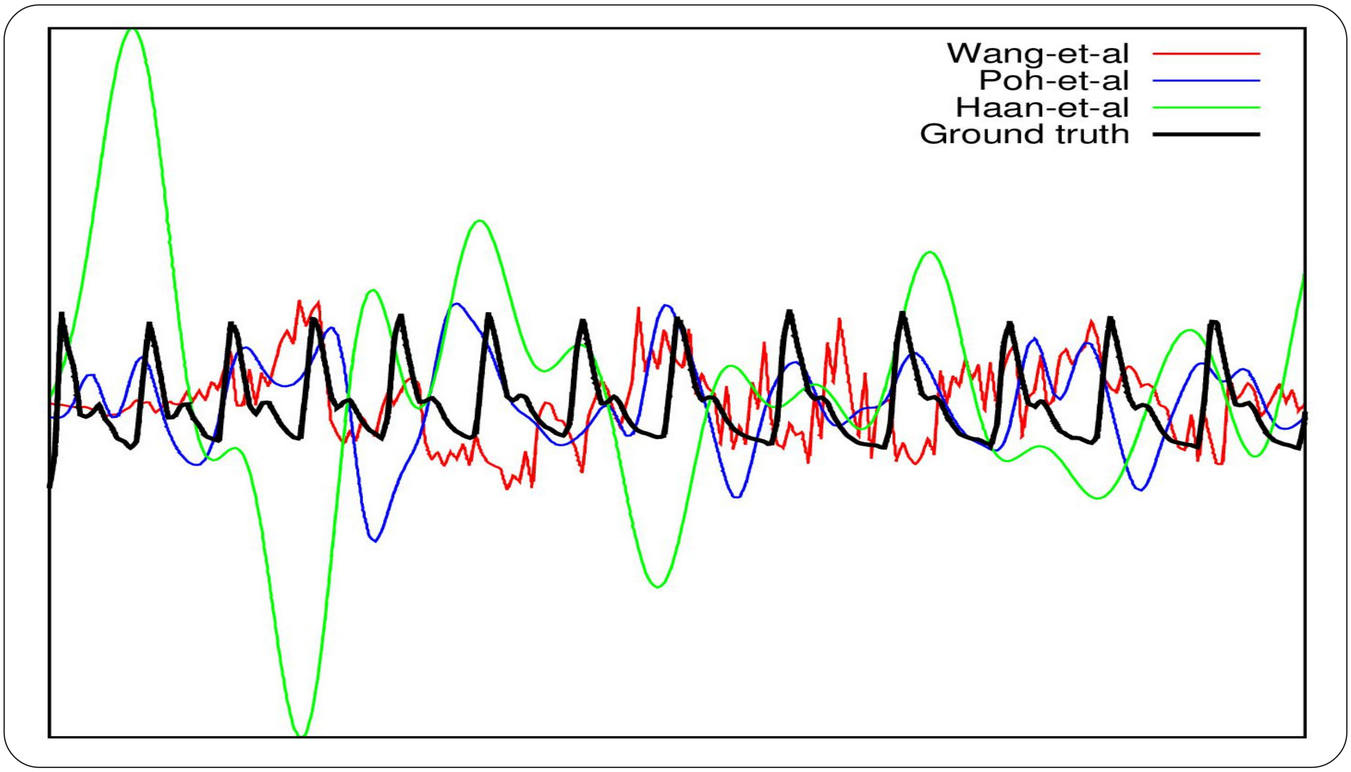

Figure 3. First 10 seconds of the color-based methods

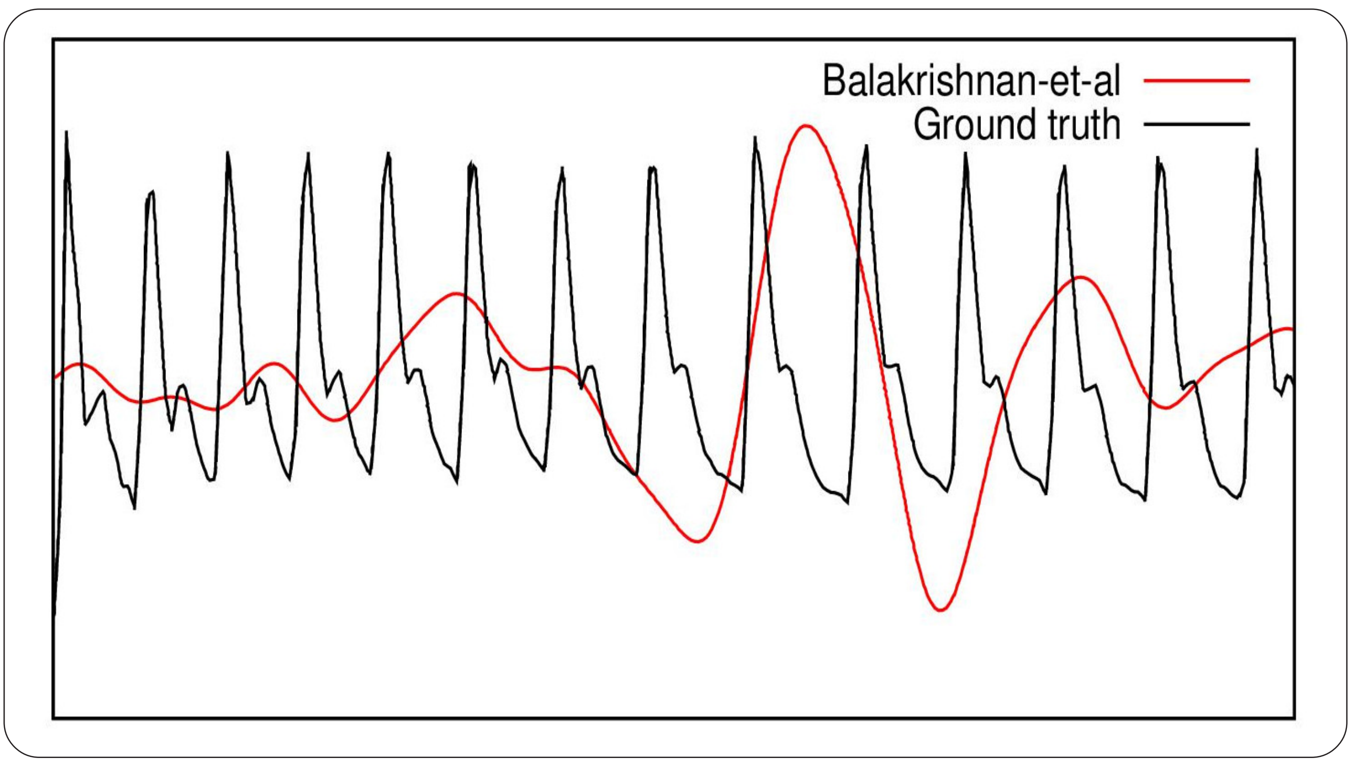

Figure 4. First 10 seconds of the motion-based method

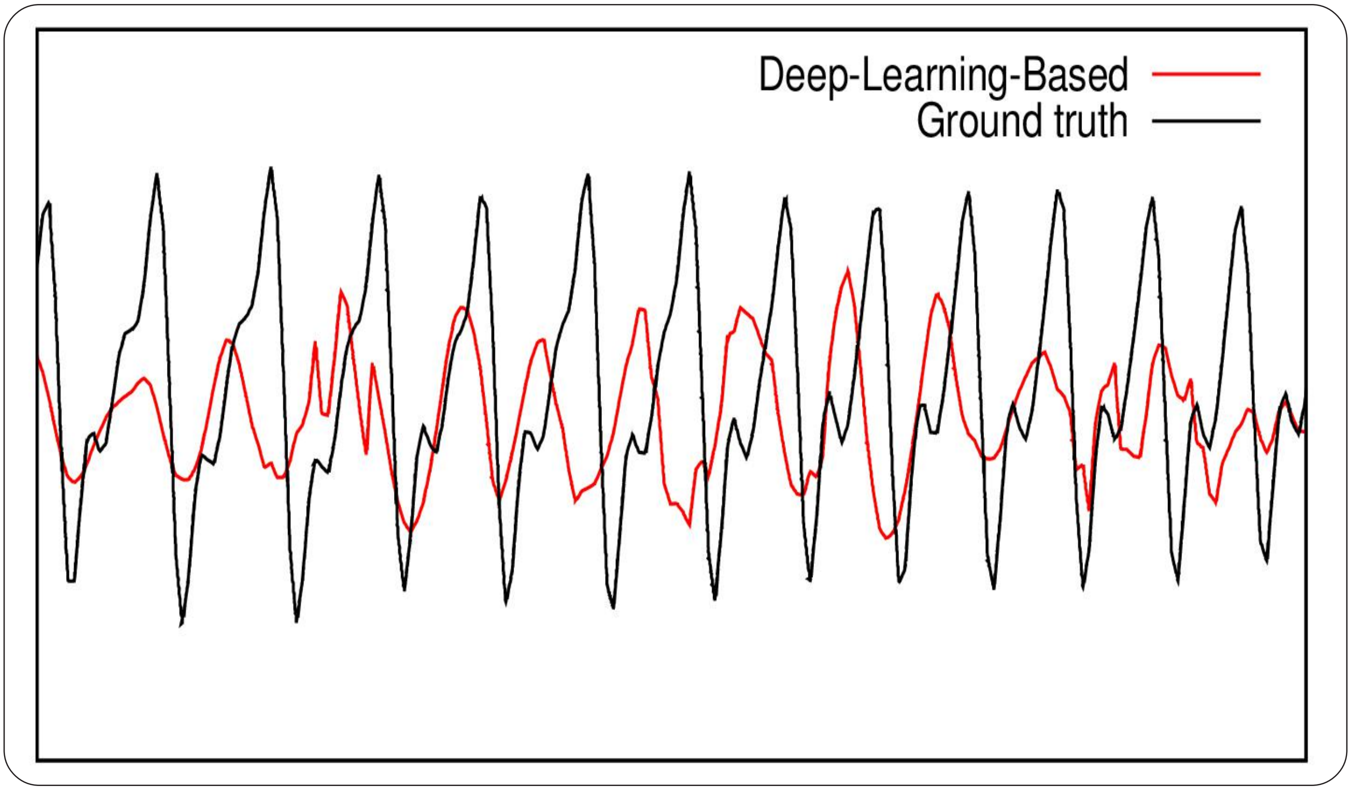

Figure 5. First 10 seconds of the Deep-Learning-Based method

Acknowledgments

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 780819.

References

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

[2] de Haan, G., Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering*, 60 (10) 2878-2886, (October).

[3] Heusch, G., Anjos, A., Marcel, S. (2017). A reproducible study on remote heart rate measurement. CoRR, abs/1709.00962.

[4] Hyvarinen, A. (1999). Fast ICA for noisy data using Gaussian moments. *In*: ISCAS'99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems VLSI, volume 5, p. 57-61, (May).

[5] Lewandowska, M., Rumi–nski, J., Kocejko, T., Nowak, J. (2011). 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), p. 405-410 (September) .

[6] Lucas, B. D., Kanade, T. An iterative image registration technique with an application to stereo vision. *In*: Proceedings of the 7th International Joint Conference on Articial Intelligence - Volume 2, IJCAI'81, p. 674-679, San Francisco, CA, USA, 1981. *Morgan Kaufmann Publishers Inc*.

[7] Patil, O. R., Gao, Y., Li, B., Jin, Z. (2017). 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, p. 524-529, New York, NY, USA, 2017. ACM.

[8] Patterson, J., Gibson, A. (2017). Deep Learning: A Practitioner's Approach. O'Reilly, Sebastopol.

[9] Poh, M., McDu, D. J., Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58 (1) 7-11 (January).

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

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

[12] Wang, W., Stuijk, S., de Haan, G. (2016). A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Transactions on Biomedical Engineering*, 63 (9), 1974-1984, (September).

[13] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W. (2012). Eulerian video magnication for revealing subtle changes in the world. *ACM Trans. Graph*., 31 (4), 65, 1-65, 8, (July).