Towards an Optimal Remote Photo-Plethysmographic Framework using Weighted Ensemble Models

Swapnil Sayan Saha, Suparno Pal, Rishav Guha and Paawan Garg
University of California - Los Angeles
{swapnilsayan, suparnopal, rishavgh97, pwngarg27}@g.ucla.edu

Abstract

Remote photo-plethysmography (rPPG) is a non-contact means of blood volume pulse (BVP) detection through light transmittance variations from the skin, measured using consumer-level cameras. In this paper, we propose a complementary approach of fusing rPPG algorithms using weights extracted via linear optimization on the residuals and absolute errors on two standard rPPG video datasets. Three state-of-the-art (SOTA) rPPG techniques, namely CHROM, POS and ICA, were fused and observed to reduce average root-squared and absolute deviations by 24% and 18% respectively over individual models and vanilla fusion techniques, achieving an upper quartile of $\pm$ 5.0 bpm.

1. Introduction

Photo-plethysmographic cardiac pulse measurement allows for an inexpensive, non-invasive, non-abrasive and comfortable means of measuring heart rate and associated cardiovascular parameters using commodity imaging systems [17] void of specialized hardware. Changes in blood volume alters light transmittance and reflectance characteristics of the skin, inducing subtle detectable color variations [21] [23]. Applications of rPPG include non-contact healthcare monitoring (heart rate (HR), respiratory rate, blood pressure, pulse oxymetry, arterial assessment and pulsatility), tele-medicine, fitness training, security (e.g. face anti-spoofing), sleep and wakefullness analysis and psychological assessment [21][23][25][2]. Several notable analytical rPPG techniques have been proposed in literature over the recent years, namely GREEN [21], ICA [17], CHROM [4], BCG (also called PCA) [1] and POS [23], with recent neural architectural advancements in deep-learning catalyzing a rise in learning-enabled rPPG frameworks [19][18][16].

In this paper, we illustrate a near-optimal weighted rPPG algorithmic fusion framework inspired from the complementary and Kalman filter sensor fusion techniques [5][8], aimed to mitigate the weaknesses of individual methods and amplify their strengths by taking the HR error variances into account. Specifically, given individual HR estimation errors, a linear optimization program can yield optimal weights for fusing several rPPG algorithms, reducing both the mean and variance of deviations over naive fusion or solo models. We parametrize the performance of five SOTA analytical rPPG algorithms mentioned in previous paragraph on two standard rPPG video datasets totaling upto 60 minutes of video data, in terms of window size, color distortion filtering and skin segmentation, and fuse three of them to yield a stable and accurate ensemble of rPPG BVP detectors. In addition, we provide intuition of how camera and ambient lighting parameters may affect HR estimation. The code for the paper (implemented in MATLAB) is available at: `https://git.io/JItjr`.

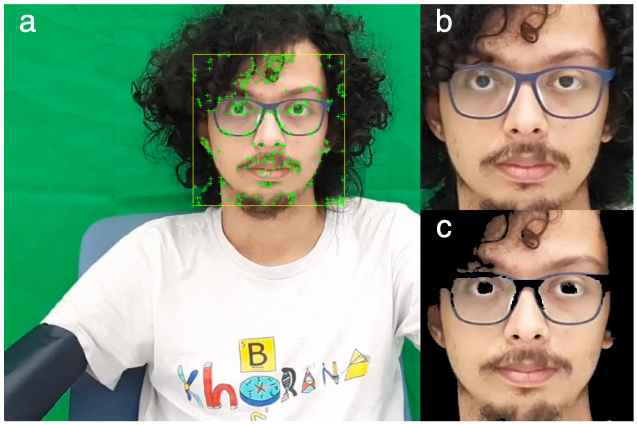

Figure 1. Extracting ROI from video frames for rPPG (a) Using Viola-Jones cascade for face detection and KLT for tracking facial features (characterized by green points) (b) ROI without skin segmentation (c) ROI after applying RGB skin segmentation.

2. Technical Approach

The first step in remote HR detection is to track faces and extract measurement regions of interest (ROI) in the

video frames [17], supplemented via algorithm-specific post-processing such as spatial averaging over pre-defined time windows [21], color distortion filtering [23] or skin tissue segmentation [3]. The extracted RGB traces (whose intensities are time-varying) are then analyzed in the frequency domain with the goal of obtaining HR-specific harmonics.

2.1. Face Detection and Tracking

For robust face detection in the first frame, we use the widely-applicable Viola-Jones object detection framework, exploiting Haar-like regularities in frontal upright facial images in the form of an increasingly complex binary classifier cascade trained using AdaBoost [22]. In subsequent frames, we use Kanade-Lucas-Tomasi (KLT) feature tracker [12][20] to track salient features [7] using spatial intensities extracted from the cardinal frame. Features are selected if the eigenvalues of weighted image gradient matrix exceed a pre-defined threshold, while tracking is formulated as a local optimization problem with respect to transformation variables. The resulting framework, shown in Figure 1(a) and 1(b), is computationally inexpensive and robust to facial and photometric transformations, noise and clutter.

2.2. Skin Segmentation

For GREEN and ICA, we implemented a threshold-based RGB skin tissue segmentation to remove unwanted artifacts, shown in Figure 1(c), which is given as:

$$\mathbf{I}_s(x, y, c_k) = \mathbf{I}(x, y, c_k) \odot \mathbf{M}(x, y) \tag{1}$$

where, I = ROI, $k \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ and M is given by:

$$\mathbf{M}(i, j) = (\phi \rightarrow 1) \wedge (\neg\phi \rightarrow 0) \tag{2}$$

$$\phi = [(\mathbf{R}_{i,j} > \alpha) \wedge (\mathbf{G}_{i,j} > \beta) \wedge (\mathbf{B}_{i,j} > \gamma) \wedge (\mathbf{R}_{i,j} > \mathbf{G}_{i,j}) \wedge (\mathbf{R}_{i,j} > \mathbf{B}_{i,j}) \wedge (|\mathbf{R}_{i,j} - \mathbf{G}_{i,j}| > \kappa) \wedge (\max(\mathbf{R}_{i,j}, \mathbf{G}_{i,j}, \mathbf{B}_{i,j}) - \min(\mathbf{R}_{i,j}, \mathbf{G}_{i,j}, \mathbf{B}_{i,j}) > \delta)] \tag{3}$$

where, $\alpha, \beta, \gamma, \kappa$ and δ are thresholds set empirically.

2.3. Processing the Green Channel (GREEN)

Verkruysse et al. [21] first proposed the use of commodity cameras for rPPG under normal environmental conditions. A series of RGB pixel values $PV(x, y, t)$ from t ROI video frames are spatially averaged either directly or projected onto a coarse grid to improve signal-to-noise ratio (SNR), yielding $PV_{raw}(x, y, t)$. Fast Fourier transform (FFT) on $PV_{raw}(x, y, t)$ generates phase and power spectra maps, which can be used to find signals with HR periodicity (0.6 to 4 Hz), most pronounced in the green channel. The algorithm is not robust to movement artifacts, photon shot noise or varying lighting conditions, and is often complemented with Butterworth filtering or static DC bias removal to remove unwanted frequency artifacts.

2.4. Independent Component Analysis (ICA)

To improve the robustness of GREEN against noise and motion artifacts that fall within HR bands, Poh et al. [17] illustrated blind source separation (BSS) using ICA (assuming linear and independent signal mixture for short time windows) to uncover uncorrupted PPG signals. If $\mathbf{s} = [c_R(t), c_G(t), c_B(t)]$ is the ideal PPG signal and $\mathbf{x}$ is the observed image:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{4}$$

The goal is to find a demixing matrix $\mathbf{W}$ ($\sim \mathbf{A}^{-1}$) that maximizes non-Gaussianity of $\mathbf{s}$ such that:

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \tag{5}$$

The joint approximate diagonalization of eigenmatrices (JADE) is used to find fourth-order cumulants of $\mathbf{x}$, and a cost function is used to find an approximate value of $\mathbf{W}$ via joint diagonalization of cumulant matrices. The rest of the algorithm follows GREEN. The assumptions do not hold for longer windows (due to Beer-Lambert law and specular reflections) and not robust to pronounced head jerks.

2.5. Chrominance-Based PPG (CHROM)

BSS assumes strong periodicity of HR signals, which can break down for periodic head movements as well as specular reflections under eclectic lighting conditions. de Haan et al. [4] modeled the reflected light from skin on camera sensor as:

$$C_i = I_{C_i}(\rho C_{dc} + \rho C_i + s_i) \tag{6}$$

where, $C \in \{R, G, B\}$, I_{C_i} = pixel intensity integrated camera exposure time in image i, ρC_{dc} = DC reflection coefficient, ρC_i = AC reflection coefficient due to BVP, and s_i = additive specular reflection. To eliminate dependence of C_i on color and intensity of ambient light, C_i is normalized over the running average of C_i over a temporal interval:

$$C_{ni} = \frac{C_i}{\mu(C_i)} \tag{7}$$

The specular reflection component can be eliminated via signal chrominance, and the ratio of two orthogonal chrominance components yields ideal rPPG signal:

$$S = \frac{X_n}{Y_n} - 1 = \frac{R_n - G_n}{0.5R_n + 0.5G_n - B_n} - 1 \tag{8}$$

Equation (8) is complemented with skin-tone standardization by dividing the color channels by their mean and multiplying them [R_n,G_n,B_n] with [0.7682, 0.5121, 0.3841]. The resulting model of S, when approximated using the logarithmic Taylor series, yields $S \approx X_s - Y_s$, which follows equation (5) from ICA closely but without requiring additional heuristics. To account for small BVP, the model is written as:

$$S = X_f - \frac{\sigma(X_f)}{\sigma(Y_f)} Y_f \tag{9}$$

2.6. Plane-Orthogonal-to-Skin (POS)

Algorithm 1: Plane-Orthogonal-to-Skin (POS)

input : Video sequence of N frames
Initialize: $\mathbf{H}$ = **zeros**(1,N), l (depends on camera)
for $n = 1, 2, ..., N$ **do**
  $\mathbf{C}(n) = [R(n), G(n), B(n)]^T \leftarrow$ spatial avg.
  if $m = n - l + 1 > 0$ **then**
    $\mathbf{C_n^i} = \frac{\mathbf{C^i_{m \to n}}}{\mu(\mathbf{C^i_{m \to n}})} \leftarrow$ temp. norm.
    $\mathbf{S} = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix} \cdot \mathbf{C_n} \leftarrow$ projection
    $\mathbf{h} = \mathbf{S}_1 + \frac{\sigma(\mathbf{S}_1)}{\sigma(\mathbf{S}_2)}\mathbf{S}_2 \leftarrow$ alpha tuning
    $\mathbf{H_{m \to n}} = \mathbf{H_{m \to n}} + (\mathbf{h} - \mu(\mathbf{h})) \leftarrow$ ovlp., add
  end
end
output: rPPG signal $\mathbf{H}$

While mostly similar to CHROM, POS [23] focuses on hue-change rather than intensity and alters the order of distortion compensation. CHROM compensates specular distortions first and uses intensities for alpha tuning (equation (9)) while POS does the opposite. POS softens CHROM's standardized skin-tone prior by using a projection plane normal to the normalized momentary skin-tone direction, where as CHROM's projection plane is orthogonal to the specular variation direction. While CHROM is not robust to subject-dependent changes in specular distortions, POS is not robust heterogeneous illumination spectra, having complementary weaknesses.

2.7. Principal Component Analysis (BCG / PCA)

Balakrishnan [1] proposed another BSS technique for HR extraction through PCA applied on head movements instead of color channels, often referred to as ballistocardiography. The cyclical movement of blood from the heart to the head causes the head to move periodically. BCG extracts head pose using feature tracking and uses PCA to isolate the motion due to BVP, which is then projected onto a 1-D signal to extract individual beat boundaries from trajectory peaks. PCA selects the component whose temporal power spectrum is strongly correlated to typical BVP.

2.8. The Optimization Program

Consider n rPPG frameworks (each denoted by S), each making independent inferences on N videos, with the goal of reporting an average heart rate over the entire video while the ground truth heart rate being H. The weights w for the optimal fusion of these models can be found via the following optimization program:

$$\min_{w} \frac{1}{N}\sum_{j=1}^{N}|(H_j - \sum_{i=1}^{n} w_i S_{i,j})| + \sum_{j=1}^{N}\sqrt{\frac{(H_j - \sum_{i=1}^{n} w_i S_{i,j})^2}{N}}$$
$$\text{s.t.} \quad \sum_{i=1}^{n} w_i = 1, \quad 0 \leq w_i \leq 1 \tag{10}$$

The first part of the equation models the mean absolute deviation (MAE) (accuracy) while the second part models the root mean-squared error (RMSE) (stability) of ground truth heart rate versus estimated heart rate, resulting in weights that can minimize the sum of both errors.

3. Evaluation

We evaluated the algorithms on two SOTA rPPG video datasets, benchmarking the MAE and RMSE in terms of window length, skin segmentation and color distortion filtering. Afterwards, we fused CHROM, POS and ICA with weights obtained via equation (10) and compared its performance against solo and naive fusion algorithms. We reused and modified code for the 5 algorithms from the iPhys Physiological Measurement Toolbox [14]. The UBFC-RPPG dataset [3] contains 1-minute 30-FPS clips from 50 subjects (90,000 uncompressed video frames, 50 minutes) captured under varying lighting conditions using a commonplace webcam, with a transmissive pulse oximeter offering ground truth HR. The Ostankovich-Prathap-Afanasyev (OPA) dataset [15] contains 20-second 25-FPS clips from 15 subjects (15,000 compressed video frames, 10 minutes), with each subject offering BVP under normal conditions and after physical activity. The ground truth HR was calculated from ECG data using a portable USB electrocardiograph.

3.1. Results

Figure 2 (top) shows the lowest possible MAE and RMSE (in BPM, with respect to ground truth HR) obtained by the 5 algorithms on OPA dataset. We notice that BCG performs poorly compared to other algorithms, and hypothesize that this is due to its inability to properly filter out head motion caused by BVP from physical movements. CHROM and POS provide lowest MAE (highest accuracy) but are less stable than ICA in terms of error spread. Based on individual performance, we fused CHROM, POS and ICA, with [0.1478, 0.0687, 0.7834] as weights obtained from (10) and benchmarked the performance of chosen algorithms on both datasets. From Figure 2 (bottom), we see that our fusion framework, on average, reduces MAE and RMSE by 18% and 24% respectively, achieving optimal balance between accuracy (5.45 bpm) and stability (9.9

bpm). We also observe that our fusion framework successfully amplifies strengths of each algorithm, coupling stability of ICA with accuracy of CHROM and POS. In addition, we plotted the histogram of residuals and noticed 58% of estimated values lie within $\pm$ 2.5 bpm, while 75% of estimated values lie within $\pm$ 5 bpm, highest among all models. Figure 3(a) shows the effects of window sizes on MAE

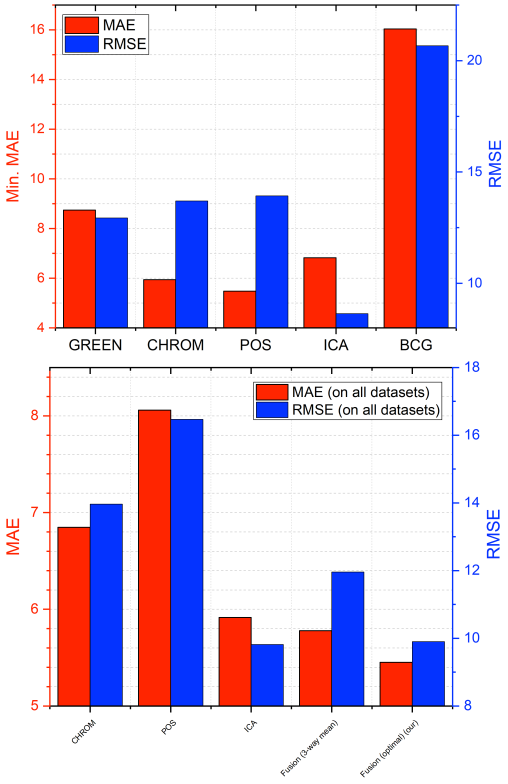

Figure 2. (Top) Lowest MAE and RMSE achieved by algorithms on OPA dataset. (Bottom) MAE and RMSE achieved by solo and fusion frameworks on OPA and UBFC-RPPG dataset.

of solo algorithms. The MAE of GREEN increases with smaller windows as GREEN requires a large number of frames to achieve a high SNR during spatial averaging. In contrary, for CHROM, POS and ICA, a larger time window breaks linear independence assumption for signal mixture and causes performance degradation. In our final framework, we chose 2 second windows for all three algorithms. Figure 3(b) illustrates the effects of skin segmentation on MAE. CHROM and POS came prepackaged with their own YbCbCr skin segmentation models, while we applied our RGB skin segmentation on GREEN and ICA. We observe that GREEN does not respond to the use of skin segmentation, supposedly due to skin-agnostic formulation as well as loss in SNR due to partially imperfect skin rendering. In addition, we tested the optional color distortion filtering [24] that came with POS but noticed that it increases MAE.

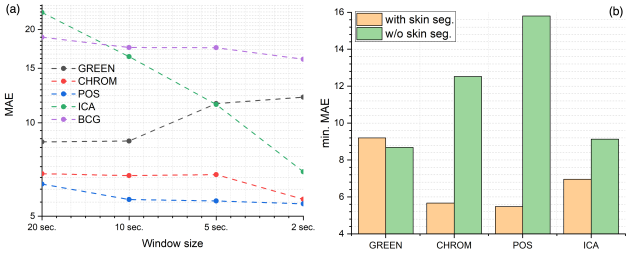

Figure 3. (a) MAE of algorithms with varying window sizes (b) Effect of skin segmentation on MAE of algorithms (BCG does not require it because it operates on head motion.)

3.2. Effects of camera control (comp. imaging)

Laurie et al. [10] proposed dedicated exposure control for rPPG frameworks to improve SNR by evaluating the maximum exposure time possible without saturating the ROI. As shutter speed decreases, exposure time increases, causing high-value pixels to get saturated. This appears as soft-clipping distortion in the rPPG signal in frames sampling close to rPPG peaks. On the other hand, since ISO (gain) is applied directly to analog sensor output before quantization, increasing ISO improves signal-to-quantization-noise ratio. However, random noise is also amplified with the signal. Laurie et al. showed that increasing ISO works until a critical point, after which noise power starts to dominate, and suggested shutter speed as the most appropriate control for rPPG. Ma et al. [13] mentioned that a small aperture camera adds random noise to rPPG signals under low-light, while a large aperture may cause pixel saturation [10]. Increasing the aperture size also induces Bokeh effect, which is beneficial for focusing specifically on ROI and blur out unwanted artifacts [10], reducing their tendency to induce unwanted artifacts in HR bands. Lin et al. [11] showed that light green filter or illumination improves SNR of rPPG green channel by $2\times$ due to high absorptivity of oxyhaemoglobin at $\sim$ 550 nm. Fukunishi et al. [6] demonstrated that specular distortion of rPPG signal caused by surface reflection can be removed with a polarizing filter. Kipge et al. [9] illustrated that HDR imaging coupled with high FPS increases temporal resolution of BVP. We observed that camera position affects MAE depending on useful ROI availability (front better than bottom).

4. Conclusion

In this paper, we provided a complementary technique for near-optimal weighted fusion of multiple rPPG models using linear optimization on residuals and absolute deviation. We fuse three classical rPPG frameworks and observe 18-24% improvement in MAE and RMSE over solo and simple fusion frameworks, with 75% estimates lying within $\pm$ 5.0 bpm (Figure 4).

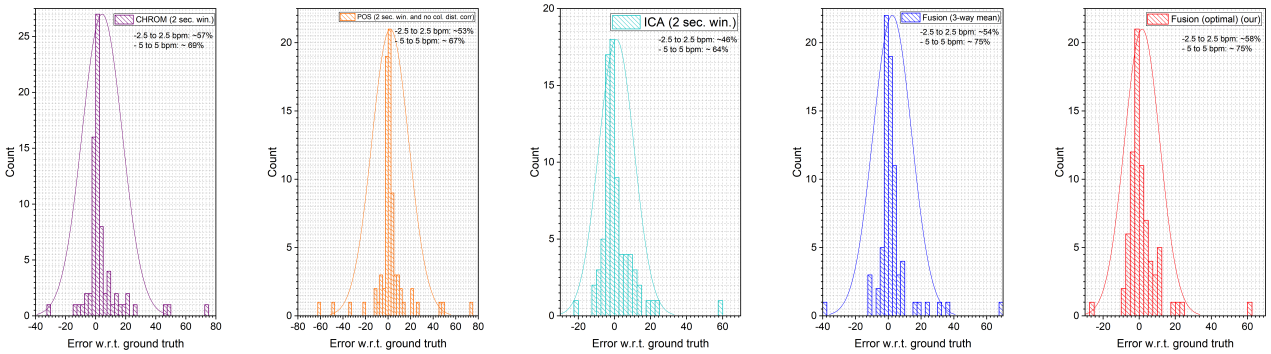

Figure 4. Histogram plots for CHROM, POS, ICA, naive fusion and our fusion on both datasets.

References

[1] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[2] Simone Benedetto, Christian Caldato, Darren C Greenwood, Nicola Bartoli, Virginia Pensabene, and Paolo Actis. Remote heart rate monitoring-assessment of the facereader rppg by noldus. *PloS one*, 14(11):e0225592, 2019.

[3] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82 – 90, 2019. Award Winning Papers from the 23rd International Conference on Pattern Recognition (ICPR).

[4] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[5] M. Euston, P. Coote, R. Mahony, J. Kim, and T. Hamel. A complementary filter for attitude estimation of a fixed-wing uav. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 340–345, 2008.

[6] Munenori Fukunishi, Kouki Kurita, Shoji Yamamoto, and Norimichi Tsumura. Video based measurement of heart rate and heart rate variability spectrogram from estimated hemoglobin information. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[7] Jianbo Shi and Carlo Tomasi. Good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[8] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.

[9] E. Kviesis-Kipge and U. Rubīns. Portable remote photoplethysmography device for monitoring of blood volume changes with high temporal resolution. In *2016 15th Biennial Baltic Electronics Conference (BEC)*, pages 55–58, 2016.

[10] J. Laurie, N. Higgins, T. Peynot, and J. Roberts. Dedicated exposure control for remote photoplethysmography. *IEEE Access*, 8:116642–116652, 2020.

[11] Yu-Chen Lin and Yuan-Hsiang. Lin. A study of color illumination effect on the snr of rppg signals. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4301–4304, 2017.

[12] B Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh Int'l Joint Conf. Artificial Intelligence*, pages 674–679, 1981.

[13] Xiaocong Ma, Diana P. Tobón, and Abdulmotaleb El Saddik. Remote photoplethysmography (rppg) for contactless heart rate monitoring using a single monochrome and color camera. In Troy McDaniel, Stefano Berretti, Igor D. D. Curcio, and Anup Basu, editors, *Smart Multimedia*, pages 248–262, Cham, 2020. Springer International Publishing.

[14] D. McDuff and E. Blackford. iphys: An open non-contact imaging-based physiological measurement toolbox. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6521–6524, 2019.

[15] V. Ostankovich, G. Prathap, and I. Afanasyev. Towards human pulse rate estimation from face video: Automatic component selection and comparison of blind source separation methods. In *2018 International Conference on Intelligent Systems (IS)*, pages 183–189, 2018.

[16] Olga Perepelkina, Mikhail Artemyev, Marina Churikova, and Mikhail Grinenko. Hearttrack: Convolutional neural network for remote video-based heart rate monitoring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[17] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010.

[18] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen. Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 69(10):7411–7421, 2020.

[19] Chuanxiang Tang, Jiwu Lu, and Jie Liu. Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[20] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, Carnegie Mellon University, CMU-CS-91-132, 1991.

[21] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, Dec 2008.

[22] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001.

[23] 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.

[24] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Color-distortion filtering for remote photoplethysmography. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 71–78, 2017.

[25] Changchen Zhao, Chun-Liang Lin, Weihai Chen, and Zhengguo Li. A novel framework for remote photoplethysmography pulse extraction on compressed videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.