MMPD: Multi-Domain Mobile Video Physiology Dataset

Jiankai Tang[1], Kequan Chen[1], Yuntao Wang[1*], Yuanchun Shi[1], Shwetak Patel[2], Daniel McDuff[2], Xin Liu[2]
* Corresponding Author

***Abstract*— Remote photoplethysmography (rPPG) is an attractive method for noninvasive, convenient and concomitant measurement of physiological vital signals. Public benchmark datasets have served a valuable role in the development of this technology and improvements in accuracy over recent years. However, there remain gaps in the public datasets. First, despite the ubiquity of cameras on mobile devices, there are few datasets recorded specifically with mobile phone cameras. Second, most datasets are relatively small and therefore are limited in diversity, both in appearance (e.g., skin tone), behaviors (e.g., motion) and environment (e.g., lighting conditions). In an effort to help the field advance, we present the Multi-domain Mobile Video Physiology Dataset (MMPD), comprising 11 hours of recordings from mobile phones of 33 subjects. The dataset is designed to capture videos with greater representation across skin tone, body motion, and lighting conditions. MMPD is comprehensive with eight descriptive labels and can be used in conjunction with the rPPG-toolbox [1]. The reliability of the dataset is verified by mainstream unsupervised methods and neural methods. The GitHub repository of our dataset: https://github.com/THU-CS-PI/MMPD_rPPG_dataset.**

I. INTRODUCTION

Remote photoplethysmography (rPPG) is an optical technique for measuring the cardiac pulse, or photoplethysmograph (PPG), via subtle changes in light reflected from the skin [2]. Unobtrusive measurement of vital signs, such as heart rate, is a crucial technology for remote health monitoring and could be particularly useful for screening, and monitoring. individuals with chronic cardiovascular diseases. However, the high cost and complicated operation of traditional medical devices make regular measurements infeasible. While rPPG offers many benefits, the performance of existing video-based measurement is often brittle and can be sensitive to changes in i) appearance (e.g., skin tone), ii) the environment (e.g., lighting) and iii) activities (e.g., types of body motion). Research has shown that it is harder to extract pulse signals from individuals with darker skin tones due to the lower signal-to-noise ratio in the reflected light [3]. Changes in lighting can significantly alter the appearance of a person's face and make it harder to detect subtle changes in reflectance due to blood flow [4]. Dim or bright lighting can also lead to under or overexposure and create unwanted specular reflections, which can further obscure the signal. Motion artifacts in videos present severe challenges, and current state-of-the-art models struggle to generate precise pulse waveforms and heart rates when people are moving. Most models have not been extensively tested when users engage in naturalistic activities, such as talking or walking [4], [5], [6].

Public benchmark datasets are an extremely valuable resource to the scientific community; however, all datasets are finite. In the case of rPPG, existing datasets do not contain examples that allow researchers to systematically test models across all the aforementioned dimensions (appearance, environment and activity). For example, the widely used UBFC-rPPG [7] dataset primarily includes videos of stationary subjects with Fitzpatrick skin types 2-3. The PURE [8] dataset includes head motions that are relatively unnatural and it was also collected primarily from subjects with Fitzpatrick skin types 2-3. Finally, many of the existing public rPPG datasets were recorded using digital single-lens reflex (DSLR) cameras or devices from specialist imaging companies. This is in contrast with the most ubiquitous camera types, namely smartphone cameras.

To address gaps in existing public rPPG datasets, we introduce the multi-domain mobile video physiology dataset (MMPD). Our dataset includes 33 subjects with Fitzpatrick skin types 3-6, four different lighting conditions (LED-high, LED-low, incandescent, natural), and four different activities (stationary, head rotation, talking, and walking). All videos in MMPD are captured using mobile phones. Our paper presents the following contributions: 1) we introduce the MMPD dataset, the first public dataset that includes subjects with diverse skin types (Fitzpatrick scale of 3-6), different lighting conditions, and various real-world motion scenarios. 2) we conduct a comprehensive quantitative analysis to evaluate the performance of existing state-of-the-art neural and unsupervised signal processing methods on our dataset. Our goal is to provide researchers with a dataset that enables the development of algorithms that can handle complex and realistic scenarios, as well as address bias in camera-based physiological measurements.

II. RELATED WORKS

There are a number of commonly used rPPG datasets (PURE [8], MAHNOB-HCI [9], BP4D [10], VIPL-HR [11], COHFACE [12], UBFC-RPPG [7], MR-NIRP [13], VicarPPG-2/CleanerPPG [14], Scamps [15]). Some of these datasets were collected with the explicit purposes of rPPG in mind, while others were collected for generic physiological and computer vision research. From these remarkable datasets, we picked the three most frequently used datasets for analysis and comparison.

[1] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi are with the Tsinghua University, China yuntaowang@tsinghua.edu.cn.
[2] Xin Liu, Shwetak Patel, Daniel McDuff are with the University of Washington, Seattle, WA, USA {xliu0, shwetak, dmcduff}@cs.washington.edu.

TABLE I

DATASET COMPARISON

Dataset	Frames	Subjects	Camera	Sensor	Skin Tone	Motion	Lighting	Exercise
UBFC	57,420	42	Logitech C920 HD Pro	CMS50E	✗	✗	✓	✗
PURE	168,120	10	eco274CVGE	CMS50E	✗	✓	✗	✗
Scamps*	1,296,000	2800	/	/	✓	✓	✓	✗
MMPD	1,188,000	33	Galaxy S22 Ultra	HKG-07C+	✓	✓	✓	✓

As different datasets contain videos with different durations, size was computed here in terms of the number of video frames. *Scamps is a synthetic dataset and therefore is not directly comparable to other datasets.

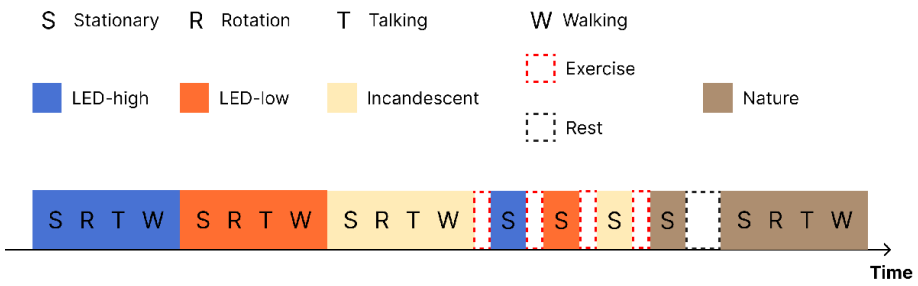

Fig. 1. A visual illustration of our data collection protocol. Video recordings of each participant were collected under different lighting configurations (blue), activities (S, R, T, W) and before and after exercise (red vs. black box).

UBFC [7]. The UBFC-RPPG dataset is captured using a Logitech C920 HD Pro webcam with a resolution of 640x480 at a frame rate of 30fps. Ground truth PPG data, including the PPG waveform and heart rate, is obtained using a CMS50E transmissive pulse oximeter. The subjects are seated approximately 1 meter away from the camera with their face visible, and the experiments are conducted indoors with varying levels of sunlight and indoor lighting. While the dataset is reliable and widely used as a baseline, it has limited diversity due to the range of skin tones and motions represented.

PURE [8]. The PURE database includes 60 one-minute sequences captured using an eco274CVGE camera at 30fps with a resolution of 640x480 pixels, and the PPG data is acquired in parallel by a CMS50E transmissive pulse oximeter at a sampling rate of 60Hz. The dataset is widely used due to its diversity of motions, including talking, translation, and head rotation, but it lacks variety in skin tones, real-world motion tasks, and lighting conditions.

Scamps [15]. The Scamps dataset provides frame-level ground-truth labels for PPG, inter-beat intervals, breathing waveforms, breathing intervals, and 10 facial actions in 2,800 video sequences. Each video is rendered using the corresponding waveforms, action unit intensities, and randomly sampled appearance properties. Although the dataset has demonstrated its potential for various applications, models trained on SCAMPS tend to have poor performance due to overfitting because of the simplistic nature of the vitals.

III. DATASET

In an effort to create a dataset that captures some of the diversity and complexity of videos seen in real-world applications, we recruited subjects from different countries and conducted experiments under various lighting configurations. A total of 660 one-minute videos were recorded using a Samsung Galaxy S22 Ultra, while gold-standard PPG signals were simultaneously recorded using an HKG-07C+ oximeter. In this section, we will describe the data collection protocol, data processing techniques and dataset organization.

A. Data Collection

As previously noted, lighting and motion can greatly affect the extraction of PPG signals from videos. To further study these factors, we designed an experiment that simultaneously collects face videos and finger PPG signals. The experimental procedure is illustrated in Figure 1, and all the videos were captured at a distance that allows for touch.

The experiment involved four levels of light intensity and three types of light sources, including low LED light (100 lumens on the face region), mid-level incandescent light (200 lumens on the face region), high LED light (300 lumens on the face region), and natural light (varying from 300-800 lumens intensity on the face region). For motion, we designed four tasks of varying difficulty, including remaining stationary while staring at a screen, head rotation, talking while keeping the head stationary, and taking a selfie video while holding the phone. In addition, we conducted four exercises to investigate the impact of physical activity on stationary scenarios. Subjects were asked to perform high knee lifts or other strenuous exercises to raise their post-exercise heart rate before recordings. After all exercises, subjects would take enough breaks to calm down before taking the next experiment.

TABLE II

THE RESULTS OF UNSUPERVISED SIGNAL PROCESSING METHODS.

Method	ICA [16]				POS [17]				CHROM [18]			
	MAE↓	RMSE↓	MAPE↓	ρ ↑	MAE↓	RMSE↓	MAPE ↓	ρ ↑	MAE↓	RMSE↓	MAPE ↓	ρ ↑
Skin tone												
3	8.83	12.24	12.15	0.26	**5.76**	**9.67**	**8.63**	**0.48**	6.57	10.46	9.64	0.33
4	15.16	19.81	17.60	0.12	9.06	13.51	10.37	0.23	10.57	13.80	12.69	0.15
5	14.42	17.70	20.07	-0.10	12.78	16.69	19.29	-0.03	14.65	18.91	22.29	-0.12
6	17.14	21.52	19.77	-0.01	11.17	15.34	13.63	0.26	12.53	16.47	14.94	0.06
Motion												
Stationary	11.48	15.82	15.06	0.16	9.70	13.74	13.35	0.26	10.23	14.23	14.46	0.15
Rotation	11.75	15.88	15.35	0.06	7.50	11.99	10.85	0.40	9.28	14.02	13.10	0.16
Talking	13.14	17.18	16.50	0.20	8.05	12.60	10.87	0.30	9.31	13.51	12.26	0.25
Walking	26.15	30.75	27.43	-0.08	17.05	21.20	18.49	-0.06	17.61	21.07	19.15	-0.12
Light												
LED-low	12.20	16.54	15.71	0.03	9.76	14.15	13.41	0.14	10.49	14.84	14.52	0.08
LED-high	11.98	15.90	15.41	0.21	7.26	11.26	10.24	0.45	9.53	13.42	13.25	0.15
Incandescent	12.20	16.48	15.80	0.16	8.24	12.81	11.41	0.35	8.80	13.46	12.06	0.29
Nature	17.21	21.42	19.84	0.19	10.71	14.21	13.20	0.36	12.88	17.04	15.46	0.12

Method	GREEN [19]				LGI [20]				PBV [21]			
	MAE↓	RMSE↓	MAPE↓	ρ ↑	MAE↓	RMSE↓	MAPE ↓	ρ ↑	MAE↓	RMSE↓	MAPE ↓	ρ ↑
Skin tone												
3	12.37	16.48	16.67	0.15	5.99	9.83	8.10	0.45	7.94	11.36	10.96	0.38
4	23.39	26.27	27.72	0.10	14.43	19.91	16.17	-0.17	15.87	20.50	18.34	-0.01
5	15.22	18.89	20.51	0.17	14.23	18.17	19.84	-0.02	14.62	17.77	20.17	0.08
6	20.59	24.96	23.37	0.13	17.02	22.15	19.28	0.03	17.24	21.38	19.81	0.10
Motion												
Stationary	13.33	18.41	16.97	0.16	10.80	15.99	13.61	0.01	10.80	14.40	14.08	0.28
Rotation	16.67	20.48	21.33	0.07	9.38	14.87	12.26	0.15	11.18	16.21	14.64	0.08
Talking	17.16	21.14	21.48	0.06	11.36	16.26	13.87	0.20	13.44	17.43	16.69	0.19
Walking	29.81	34.41	31.56	0.06	25.48	30.24	26.82	0.04	25.66	30.19	27.09	0.08
Light												
LED-low	17.12	21.43	21.75	0.17	11.59	16.43	14.53	-0.01	11.91	16.09	15.16	0.11
LED-high	14.71	19.13	18.76	0.08	9.74	14.91	12.46	0.17	13.01	17.22	16.95	0.09
Incandescent	15.33	19.49	19.27	0.06	10.22	15.78	12.75	0.19	10.49	14.76	13.31	0.33
Nature	20.07	24.74	23.19	-0.05	16.29	20.68	18.97	0.19	15.64	19.68	18.39	0.28

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), ρ = Pearson Correlation in HR estimation.

B. Data Processing

To enhance the accessibility and usability of our dataset, we preprocessed the raw data and converted it into a convenient MAT file format compatible with both Matlab and Python. The videos were filmed at 30 frames per second with a resolution of 1280x720 pixels but were compressed to 320x240 pixels to facilitate storage and transmission. The PPG signals were downsampled from 200Hz to 30Hz to match the frame rate of the videos, resulting in 1800 frames per video. To enable researchers to explore the potential impact of various factors on rPPG, we assigned multiple labels, such as skin tone, gender, glasses, hair coverage, and makeup, to the dataset.

To ensure the synchronization of the videos and ground-truth PPG waves captured by different devices, we employed a Logitech Yeti microphone as an intermediary. Prior to each experiment, we recorded a chirp audio signal on both devices and then calculated the cross-correlation between the two recorded audio signals to determine the time delay between the phone and laptop. The timestamps of the oximeter were obtained through the USB COM port, allowing us to synchronize the PPG signals and video signals using two timestamps.

C. Data Samples

Figure 2 illustrates some samples from MMPD dataset. It includes Fitzpatrick skin types 3-6, four different lighting conditions (LED-high, LED-low, incandescent, natural), and four different activities (stationary, head rotation, talking, and walking).

IV. RESULT AND DISCUSSION

A. Unsupervised Signal Processing Methods

Six traditional unsupervised learning methods were evaluated on our dataset [19], [20], [17], [18], [16], [21]. In the skin tone comparison, we excluded the exercise, natural light, and walking conditions to eliminate any confounding factors and concentrate on the task at hand. Similarly, the motion comparison experiments excluded the exercise and natural light conditions, while the light comparison experiments excluded the exercise and walking conditions. This

TABLE III

BASELINE RESULTS ON THE MMPD DATASETS GENERATED USING THE RPPG-TOOLBOX [1]. FOR THE SUPERVISED METHODS WE SHOW RESULTS TRAINED ON THE UBFC-RPPG AND PURE.

Training Set	**UBFC** [7]				**PURE** [8]			
Testing Set	MMPD				MMPD			
	MAE↓	RMSE↓	MAPE↓	ρ ↑	MAE↓	RMSE↓	MAPE ↓	ρ ↑
Skin tone								
3	3.60	6.91	5.01	0.76	**3.06**	**6.60**	**4.06**	**0.77**
4	14.45	20.51	16.23	-0.12	8.94	15.74	9.98	0.25
5	10.06	13.72	14.11	0.45	12.39	16.51	16.74	0.12
6	14.88	20.21	16.85	0.18	15.43	20.98	17.51	0.20
Motion								
Stationary	5.34	11.17	6.32	0.56	5.91	11.56	7.13	0.54
Rotation	11.73	16.45	15.14	0.12	8.92	14.99	11.24	0.17
Talking	7.35	12.52	9.07	0.50	8.32	13.71	10.21	0.42
Walking	24.91	29.76	26.24	-0.02	27.21	31.97	28.56	0.03
Light								
LED-low	8.33	13.69	10.10	0.44	7.95	13.64	9.46	0.39
LED-high	7.92	13.18	10.14	0.40	7.80	13.37	10.00	0.37
Incandescent	8.18	13.83	10.29	0.37	7.40	13.46	9.12	0.38
Nature	10.41	16.77	12.37	0.36	11.04	17.66	12.52	0.35

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), ρ = Pearson Correlation in HR estimation.

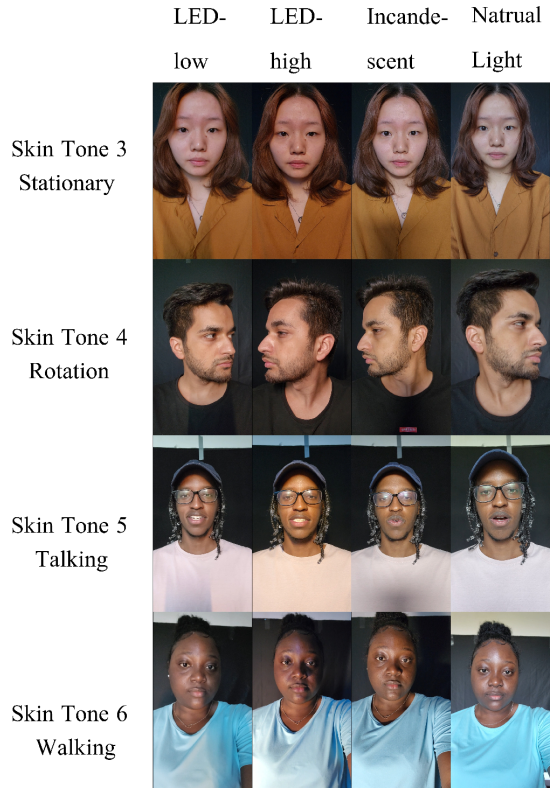

Fig. 2. Sample video frames across multiple domains such as skin tones, motions and lighting conditions.

approach enabled us to exclude confounding factors and better understand the unique challenges posed by each task.

As shown in Table 2, for unsupervised method comparison, the LGI [20] method performed best for relatively simple data from type 3 skin tone, while the POS [22] method had the best average performance for all conditions and robustness. For skin tone comparison, all the methods performed well on the data of skin type 3. However, for types 4, 5, and 6, most of the results showed a mean absolute error (MAE) greater than 10, indicating poor generalizability. For motion comparison, none of the models performed well for the hardest walking motion, but each model had its strengths for stationary, rotation, and talking tasks. For light comparison, there was no significant difference between the three types of artificial light, and all models performed poorly under natural light.

B. Supervised Deep Learning Methods

In this paper, we also investigated how a state-of-the-art supervised neural network performs on MMPD and studied the influence of skin tone, motion, and light. We used a pre-trained TS-CAN [23] model which was trained on the UBFC [7] and PURE [8] datasets. We used the same exclusion criteria as the evaluation on unsupervised methods.

Table 3 shows the results of the supervised neural network across different tasks. The results indicate that the neural network does not generalize well in all scenarios, as it only performs well on data from skin type 3 and with stationary tasks. This is because the training data (PURE and UBFC) only contains subjects in skin types 2-3 and mostly stationary videos. There is no significant difference between the models trained on the UBFC[7] and PURE[8] under the improved training framework of rPPG-toolbox [1].

C. Discussion

The discussion of our findings indicates that the performance of supervised and unsupervised methods varies depending on the similarity of test data to training data. In our study, we found that the generalizability of supervised methods is limited when tested on subjects with skin types 4-6 or under challenging motion and lighting conditions. Conversely, unsupervised methods exhibit better generalizability

as they do not rely on training. Our study also revealed that public rPPG datasets may not adequately represent real-world challenges encountered in MMPD dataset. Specifically, public datasets tend to have limited representation of skin types beyond types 2-3, mostly stationary videos, and uniform lighting conditions, leading to limited generalizability of supervised methods.

To improve the quality of data collection, we suggest using raw mode in the phone camera app to capture subtle changes in the face and properly positioning the phone. Additionally, minimizing complex signal processing and properly utilizing video processing tools such as ffmpeg can improve the quality of video frames and reduce time delays between the oximeter and phone. Face alignment and frame padding should also be considered, given the larger size of faces in mobile phone videos.

Overall, our study highlights the importance of properly selecting training and testing data and carefully considering the real-world challenges and limitations of data collection to improve the generalizability and accuracy of rPPG methods.

V. CONCLUSIONS

In this paper, we introduce the MMPD dataset, a collection of over 11 hours of video recording using a mobile phone. The dataset features subjects of four skin tones, in four motion conditions, and four lighting conditions, providing a diverse range of data for the benchmarking rPPG methods. With eight descriptive labels, the MMPD dataset aims to address the limitations of existing datasets recorded with mobile phones, particularly for videos of darker skin types and real-world motion and lighting tasks. The MMPD dataset and our evaluation of rPPG methods provide a step forward in advancing the accuracy and generalizability of this technology, with the potential to improve healthcare and other applications.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of China (NSFC) under Grant No. 62132010 and No. 62002198, Young Elite Scientists Sponsorship Program by CAST under Grant No.2021QNRC001, Tsinghua University Initiative Scientific Research Program, Beijing Key Lab of Networked Multimedia, and Institute for Artificial Intelligence, Tsinghua University. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

REFERENCES

[1] Xin Liu, Xiaoyu Zhang, Girish Narayanswamy, Yuzhe Zhang, Yuntao Wang, Shwetak Patel, and Daniel McDuff. Deep physiological sensing toolbox. *arXiv preprint arXiv:2210.00716*, 2022.

[2] Daniel McDuff. Camera measurement of physiological vital signs. *ACM Computing Surveys (CSUR)*, 2021.

[3] Ewa M Nowara, Daniel McDuff, and Ashok Veeraraghavan. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 284–285, 2020.

[4] Xin Liu, Yuntao Wang, Sinan Xie, Xiaoyu Zhang, Zixian Ma, Daniel McDuff, and Shwetak Patel. Mobilephys: Personalized mobile camera-based contactless physiological sensing. *arXiv preprint arXiv:2201.04039*, 2022.

[5] Hao Lu, Hu Han, and S Kevin Zhou. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12404–12413, 2021.

[6] Daniel McDuff. Deep super resolution for recovering physiological information from videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1367–1374, 2018.

[7] 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.

[8] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062. IEEE, 2014.

[9] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[10] Zheng Zhang, Jeff M Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.

[11] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video. *arXiv preprint arXiv:1810.04927*, 2018.

[12] Guillaume Heusch, André Anjos, and Sébastien Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017.

[13] Ewa Magdalena Nowara, Tim K. Marks, Hassan Mansour, and Ashok Veeraraghavan. Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In *Computer Vision and Pattern Recognition (CVPR), 1st International Workshop on Computer Vision for Physiological Measurement*, 2018.

[14] Amogh Gudi, Marian Bittner, and Jan van Gemert. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. *Applied Sciences*, 10(23):8630, 2020.

[15] Daniel McDuff, Miah Wander, Xin Liu, Brian L Hill, Javier Hernandez, Jonathan Lester, and Tadas Baltrusaitis. Scamps: Synthetics for camera measurement of physiological signals. *arXiv preprint arXiv:2206.04197*, 2022.

[16] Ming-Zher Poh, Daniel McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.

[17] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard De Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.

[18] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

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

[20] Christian S Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. Local group invariance for heart rate estimation from face videos in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1254–1262, 2018.

[21] Gerard De Haan and Arno Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.

[22] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[23] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-task temporal shift attention networks for on-device contactless vitals measurement. *NeurIPS*, 2020.