VitalVideos-Worldwide: A large and diverse rPPG dataset with rich ground truths

Toye Pieter-Jan
VitalVideos
West-Vlaanderen, Belgium
pj@vitalvideos.org

Abstract

We collected a new dataset consisting of 7000 unique participants from three geographical regions: Western-Europe, South-Asia and West-Africa. For every participant we recorded 60 seconds of video using two cameras simultaneously, an industrial camera and a smartphone. For ground truth we collected PPG waveform, respiration waveform, 3-lead ECG, heart rate, oxygen saturation and blood pressure. Gender, age and Fitzpatrick skin tone were also registered for every participant. The dataset was collected in a diverse set of locations to ensure a wide variety of backgrounds and lighting conditions. Request access to the dataset through vitalvideos.org

1. Introduction

The use of smartphone cameras for remote vital sign measurement has the potential to make a significant impact on a global scale. This technology is particularly promising because of the widespread availability of smartphones.

The existence of large and high-quality public datasets is essential for the rate of progress. Public datasets eliminate the need for researchers to collect their own data, which lowers the bar to entry. Furthermore, they enable us to replicate each other's work and facilitate benchmarking.

While there are already several public datasets [1, 2, 4–10, 12–25] available for researchers to use, most of these datasets are relatively small (10-100 participants) and primarily consist of young, male participants with lighter skin types. Many datasets also have a limited number of reference values, mostly limited to PPG waveform and heart rate although some datasets contain ground truths for respiration data [9, 12, 17, 19] or blood pressure [10, 25].

In an effort to address these limitations and contribute to both academic research as well as ongoing product development in this field, we have created a dataset that aims to include a large number of participants, has participants of all ages between 18-100, contains roughly equal numbers of participants of all skin tones and contains rich ground truth values. By providing researchers from both academia and industry with access to subsets of this data, we hope to enable both new research as well ongoing product development for rPPG applications.

2. Related work

As shown in Table 1, most publicly available rPPG datasets have focused on small, homogeneous participant groups with limited physiological ground truths. Popular datasets like PURE [20] AFRL [3], and UBFC-RPPG [1] are constrained to under 50 subjects, typically young adults with lighter skin tones, and often only include ground truth for the PPG waveform, heart rate and oxygen saturation. More recent datasets such as MSPM [19], VitalVideos-EU [23], and SCAMPS [12] have begun to address these limitations through larger, more diverse participant pools or synthetic data. These newer datasets also incorporate more ground truth values such as blood pressure or respiration rate. However, these datasets still fall short either in participant diversity, environmental diversity or missing ground truths (such as ECG).

Compared to this prior work, our dataset significantly expands both in breadth and depth. It includes over 7000 participants across three continents, ensuring balanced representation across gender, age, and Fitzpatrick skin type. Unlike most datasets that capture under-controlled lab conditions, our data spans over 30 locations including outdoor environments, with both natural and artificial lighting. Critically, each recording contains synchronized waveforms for PPG, respiration, and 3-lead ECG, alongside oxygen saturation, heart rate, and paired pre/post blood pressure values, making it the only large-scale public dataset with such comprehensive ground truth. This combination makes our dataset uniquely suited for training and benchmarking robust rPPG models in real-world conditions.

Table 1. Comparison of rPPG datasets (adapted from [11]).

Dataset	Subjects	Resolution	FPS	PPG	HR	SpO_2	BP	RR	ECG
MAHNOB-HCI [17]	27	780×580	61					✓	✓
BP4D [25]	140	1040×1392	24				✓		
VIPL-HR [14]	107	960×720	25	✓	✓	✓			
COHFACE [5]	40	640×480	20	✓					
UBFC-RPPG [1]	42	640×480	30	✓	✓				
UBFC-PHYS [13]	56	1024×1024	35	✓					
RICE CameraHRV [16]	12	1920×1200	30	✓					
MR-NIRP [15]	18	640×640	30	✓					
PURE [20]	10	640×480	60	✓		✓			
rPPG [8]	8	1920×1080	15		✓	✓			
OBF [9]	10	1920×1080	60	✓				✓	✓
PFF [6]	13	1280×720	50		✓				
VicarPPG [22]	20	720×1280	30	✓					
CleanerPPG [4]	10	1280×720	60	✓					✓
CMU [2]	14	25×25	15		✓				
SCAMPS [12]	2800	320×240	30	✓	✓			✓	
VitalVideos-EU [23]	850	1920×1080	30	✓	✓	✓	✓		
RLAP [24]	58	640x480	30	✓					
SUMS [10]	10	1280x720	60	✓		✓		✓	
MMPD [21]	70	1920x1080	90	✓	✓	✓			
MSPM [19]	103	1920x1080	30	✓	✓	✓	✓		
iBVP [7]	30	640x480	30	✓					
AFRL [3]	10	640x480	30	✓	✓	✓			
Ours	**7000**	**960×1056 / 1920×1080**	**60**	✓	✓	✓	✓	✓	✓

3. Methodology

3.1. Data collection procedure

Participants were seated at a distance of 40 to 70 cm in front of a tripod with a camera and smartphone. The blood pressure cuff, respiration belt, ECG leads and pulseoximeters were attached to the participant.

After some rest, blood pressure is measured two times to check for stability. The second result is registered if results indicate stability. The participant is then instructed to rest the arms and fingers and look straight into the camera. When the participant has complied, we record 60 videos from both cameras simultaneously as well as the reference values: both pulseoximeters, heart rate, oxygen saturation, respiration rate and ECG. As soon as the recording is finished a final blood pressure measurement is performed and registered.

3.2. Equipment

3.2.1. Waveform sensors

Two pulse oximeters were used. The first was a Contec CMS50D+ with a sampling frequency is rated at 60 Hz, with practical performance between 57–60 Hz, while HR and SpO_2 are sampled at 1 Hz. The second was a raw BVP sensor with a sampling frequency of 500 Hz. A respiration belt and 3-lead ECG (inner wrists and right ankle) sampling at 500 Hz completed the waveform sensor setup.

3.2.2. Blood pressure monitor

Blood pressure was measured using a MicroLife BP B2 Basic device equipped with small, medium, and large cuffs to accommodate arm circumference variation.

3.2.3. Cameras

Camera 1. Basler acA1440-220uc industrial camera with Basler C125-0418-5M lens, connected via USB 3.0. Settings: resolution 960×1056, 60 FPS, gain/ISO 0, exposure time 1/83s (12ms). Videos were recorded at 60 seconds length, using libx264 lossless compression (MP4 container), resulting in about 4GB per recording. Framing captured the entire face, chest, and shoulders.

Camera 2. iPhone 15 Pro smartphone. Settings: resolution 1920×1080, 60 FPS, SDR video. Recordings were 60 seconds, encoded in ProRes (slightly lossy) within a MOV container, yielding about 3GB per recording. Framing captured the entire face, neck, and parts of the upper chest.

3.3. Software

We made use of a custom GUI to point the camera, monitor waveform quality and enter participant metadata.

3.4. Recording environments

Recording was done in over 30 different locations, inside and outside, making use of both artificial and natural light. When light intensity was too low a ring light was added to supply extra light to the face of the participant.

3.5. Dataset statistics

Figure 1 shows the basic demographic information of participants along with the distribution of systolic and diastolic blood pressure values.

3.6. Sample data and waveform examples

Figure 2 shows blurred and anonymized sample frames from both the industrial camera and the smartphone. Figures 3 and 4 show 10-second segments of raw PPG wave-

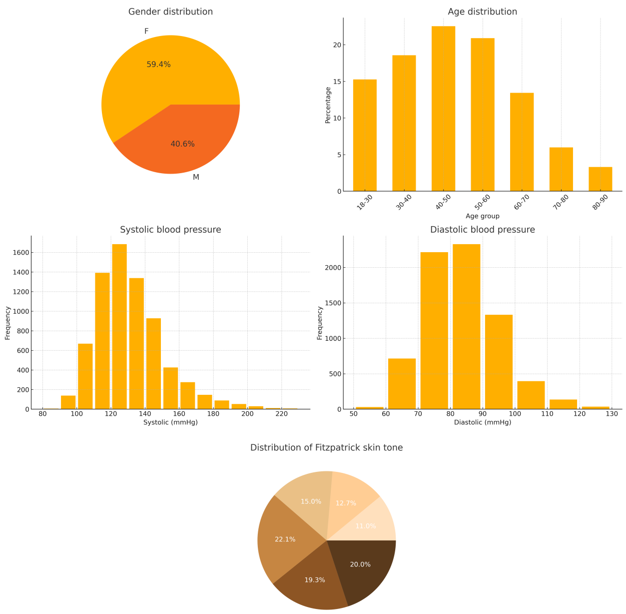

Figure 1. Dataset demographics and BP distribution.

(a) Basler frame

(b) iPhone frame

Figure 2. Left: industrial camera. Right: smartphone camera.

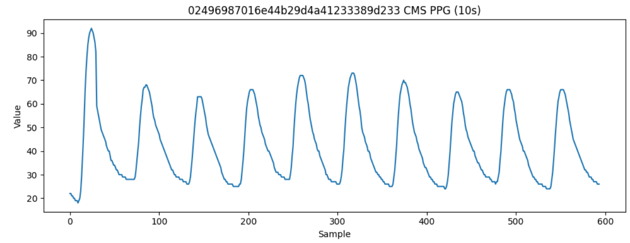

Figure 3. 10sec PPG waveform 1

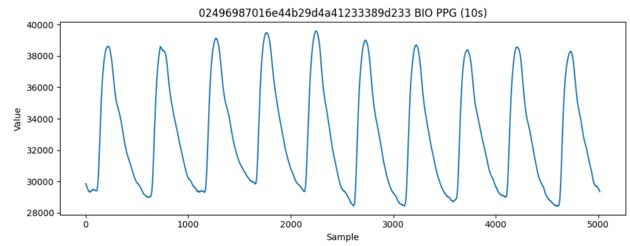

Figure 4. 10sec PPG waveform 2

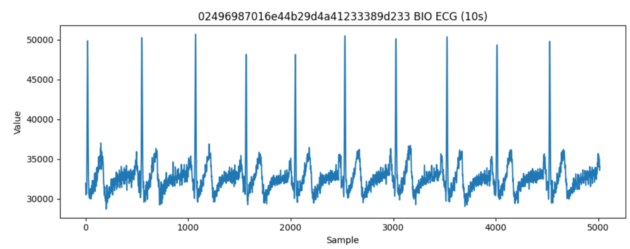

Figure 5. 10sec ECG waveform

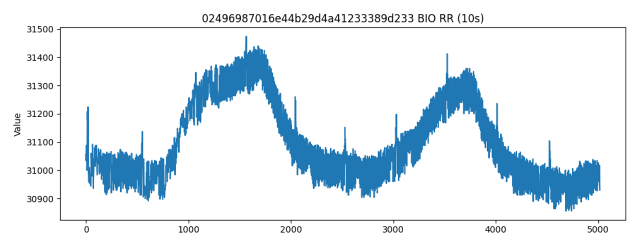

Figure 6. 10sec RR waveform

forms from two participants. Figures 5 and 6 display 10-second excerpts of ECG and respiration signals, respectively. Finally, Figure 7 illustrates the structure and contents of the ground truth metadata file associated with each recording.

4. Baseline results

For baseline evaluation, we applied three standard rPPG algorithms—G, CHROM, and POS—on a random 70 participant subset of the dataset to estimate heart rate and assess pulse signal quality. These methods were implemented using open-source code. We report the mean absolute error (MAE) for heart rate and signal-to-noise ratio (SNR) for pulse waveform quality. Among the tested methods, POS achieved the lowest mean absolute error (0.915 BPM) and the best SNR (10.513 dB).

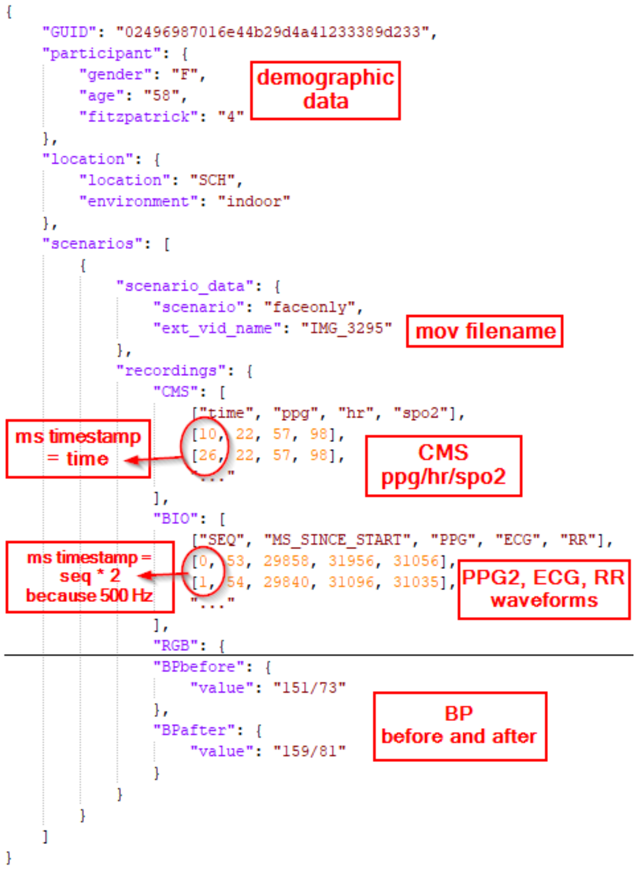

Figure 7. JSON ground truth file structure

Table 2. Preliminary results on a subselection of the dataset

Method	**Heart rate** MAE	**Pulse wave** SNR
G	9.000	-10.483
CHROM	1.840	7.548
POS	0.915	10.513

5. Discussion

5.1. Reflection and suggestions for future datasets

The inclusion of an iPhone 15 Pro smartphone in this dataset improves ecological validity by introducing high-quality mobile recordings. Future expansions could benefit from also incorporating an Android device, which would help expose models to differences in the Android video processing pipeline. Adding a near-infrared (NIR) camera could further extend modality diversity, enabling exploration of cross-spectral approaches.

Another valuable addition would be blood sample data, including parameters such as hemoglobin or glucose levels. These could enable investigation into whether metabolic or hematological markers can be inferred from facial video.

Lastly, while we included pre- and post-recording blood pressure measurements, continuous beat-to-beat BP recording (e.g., via Finapres) would offer more precise temporal alignment with facial signals. However, such equipment is costly and time-consuming to set up, which currently limits its feasibility in large-scale data collection.

5.2. Limitations

While the dataset offers substantial scale and diversity, a few limitations remain. The gender distribution is slightly skewed, with females being overrepresented relative to males. Despite recordings being captured under a wide range of lighting conditions, from 200 lux to 5000 lux, certain edge cases such as very low-light environments or direct sunlight were not included. Finally, some ECG and respiration signals exhibit noise artifacts in a subset of participants, which may stem from individual skin characteristics, ambient electromagnetic interference, or suboptimal fit of the respiration belt around the waist.

6. Ethics and availability

6.1. Ethical approval and consent

This study underwent Institutional Review Board (IRB) review in each of the 3 locations which resulted in two approval letters and one formal determination that IRB approval was not required.

All participants provided consent to participate in the dataset. This consent form grants permission to share the collected data with universities, research institutes, and companies for the purposes of research and development (product creation/validation).

6.2. Availability

Anyone can request to use a part of the dataset for research or model development purposes.

If you or your organization will directly or indirectly profit from the use of this dataset then an appropriate fee will be determined. If you or your organization doesn't directly or indirectly profit from the use of this dataset and you promise to publish your work publicly so that others can replicate your work then you may apply to use part of the dataset without any charge.

You can request access through vitalvideos.org.

7. Funding

The creation of this dataset was funded from personal savings and revenues from industrial licenses.

8. Thanks

We thank the numerous people who helped to create this dataset through their participation, assistance or funding.

We also thank Philipp Rouast for the open source implementations of the G, CHROM and POS algorithms used in calculating the baseline results.

9. Conclusion

We believe this dataset can serve as a valuable foundation for future research and product development in remote physiological measurement. Its key strengths include its unprecedented scale, as well as its demographic diversity, with balanced representation across age, gender, and skin tone. The recordings span over 30 locations, both indoor and outdoor, capturing a wide range of backgrounds and lighting conditions. Each session includes synchronized high-quality waveforms for PPG, respiration, and ECG, along with pre- and post-recording blood pressure measurements. Finally, the dataset offers high-bitrate video from both an industrial camera and a modern smartphone, providing data suitable for a wide variety of modeling pipelines.

References

[1] 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. 1, 2

[2] Ananyananda Dasari, Sakthi Kumar Arul Prakash, László A. Jeni, and Conrad S. Tucker. Evaluation of biases in remote photoplethysmography methods. *NPJ Digital Medicine*, 4: 1–13, 2021. 1, 2

[3] J.R. Estepp, E.B. Blackford, and C.M. Meier. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, page 1462–1469, 2014. 1, 2

[4] 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. 1, 2

[5] Guillaume Heusch, André Anjos, and Sébastien Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017. 2

[6] Gee-Sern Hsu, Arul Murugan Ambikapathi, and Ming-Shiang Chen. Deep learning with time-frequency representation for pulse estimation from facial videos. In *Proceedings of the 2017 IEEE International Joint Conference on Biometrics*, pages 383–389, 2017. 2

[7] Jitesh Joshi and Youngjun Cho. ibvp dataset: Rgb-thermal rppg dataset with high resolution signal quality labels. *Electronics*, 13(7):1334, 2024. 2

[8] Mikhail Kopeliovich and Mikhail Petrushan. Color signal processing methods for webcam-based heart rate evaluation. In *Proceedings of the SAI Intelligent Systems Conference*, pages 703–723, 2019. 2

[9] Xiaobai Li, Iman Alikhani, Jingang Shi, Tapio Seppanen, Juhani Junttila, Kirsi Majamaa-Voltti, Mikko Tulppo, and Guoying Zhao. The obf database: A large face video database for remote physiological signal measurement and atrial fibrillation detection. In *Proceedings of the 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 242–249, 2018. 1, 2

[10] Ke Liu, Jiankai Tang, Zhang Jiang, Yuntao Wang, Xiaojing Liu, Dong Li, and Yuanchun Shi. Summit vitals: Multi-camera and multi-signal biosensing at high altitudes. *arXiv preprint arXiv:2409.19223*, 2024. 1, 2

[11] Daniel McDuff. Camera measurement of physiological vital signs. *ACM Computing Surveys*, 55(9):1–40, 2023. 2

[12] 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. 1, 2

[13] Rita Meziatisabour, Yannick Benezeth, Pierre De Oliveira, Julien Chappe, and Fan Yang. Ubfc-phys: A multimodal database for psychophysiological studies of social stress. *IEEE Transactions on Affective Computing*, 2021. Early access. 2

[14] 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. 2

[15] 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 *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018. 2

[16] Amruta Pai, Ashok Veeraraghavan, and Ashutosh Sabharwal. Camerahrv: Robust measurement of heart rate variability using a camera. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, page 105010S. International Society for Optics and Photonics, 2018. 2

[17] 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. 1, 2

[18] Jeremy Speth, Nathan Vance, Adam Czajka, Kevin W. Bowyer, Diane Wright, and Patrick J. Flynn. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. *arXiv preprint arXiv:2106.06583*, 2021.

[19] Jeremy Speth, Nathan Vance, Benjamin Sporrer, Lu Niu, Patrick J. Flynn, and Adam Czajka. Mspm: A multi-site physiological monitoring dataset for remote pulse, respiration, and blood pressure estimation. *arXiv preprint arXiv:2402.02224*, 2024. 1, 2

[20] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062, 2014. 1, 2

[21] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi domain mobile video physiology dataset. *arXiv preprint arXiv:2302.03840*, 2023. 2

[22] H. Emrah Tasli, Amogh Gudi, and Marten Den Uyl. Remote ppg based vital sign measurement using adaptive facial regions. In *Proceedings of the 2014 IEEE International Conference on Image Processing*, pages 1410–1414, 2014. 2

[23] Pieter-Jan Toye. Vitalvideos-europe: A dataset of face videos with ppg and blood pressure ground truths. *arXiv preprint arXiv:2306.11891*, 2023. 1, 2

[24] Kegang Wang, Yantao Wei, Jiankai Tang, Yuntao Wang, Mingwen Tong, Jie Gao, Yujian Ma, and Zhongjin Zhao. Camera based hrv prediction for remote learning environments. *arXiv preprint arXiv:2305.04161*, 2024. 2

[25] Zheng Zhang, Jeff M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, and 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. 1, 2