Camera Viewpoint Effects on Remote Photoplethysmography

Maximilian P. Oppelt[1,2], Muthu Sai Charan Venkata Raman[1], Anne-Marie Lehmeyer[1], Nadine R. Lang-Richter[1], Thomas Wittenberg[1]

[1] Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Department Digital Health and Analytics, Erlangen, Germany
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, Department Artificial Intelligence in Biomedical Engineering, Erlangen, Germany

Abstract

Remote photoplethysmography (rPPG) enables unobtrusive cardiovascular monitoring in clinical settings where cameras rarely achieve optimal frontal alignment. Given that rPPG signal quality depends on the geometric relationship between facial surface orientation, illumination, and camera position, lateral viewpoints may degrade heart rate estimation accuracy. However, the isolated effect of static camera perspective, independent of head motion, remains poorly characterized. We systematically evaluated eight state-of-the-art deep learning rPPG models using recordings from 49 healthy adults captured with six synchronized cameras spanning frontal to lateral perspectives. Synchronized photoplethysmography (PPG) provided ground truth. All models were evaluated zero-shot, with performance stratified by residual facial motion. Results demonstrate consistent performance degradation for lateral viewpoints across all model-dataset combinations. The best-performing model, iBVP-Net pretrained on PURE, achieved 1.81 bpm Mean Absolute Error (MAE) frontally, degrading to 2.69 bpm for side cameras. EfficientPhys pretrained on UBFC-rPPG showed the lowest degradation ratio (1.37×), while DeepPhys reached up to 3.15×. Direct grouped comparison confirmed significantly higher MAE for lateral versus frontal cameras at all motion levels (Low Mean Landmark Displacement (MLD): Δ=1.56 bpm, p=0.0014; Mid: Δ=1.49, p=0.0008; High: Δ=1.39, p=0.0014; Wilcoxon signed-rank, BH-corrected), confirming that static camera angle alone degrades estimation accuracy by approximately 20–27% independently of subject movement.

These findings indicate that camera placement critically affects rPPG reliability for clinical cardiovascular monitoring and underscore the need for viewpoint-aware training strategies to enable robust deployment in real-world healthcare environments.

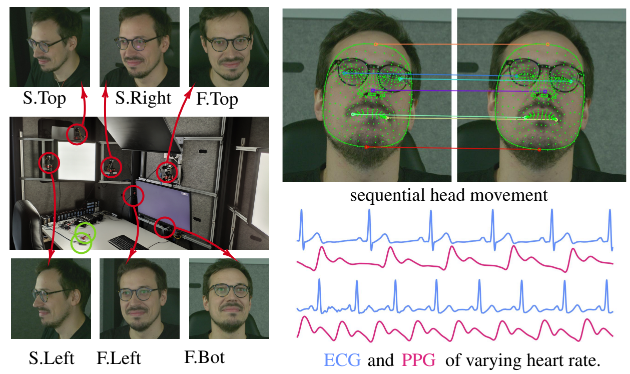

Figure 1. Multi-view recording setup: camera arrangement, exemplary frames from six perspectives with landmark-based motion tracking, and synchronized ECG and PPG waveforms.

1. Introduction

Camera-based remote photoplethysmography (rPPG) enables unobtrusive cardiovascular monitoring in diverse clinical and care settings such as waiting rooms, bedside monitoring, and robot-assisted nursing [1]. In these settings, patients are rarely observed by a camera with perfect frontal alignment. Instead, cameras often capture the face from varying perspectives and distances. Because rPPG signal depends on the geometric relationship between facial surface orientation, illumination, and camera position [2, 3], non-frontal viewpoints may systematically affect signal quality. Yet the extent to which static camera perspective, independent of head motion, degrades heart rate (HR) estimation accuracy remains poorly characterized. This study addresses that gap by evaluating state-of-the-art deep rPPG models zero-shot across six fixed camera angles under controlled conditions, with results stratified by residual motion.

Prior rPPG work has extensively studied motion artifacts, illumination changes, and cross-dataset domain shift [3–5], with physics-based models confirming inherent viewpoint sensitivity [2, 3]. Head rotation degrades single-camera performance via yaw-induced face occlusion [5–7], yet these studies conflate subject motion with angular variation. Recently, Egorov et al. [8] introduced a multi-view dataset and reported substantial accuracy degradation for side-view cameras compared to frontal views, yet their analysis aggregated viewpoints coarsely and did not stratify by motion intensity. To the best of our knowledge, the effect of static camera perspective on rPPG signal quality, under tightly controlled conditions with minimal head movement, remains underexplored. In this work, we systematically evaluate state-of-the-art deep rPPG models zero-shot across six fixed camera angles and stratify performance by residual motion level, isolating viewpoint as an independent factor affecting estimation accuracy.

2. Methods

We recorded facial video data using six synchronized Basler acA1920-150uc industrial cameras equipped with Fujinon HF25XA-5M lenses (f/1.6, 25 mm focal length). All cameras captured RGB video at 1920×1080 resolution and 25 fps, subsequently resampled to 30 fps via nearest-neighbor interpolation to match the input requirements of standard rPPG models. Cameras were arranged at varying azimuth and elevation angles relative to the subject's approximate nasion position, as detailed in Table 1. Diffuse, non-pulsed LED panels provided controlled illumination to avoid interference with rPPG signal extraction. Physiological ground truth was acquired using a BIOPAC MP160 system (2000 Hz). Three-lead ECG and finger-clip photoplethysmography (PPG) were recorded simultaneously. A hardware trigger pulse synchronized all six camera frames with physiological recordings at sub-frame temporal precision. Forty-nine healthy adults (27 male, 22 female; 32.4 ± 12.0 years) participated. Subjects were seated and performed an image-based arousal rating task to elicit natural HR variability while maintaining a forward-facing posture (32.6 ± 12.2 min sessions). Ethics board approval and written informed consent were obtained. Faces were detected and cropped per frame using YOLOv5 [9]. Bounding boxes were scaled to include peripheral regions and resized to model-specific inputs (72×72 to 128×128 px). Frames without detections were excluded. Non-overlapping 30-second

Table 1. Camera positions relative to approximate nasion

Camera	Azimuth (°)	Elevation (°)	Distance (mm)
Frontal-Bottom	81	27	1323
Frontal-Top	74	15	1093
Frontal-Left	58	17	836
Side-Right	35	1	813
Side-Top	0	19	867
Side-Left	0	−22	882

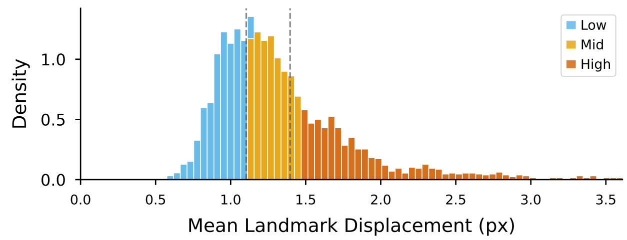

Figure 2. Distribution of MLD across all valid segments. Dashed vertical lines indicate tertile thresholds defining subsets of low/mid/high pixel displacement.

segments were extracted via greedy sliding-window selection to maximize independent segments per subject. A segment was valid if the PPG-based and ECG-based HR differed by $\leq$ 1 bpm, ensuring high-quality labels. This criterion, together with outlier removal outside 40–120 bpm, removed segments corrupted by contact PPG artifacts. To quantify residual facial motion within each segment, we extracted facial landmarks per frame using MediaPipe Face Mesh [10] and computed the MLD $\text{MLD} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\left\|\mathbf{p_i^{(t+1)}} - \mathbf{p_i^{(t)}}\right\|_2$, where $N = 478$ denotes the number of landmarks, T the number of frames in the segment, and $\mathbf{p_i^{(t)}} \in \mathbb{R}^\mathbf{2}$ the pixel coordinates of landmark i at frame t. MLD thus represents the cumulative displacement averaged across all facial landmarks, expressed in pixels. In order to enable subgroup analysis, we performed motion stratification by globally dividing segments into tertiles (low/mid/high) based on MLD values (Figure 2). We evaluated eight state-of-the-art end-to-end deep learning models for rPPG: TS-CAN, PhysNet, DeepPhys, PhysFormer, EfficientPhys, RhythmFormer, PhysMamba, and iBVPNet. All models were obtained from the rPPG-Toolbox framework [11] with pre-trained weights from public datasets including PURE, SCAMPS, UBFC-rPPG, and BP4D. We evaluated all model–dataset combinations for which pre-trained weights were available in the rPPG-Toolbox, yielding the entries shown in Table 2. Evaluation followed a zero-shot cross-dataset protocol: models were applied directly to our multi-perspective recordings without fine-tuning. Inference window length was set according to model-specific defaults as implemented in rPPG-Toolbox. For each segment, HR was estimated from the predicted rPPG signal using the systolic peak detection algorithm proposed by Elgendi et al. [12], and compared against the PPG-derived ground-truth HR. The MAE between the rPPG-estimated HR and the PPG-derived ground-truth HR was computed, averaged across the test set and reported with standard error, separately for each camera perspective and stratification subgroup to characterize viewpoint-dependent performance degradation and its interaction with motion intensity.

3. Results

MAE values across six camera perspectives for eight supervised rPPG models pretrained on four public datasets are presented in Table 2. Across all model-dataset combinations, frontal views consistently outperformed lateral perspectives. iBVPNet pretrained on PURE achieved the lowest overall errors, with MAE of 1.81 bpm (Frontal-Bottom) degrading to 2.69 bpm (Side-Top).

EfficientPhys was the most stable across viewpoints: UBFC-rPPG pretraining yielded the lowest degradation ratio (1.37×; 2.40 bpm frontal to 3.28 bpm lateral), while PURE reached 1.58×. DeepPhys showed the highest sensitivity (3.11× on UBFC-rPPG; 3.15× on PURE).

Models pretrained on SCAMPS, a fully synthetic dataset, exhibited substantially elevated errors across all projections, likely reflecting the domain shift between rendered avatars and real skin tissue. Performance degradation patterns differed systematically by pretraining corpus: PURE- and UBFC-rPPG-trained models showed strongest deterioration for Side-Top and Side-Left projections, whereas BP4D-pretrained models degraded most severely for Side-Right views, potentially reflecting viewpoint biases in the respective training distributions.

MAE stratified by residual motion level (MLD tertiles) is shown in Figure 3. Wilcoxon signed-rank tests revealed significant performance differences across motion strata for frontal projections. Side projections exhibited less consistent stratification, with significance primarily in the mid-to-high transition. Direct comparison of grouped frontal versus lateral cameras confirmed significantly higher MAE at all motion levels (Low MLD: median MAE 5.84 vs. 7.40 bpm, Δ=1.56, W=19, p=0.0014; Mid: 5.87 vs. 7.37, Δ=1.49, W=12, p=0.0008; High: 7.24 vs. 8.63, Δ=1.39, W=20, p=0.0014; Wilcoxon signed-rank, BH-corrected). All 27 pairwise frontal–lateral camera comparisons were likewise significant (all $p < 0.026$, BH-corrected), confirming that static camera angle alone degrades estimation accuracy by approximately 20–27%.

4. Discussion

Viewpoint-dependent degradation has direct clinical implications: camera placement critically affects rPPG reliability independent of subject motion. In bedside or waiting-room scenarios, current models may yield non-viable estimates from lateral perspectives, with some configurations exceeding 20 bpm MAE.

Prior work has characterized motion artifacts and cross-dataset domain shift extensively, yet viewpoint effects remained confounded with head rotation. Our controlled setup isolates static perspective as an independent degradation factor: even under minimal motion, lateral cameras exhibited systematically elevated errors. Motion-robust architectures alone are therefore insufficient and explicit viewpoint invariance requires separate consideration.

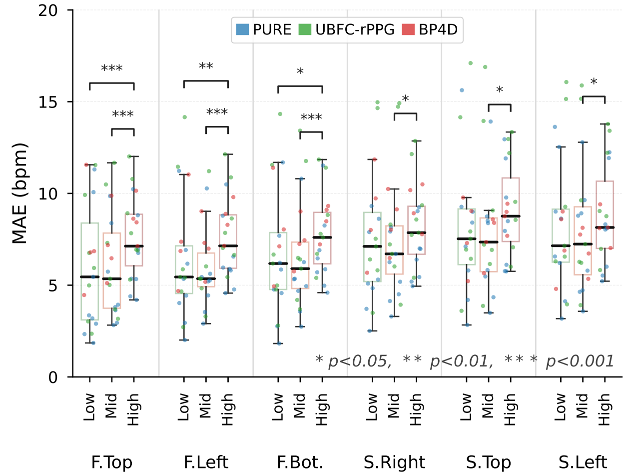

Figure 3. MAE (bpm) per projection stratified by MLD tertiles. Box plots aggregate eight rPPG models across training datasets. Two-sided Wilcoxon signed-rank tests matched on model and dataset.

Divergent degradation patterns across pretraining datasets show that training viewpoint composition shapes generalization predictably. Future work should incorporate diverse angular coverage through multi-view augmentation or synthetic generation, and develop unified frameworks that disentangle projection from motion effects.

Several limitations should be noted. Our cohort comprised 49 healthy young adults recorded under controlled diffuse illumination; generalization to elderly or cardiovascular-impaired populations, variable lighting, and naturalistic clinical environments remains to be validated. Furthermore, we did not stratify performance by skin tone, which is known to affect rPPG signal quality and may interact with viewpoint-dependent degradation.

Acknowledgments

The authors gratefully acknowledge Fabian Grosch and Stephan Göb for their invaluable technical assistance in establishing the hardware infrastructure for dataset acquisition. Funding: Bavarian Research Society ”FOR Social Robots”, AZ-1594-23.

References

[1] Wittenberg T, Ernst A, Hauenstein T, Merz N, Sassen S, André E, Lu S, Pest C, Hutschenreiter S, Berger J, Franke J, Reitelshöfer S, Garbas J, Münzenmayer C, Oppelt M, Lang-Richter N. A Concept of Emotion-Driven Human-Robot-Interaction for Nursing Care Scenarios. In 2024 2nd International Conference on Integrated Systems in Medical

Table 2. MAE (bpm) per model and camera projection, reported as mean $\pm$ standard error.

Training	Model	Frontal-Top	Frontal-Left	Frontal-Bottom	Side-Right	Side-Top	Side-Left
PURE	TS-CAN	3.00 ± 0.16	3.42 ± 0.16	4.19 ± 0.21	4.81 ± 0.22	5.71 ± 0.34	6.09 ± 0.35
	PhysNet	5.23 ± 0.25	5.53 ± 0.24	5.60 ± 0.26	8.09 ± 0.35	8.56 ± 0.35	8.80 ± 0.35
	DeepPhys	2.59 ± 0.15	3.85 ± 0.18	3.73 ± 0.20	4.58 ± 0.21	7.04 ± 0.35	8.15 ± 0.38
	PhysFormer	10.04 ± 0.45	10.99 ± 0.43	11.75 ± 0.45	8.17 ± 0.36	8.33 ± 0.38	13.72 ± 0.44
	EfficientPhys	2.58 ± 0.13	2.76 ± 0.15	2.52 ± 0.14	3.25 ± 0.17	3.91 ± 0.19	3.98 ± 0.19
	RhythmFormer	9.29 ± 0.59	5.74 ± 0.47	6.06 ± 0.49	7.56 ± 0.51	6.31 ± 0.47	5.04 ± 0.40
	PhysMamba	2.41 ± 0.14	5.08 ± 0.25	5.17 ± 0.27	4.12 ± 0.25	12.09 ± 0.49	10.48 ± 0.48
	iBVPNet	1.90 ± 0.14	1.95 ± 0.13	1.81 ± 0.12	2.10 ± 0.13	2.69 ± 0.17	2.62 ± 0.16
SCAMPS	TS-CAN	10.04 ± 0.34	10.69 ± 0.35	11.38 ± 0.38	11.98 ± 0.40	11.68 ± 0.39	11.67 ± 0.38
	PhysNet	20.98 ± 0.52	18.84 ± 0.50	19.07 ± 0.50	18.24 ± 0.48	18.24 ± 0.46	18.36 ± 0.47
	DeepPhys	4.27 ± 0.20	3.91 ± 0.17	3.68 ± 0.18	4.19 ± 0.19	5.27 ± 0.27	5.82 ± 0.27
	PhysFormer	18.04 ± 0.61	21.09 ± 0.62	18.99 ± 0.58	20.82 ± 0.61	24.17 ± 0.65	18.78 ± 0.54
	EfficientPhys	14.68 ± 0.42	15.13 ± 0.43	14.74 ± 0.43	14.31 ± 0.42	16.12 ± 0.56	16.03 ± 0.54
UBFC-rPPG	TS-CAN	4.50 ± 0.21	5.25 ± 0.23	5.88 ± 0.25	6.83 ± 0.27	7.62 ± 0.48	7.66 ± 0.48
	PhysNet	11.94 ± 0.46	11.25 ± 0.44	11.63 ± 0.45	14.45 ± 0.49	14.34 ± 0.47	15.19 ± 0.49
	DeepPhys	2.70 ± 0.15	4.04 ± 0.19	4.26 ± 0.22	5.42 ± 0.25	7.37 ± 0.31	8.39 ± 0.36
	PhysFormer	10.42 ± 0.43	13.71 ± 0.51	13.95 ± 0.50	15.01 ± 0.48	17.03 ± 0.52	15.84 ± 0.50
	EfficientPhys	2.40 ± 0.13	2.54 ± 0.12	2.67 ± 0.15	3.01 ± 0.15	3.28 ± 0.15	3.28 ± 0.16
	RhythmFormer	5.02 ± 0.38	4.42 ± 0.33	3.81 ± 0.32	6.51 ± 0.43	6.60 ± 0.42	6.33 ± 0.38
	PhysMamba	6.62 ± 0.34	5.30 ± 0.30	6.15 ± 0.35	7.11 ± 0.35	5.99 ± 0.30	6.02 ± 0.30
BP4D	TS-CAN	6.59 ± 0.30	6.45 ± 0.29	6.73 ± 0.30	7.71 ± 0.30	7.32 ± 0.38	6.78 ± 0.39
	PhysNet	10.49 ± 0.52	10.14 ± 0.54	10.16 ± 0.50	11.79 ± 0.55	9.40 ± 0.43	8.85 ± 0.42
	DeepPhys	5.98 ± 0.29	6.88 ± 0.29	7.70 ± 0.30	9.02 ± 0.33	8.15 ± 0.35	8.23 ± 0.35
	EfficientPhys	4.39 ± 0.23	4.23 ± 0.23	4.44 ± 0.24	5.40 ± 0.24	4.65 ± 0.24	4.42 ± 0.22

Technologies (ISMT). Erlangen, Germany: IEEE. ISBN 979-8-3503-7552-7, September 2024; 1–6.

[2] Wong KL, Wei Chin J, Chan TT, Odinaev I, Suhartono K, Tianqu K, So RHY. Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, LA, USA: IEEE. ISBN 978-1-6654-8739-9, June 2022; 2164–2170.

[3] Maity AK, Wang J, Sabharwal A, Nayar SK. RobustPPG: Camera-based robust heart rate estimation using motion cancellation. Biomedical Optics Express October 2022; 13(10):5447. ISSN 2156-7085, 2156-7085.

[4] Vance N, Speth J, Sporrer B, Flynn P. Promoting Generalization in Cross-Dataset Remote Photoplethysmography. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, BC, Canada: IEEE. ISBN 979-8-3503-0249-3, June 2023; 5985–5993.

[5] Li J, Yu Z, Shi J. Learning Motion-Robust Remote Photoplethysmography through Arbitrary Resolution Videos. Proceedings of the AAAI Conference on Artificial Intelligence June 2023;37(1):1334–1342. ISSN 2374-3468, 2159-5399.

[6] Li X, Chen J, Zhao G, Pietikainen M. Remote Heart Rate Measurement from Face Videos under Realistic Situations. In 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE. ISBN 978-1-4799-5118-5, June 2014; 4264–4271.

[7] Zheng K, Ci K, Cui J, Kong J, Zhou J. Non-Contact Heart Rate Detection When Face Information Is Missing during Online Learning. Sensors December 2020;20(24):7021. ISSN 1424-8220.

[8] Egorov K, Botman S, Blinov P, Zubkova G, Ivaschenko A, Kolsanov A, Savchenko A. Gaze into the Heart: A Multi-View Video Dataset for rPPG and Health Biomarkers Estimation, 2025.

[9] Qi D, Tan W, Yao Q, Liu J. YOLO5Face: Why Reinventing a Face Detector. In Karlinsky L, Michaeli T, Nishino K (eds.), Computer Vision – ECCV 2022 Workshops, volume 13805. Cham: Springer Nature Switzerland. ISBN 978-3-031-25071-2 978-3-031-25072-9, 2023; 228–244.

[10] Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, Zhang F, Chang CL, Yong MG, Lee J, Chang WT, Hua W, Georg M, Grundmann M. MediaPipe: A Framework for Building Perception Pipelines, 2019.

[11] Liu X, Narayanswamy G, Paruchuri A, Zhang X, Tang J, Zhang Y, Sengupta R, Patel S, Wang Y, McDuff D. rPPG-Toolbox: Deep remote PPG toolbox. In Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S (eds.), Advances in Neural Information Processing Systems, volume 36. Curran Associates, Inc., 2023; 68485–68510.

[12] Elgendi M, Norton I, Brearley M, Abbott D, Schuurmans D. Systolic Peak Detection in Acceleration Photoplethysmograms Measured from Emergency Responders in Tropical Conditions. PLoS ONE October 2013;8(10):e76585. ISSN 1932-6203.

Address for correspondence:

Am Wolfsmantel 33, 91058 Erlangen, Germany
maximilian.oppelt@iis.fraunhofer.de